ORIGINAL RESEARCH ARTICLE

Drug Saf 2012; 35 (8): 645-653
0114-5916/12/0008-0645/$49.95/0

Comparison of Three Pharmacovigilance Algorithms in the ICU Setting

A Retrospective and Prospective Evaluation of ADRs

Sandra L. Kane-Gill,[1,2,3] *Elizabeth A. Forsberg,*[4] *Margaret M. Verrico*[1,2] and *Steven M. Handler*[5,6,7,8]

1 Department of Pharmacy and Therapeutics, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA
2 Department of Pharmacy, University of Pittsburgh Medical Center – Presbyterian/Shadyside, Pittsburgh, PA, USA
3 Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, PA, USA
4 Department of Pharmacy, University of Pittsburgh Medical Center – St Margaret Hospital, Pittsburgh, PA, USA
5 Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
6 Geriatric Pharmaceutical Outcomes and Geroinformatics Research and Training Program, Pittsburgh, PA, USA
7 Division of Geriatric Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
8 Geriatric Research, Education and Clinical Centers (GRECC) and The Center for Health Equity Research and Promotion, Veterans Affairs Pittsburgh Healthcare System, Pittsburgh, PA, USA

Abstract

Background: Pharmacovigilance algorithms are used to assess the likelihood of adverse drug reaction (ADR) occurrence. The preferred instrument for use in the intensive care unit (ICU) is not established.

Objective: The primary objective of this study was to compare the agreement between the Kramer algorithm, Naranjo criteria and Jones algorithm for the evaluation of ADRs in the ICU. A secondary objective was to compare the agreement between the same pharmacovigilance algorithms for ADR determination when applied in a retrospective versus concurrent fashion in the ICU.

Study Design: There were two phases in this study. Phase I was the retrospective evaluation (i.e. after the patient was discharged from the hospital) conducted in patients admitted during July 2005 to June 2006. Phase II was the concurrent phase (i.e. while the patient was in the hospital) conducted over 6 weeks in 2008. Both phases were conducted at the University of Pittsburgh Medical Center and included adult patients admitted to the medical ICU.

Intervention: In phase I, a random sample of 261 medication signals were evaluated individually for potential ADRs using the Kramer algorithm, Naranjo criteria and Jones algorithm. In phase II, an active medication monitoring system was used to detect five abnormal laboratory values, resulting in a random sample of 253 signals that were evaluated using the same three algorithms.

Main Outcome Measure: Percentage agreement among the algorithms for all levels of causality was estimated using a kappa statistic for both phases of the study.

Results: For phase I, the kappa values were all >0.7 ranging from 0.721 to 0.855 between instruments, with Naranjo versus Kramer having the highest kappa, which is considered excellent agreement. The kappa statistic between individual instruments for phase II are <0.7 ranging from 0.423 to 0.635, which is considered moderate agreement, with Naranjo versus Jones displaying the lowest kappa while still exhibiting moderate agreement. For phase II, the Kramer algorithm had better agreement with both the Naranjo criteria and the Jones algorithm.

Conclusions: These instruments demonstrated similar results for evaluating ADRs in the ICU retrospectively, suggesting that instrument selection with any of the three instruments is reasonable. For concurrent ADR evaluations, there is greater variability in the level of causality obtained among pharmacovigilance algorithms and Kramer displayed better agreement with its comparators. A suggestion for a more definitive concurrent ADR assessment is to use more than one algorithm. This may be challenging in daily clinical practice; however, it is a reasonable expectation for research.

Background

Fatalities related to adverse drug reactions (ADRs) have been estimated to rank between the fourth and sixth leading causes of death in the US.[1] These unintended and potentially noxious effects of medicines can have a particularly profound effect for intensive care unit (ICU) patients due to the tenuous nature of this patient population.[2,3] Preventable ADRs have been documented to occur at a rate of 13–19 events per 1000 patient days in the ICU.[3,4] Among critically ill patients, those in the medical ICU have shown significantly higher rates of preventable ADRs compared with surgical ICUs.[3]

ADR evaluation in the ICU and other settings can be a complex and time-consuming endeavor because it involves the collection and assessment of multiple potential contributing factors. A clinician's past experience with a suspected ADR can impact their evaluation and subsequent resulting action.[5] Due to these challenges, inconsistency in interpretation exists regarding the causality of a given reaction and as a result a variation in the action taken to counteract the event. In 1976, a study by Karch et al.[6] showed that without the use of structured causality algorithms, the rate of agreement between three clinicians who evaluated the same ADRs was found to be as low as 50%. Almost 20 years later, a similar study found the rate of agreement between five clinicians to be only 17.3%.[7] These studies emphasize that without using a standardized approach to ADR assessments, there is significant potential for disagreement when evaluating ADRs.

Pharmacovigilance algorithms have been developed to help clinicians standardize ADR assessments and determine the likelihood that a suspected medication is responsible for a given ADR.[8] Many algorithms or instruments for ADR causality assessment are currently available,[9-34] each with a different format and ease of use. Algorithms have improved inter-rater agreement markedly, from 41% to 57% clinician agreement without the use of an algorithm, to 83–92% agreement with the use of an algorithm.[34] Although there is no established gold-standard algorithm, the Naranjo criteria is the most frequently published scale used in clinical practice and is required as an ADR assessment tool by at least one journal.[35,36]

ADR determination tools are used to evaluate the causal relationships between drugs and unwanted effects. A clinically useful application of these tools is in the evaluation of suspected ADRs during a patient's hospital admission. For example, a signal such as an abnormal laboratory value could be evaluated in an effort to prevent or mitigate an adverse effect before it causes patient harm.[37] Signals are clues that suggest the presence of an ADR. With so many ADR pharmacovigilance algorithms available to choose from it would be helpful to know how the instruments compare in a practical evaluation as the one described above.[38] However, to date, comparisons of ADR determination algorithms have been conducted in a retrospective manner by chart review after the patient has been discharged.[8]

A retrospective study using a limited number of ADRs (n = 28) compared the Kramer and Jones algorithms and Naranjo criteria in an adult medical population and showed a modest agreement between the three scales.[39] These three pharmacovigilance algorithms have not yet been evaluated for agreement in the critically ill population, an environment where there are many more potential confounders that make it more difficult to assess ADRs. Additionally, these instruments have not been compared for use between retrospective identification of ADRs and ADRs occurring during the patient's admission. It may be expected that the time of evaluation could affect the outcome of the causality assessment due to differences summarized in table I. Agreement between ADR pharmacovigilance algorithms matters because if all instruments perform equally then selection can be discretionary. The purpose of this study is twofold: (i) to compare the agreement among three commonly used pharmacovigilance algorithms in the medical ICU; and (ii) to compare the differences in retrospective versus concurrent evaluation of ADRs.

Methods

This was a two-phase study. Phase I was the retrospective evaluation (i.e. after the patient was discharged from the hospital) and phase II was the concurrent phase (i.e. while the patient was in the hospital). Both phases were conducted in a 32-bed, adult, medical ICU.

Description of Pharmacovigilance Algorithms

A description of the pharmacovigilance algorithms is provided in table II. The Naranjo criteria, also referred to as the Adverse Reaction Probability Scale, is a 10-item questionnaire that categorizes the probability of an ADR as doubtful, possible, probable and definite.[34]

The original Kramer algorithm, also known as the ADR scoring system or Yale algorithm, uses specific rules for the operational assessment of ADRs.[9] A scoring system was incorporated into a six-axis decision process that classified reactions into definite, probable, possible and unlikely based on 56 questions. Later, a modified Kramer algorithm was constructed that still contains six axes but the questions are simplified and condensed

Table I. Advantages and disadvantages of retrospective vs concurrent analysis of adverse drug reactions

Description of evaluation mechanism	Advantages	Disadvantages
Retrospective Post-discharge	Takes less time to complete since data already exists Rechallenge information may already exist	No ability to discuss with prescriber/other clinicians Potential for missing information in chart decreases accuracy
Concurrent during patient's admission	Ability to talk to clinician/family More realistic and clinically applicable Ability to access data that are missing/incomplete in the chart Can request tests or laboratory values to aid in evaluation	Rechallenge information may not be available at the time of review, so cannot complete a comprehensive assessment using most instruments Day of evaluation dictates information available May need to wait on test results to complete an ADR assessment

ADR = adverse drug reaction; **Rechallenge** = exposing the patient to the suspected drug a second time to determine if the patient has the same response, i.e. being an ADR.

Table II. Description of instruments and scoring

Instrument	Criteria evaluated	Score range	Category
Modified Kramer[40]	ADR known Alternative causes Temporal sequence Drug level Dechallenge Rechallenge	+7 to –7	Definite: 6–7 Possible: 0–3 Probable: 4–5 Unlikely: <0
Naranjo et al.[34]	ADR known Temporal sequence Dechallenge/antidote Rechallenge Alternative causes Placebo response Drug level Effect of dose adjustment Exposure to similar drug Objective evidence	+13 to –3	Definite: ≥9 Probable: 5–8 Possible: 1–4 Doubtful: ≤0
Jones[10]	Temporal association Dechallenge Rechallenge	A numeric scale is not used	Remote Possible Probable Highly probable

ADR = adverse drug reaction; **Dechallenge** = discontinuing the suspected drug to determine if the adverse drug reaction dissipates; **Rechallenge** = exposing the patient to the suspected drug a second time to determine if the patient has the same response, i.e. being an ADR.

into a table that is easier to use.[40] The questions assess similar domains described for the Naranjo criteria and the probability categories are similar, i.e. unlikely, possible, probable and definite.

The Jones algorithm contains five questions and is constructed so as not to allow continuation to the next question without a positive response to the prior question.[10] The final assessment is not numerical like the other algorithms but does contain a probability categorization of remote, possible, probable and highly probable.

Criteria for the Selection of Pharmacovigilance Algorithms to Evaluate

The Naranjo criteria was selected for evaluation because it is the most commonly used published instrument.[36] The Kramer and Jones algorithms were selected because of their ease of use. Also, at least three studies comparing the Naranjo criteria, and Kramer and Jones algorithms have been published but not in the ICU and only using a retrospective approach, therefore this study will add to existing information and provide a useful comparison to the literature.[8,39,40]

This study was conducted in two phases: a retrospective chart review phase and a concurrent chart review phase.

Phase I

The retrospective evaluation phase occurred following the patient's discharge and the methods for the evaluation of ADRs for phase I are described further in another publication.[41] In brief, a random sample of 261 medication antidote administrations was selected to evaluate through retrospective chart review. Adult patients admitted to the medical ICU of the University of Pittsburgh Medical Center Presbyterian Hospital from 1 July 2005 to 30 June 2006 were considered eligible if they received one of five pre-determined antidote medications (dextrose 50%, diphenhydramine, methylprednisolone, phytonadione, sodium polystyrene) designated as triggers or signals. Each signal was evaluated for the probability of an ADR using the Modified Kramer algorithm,[40] Naranjo criteria[34] and the Jones algorithm[10] by one of two clinical pharmacists who were trained in the application of the algorithms. Phase I was approved as an exempt study by the University of Pittsburgh Institutional Review Board.

Phase II

The concurrent evaluation phase occurred during the patient's admission to the ICU and consisted of using an active medication monitoring system which generates an alert when a potential ADR is present, to evaluate five pre-specified abnormal laboratory value signals (elevated blood urea nitrogen, vancomycin concentration, quinidine concentration, low sodium, low glucose) in the medical ICU from 10 June 2008 to 25 July 2008.[39] Overall, 253 signals were evaluated. The ADR evaluation occurred on the day when the laboratory value met the abnormal criteria for Monday through Friday alerts, and on Monday for alerts generated on the weekend. Each signal was evaluated using the same ADR assessment tools: the Modified Kramer algorithm, Naranjo Probability Scale and the Jones algorithm by a

clinical pharmacist (SK-G) trained in the use of pharmacovigilance algorithms.[10,34,40] The use of possible or greater likelihood for the determination of an ADR is a typical criteria.[4,38,42-45] The methods of this study are described elsewhere.[38] Phase II was approved by the institution's Total Quality Council as a quality improvement evaluation.

Analysis

Statistical analysis was completed using the Statistical Package for the Social Sciences (SPSS), version 16.0 (SPSS Software, Chicago, IL, USA). Percentage agreement among the algorithms for all levels of causality was estimated using a kappa statistic for both phases of the study. Reliability categories for the kappa statistic were excellent agreement (kappa = 0.8–0.99) and moderate agreement (kappa = 0.4–0.79).[46] Chi-squared test was used to determine significance in proportions for the summary of the level of certainty in each phase. An α of 0.05 was considered statistically significant. Finally, a description of identical agreement between assessments (signals in phase I, abnormal laboratory values in phase II) for each phase is provided.

Results

A summary of the levels of certainty for each instrument in each phase is presented in table III. For phase I, use of the Naranjo criteria resulted in significantly more probable assessments than the Jones algorithm. This difference in probable assessments was also true for phase II. In addition, phase II demonstrated significantly more probable assessments when using the Naranjo criteria versus Kramer algorithm. The Jones alogrithm was significantly more conservative in the probable assessment than Kramer in the phase II evaluation. Despite the variability in the probable assessment the definite and possible categories exhibited greater consistency between instruments.

The level of agreement between algorithms for all levels of causality is provided in table IV. For phase I, the kappa values are all >0.7 between individual instruments, with the Naranjo criteria versus Kramer algorithm having the highest kappa score, which is considered excellent agreement. The kappa value between individual instruments for phase II are <0.7 ranging from 0.423 to 0.635, which is considered moderate agreement. Comparison of the Naranjo criteria versus Jones algorithms yielded the lowest kappa, which still qualified as moderate agreement.

Overall, in terms of agreement for each signal assessment, the level of certainty was identical for 87.7% (229/261), 86.6% (226/261) and 93.1% (243/261) for the Kramer versus Jones, Jones versus Naranjo and Naranjo versus Kramer algorithms, respectively, in phase I. There were 79.4% (201/253), 64.8% (164/253) and 77.1% (195/253) identical responses for each abnormal laboratory assessment for the Kramer versus Jones, Jones versus Naranjo and Naranjo versus Kramer algorithms, respectively, in phase II.

Discussion

The ICU is an environment where ADR assessment can be more challenging than general

Table III. Summary for level of certainties in each phase for each instrument

Category	Phase I – retrospective evaluation (n = 261) [% (n)]			Phase II – concurrent evaluation (n = 253) [% (n)]		
	Kramer	Naranjo	Jones	Kramer	Naranjo	Jones
Definite/highly probable	5.4 (14)	4.6 (12)	8.0 (21)	3.5 (9)	2.8 (7)	3.6 (9)
Probable	11.1 (29)	13.4 (35)[a]	6.5 (17)	10.7 (27)[b]	18.2 (46)[b]	2.8 (7)[b]
Possible	13.4 (35)	12.3 (32)	15.3 (40)	31.3 (79)	34.4 (87)	33.6 (85)
Remote/doubtful/unlikely	70.1 (183)	69.7 (182)	70.1 (183)	54.5 (138)	44.6 (113)[c]	60.1 (152)

a Naranjo vs Jones p = 0.009.

b Naranjo vs Jones p < 0.001; Naranjo vs Kramer p = 0.016; Kramer vs Jones p < 0.001.

c Naranjo vs Jones p = 0.001; Naranjo vs Kramer p = 0.026.

Table IV. Agreement between instruments for all levels of causality

Instrument	Phase I – retrospective evaluation kappa statistic (95% CI)	Phase II – concurrent evaluation kappa statistic (95% CI)
Naranjo vs Kramer	0.855 (0.794, 0.916)	0.635 (0.555, 0.715)
Naranjo vs Jones	0.721 (0.650, 0.792)	0.423 (0.337, 0.509)
Kramer vs Jones	0.743 (0.672, 0.814)	0.635 (0.553, 0.717)

medical care units because patients have more acute conditions and more drugs prescribed.[2,3] The use of ADR pharmacovigilance algorithms aids in the evaluation of ADRs, especially since relying on clinical opinion has not been shown to be reliable.[34] This study demonstrated that agreement between instruments is moderate to excellent for an ADR analysis that occurs in ICU patients after hospital discharge. A previous study that compared the three instruments for retrospective evaluations from ADRs of an unspecified patient population, demonstrated agreement to be moderate to less than moderate with kappa scores ranging from 0.28 to 0.48.[41] Another retrospective ADR evaluation demonstrated congruency between the Naranjo criteria and Kramer algorithm to be 95% and Kramer algorithm and Jones algorithm to be 56%.[32] The use of algorithms for evaluation on the day of the ADR event during hospital admission (concurrent) poses more variability between instruments than a retrospective evaluation.

The variation for the level of certainty on individual causality rankings is interesting; however, from a clinical standpoint does this matter? Since several studies[38,42-45] accept the categorization of possible or greater as an ADR, then perhaps clinicians simply need to understand how the scales perform at assessing ADRs using this widely agreed upon threshold of an ADR. The retrospective evaluation (phase I) of the study indicates that the number of ADRs using a definition of possible or greater would be similar regardless of which instrument was used. However, in the prospective evaluation (phase II), there are significantly more ADRs using a definition of possible or greater if the Naranjo criteria is applied instead of the Jones. Ten percent more ADRs were included using the Naranjo criteria instead of the Kramer algorithm; however, this difference was not statistically significant.

A possible explanation for variations observed in phase II for ADR determination is that all instruments in this evaluation contain a dechallenge question. Unless drug discontinuation occurs immediately the user may not have sufficient information to answer this question. The Kramer algorithm and Naranjo criteria allow for a 'do not know' response; however, the Jones algorithm does not. If a dechallenge does not occur, the Jones criteria only permits a 'possible' categorization; therefore, the user may not proceed to answer further questions in this algorithm. Lack of information to complete the questionnaires at the time of assessment results in additional variation in responses; however, this is a realistic application of these algorithms. Clinicians cannot wait until they have all of the information required of an instrument to perform an ADR assessment when a medication-related problem is initially suspected.

In this study, we found divergent rates of agreement between causality determination algorithms depending upon whether the evaluations were done retrospectively (post-patient discharge) or concurrently with the patient's admission to the ICU. Our data suggest that the selection of an ADR assessment algorithm may have less of an impact when the analysis is done retrospectively compared with concurrently. This finding can be due to many factors, which are summarized in table I.

A validity study of the Naranjo criteria conducted in the ICU showed only moderate concordance compared with expert opinion in the evaluation of ADRs, retrospectively.[47] In this study, using a retrospective chart review, the Naranjo criteria is comparable in ADR assessment with both the Jones and Kramer scales. This finding suggests that testing against other instruments for the detection of an ADR, using a

category of possible or greater as the definition appears moderate to excellent. The major assumption with this comparison is that all of the instruments are valid (i.e. measuring an ADR) because good inter-rater agreement does not prove validity. Again, of note, the Naranjo criteria and the Kramer algorithm have been tested for validity outside of the ICU.

The intent of the three algorithms used in this study was to evaluate ADR and not harm associated with a medication referred to as an ADR.[48] An ADR in both of the phases was defined as "an undesirable clinical manifestation that is consequent to and caused by the administration of a particular drug" according to Kramer et al.[9] The moderate validity compared with expert opinion associated with the Naranjo criteria[47] has resulted in researchers applying a more rigorous method requiring agreement in rank of possible or greater between multiple algorithms to be considered an ADR.[38] A previous study showed that applying criteria of three out of three, two out of three and one out of three agreements between algorithms resulted in ADR rates of 36%, 44% and 60%, respectively.[39] It is known that clinical assessment has substantial variability and instruments clearly contain less variability than ranking by clinicians, therfore the use of instruments is still an advantage.[34]

With our experience using the algorithms we do have some suggestions for potential improvements and areas that require further clarification to assist with use. For the Jones algorithm, having the dechallenge as the second step is difficult since it is not realistic for many medications (e.g. insulin, β-blockers) to be abruptly discontinued based on suspicion and no definitive evidence in a critically ill population. As such, interchanging the dechallenge with de-escalation may be more appropriate. Due to the dechallenge being a stop point in the algorithm, it is difficult to score higher than possible for a concurrent assessment compared with other algorithms, which may potentially explain the findings in phase II. Unique to the Jones criteria, the answers are discrete instead of giving the option of 'do not know'. Rechallenges are not always clinically practical for any of the algorithms. For example, you would not want to intentionally reintroduce an episode of drug-induced hypotension in a critically ill patient.

The Kramer algorithm uses a numerical scale for categorization, which includes a score of zero, indicating a 'possible' association between the adverse event and drug treatment being evaluated. This categorization can be confusing since zero is typically associated with a negative response and could be mistakenly interpreted as an 'unlikely' association. In addition, the Kramer algorithm has more than one option per axis, which can create variability between reviewers depending on subjective interpretation.

The Naranjo criteria requests conclusive reports for the reaction; however, it is uncertain how many case reports would be deemed conclusive. One question in the Naranjo instrument pertains to objective evidence and another inquires about drug concentrations. However, these may be synonymous depending on the user's interpretation. It may be ideal if the concept of objective evidence represented as signs and symptoms is presented as mutually exclusive from the question about drug levels to prevent misinterpretation. Also, there may be undue emphasis on blood levels since many drug concentrations are not routinely obtained,. Additionally, not all ADRs involved toxic or supratherapeutic drug concentrations.

Limitations

It is possible that the evaluation of ADRs detected using antidote medications in phase I induces less variability in assessment using instruments than in phase II, when ADRs were detected using abnormal laboratory values as a signal. However, ADR causality assessment tools were developed as general instruments to assess all types of ADRs. It is possible that different evaluators between phase I and phase II could have influenced the results but the Naranjo criteria and Kramer algorithm have been tested for reliability. We focused on three instruments; however, other ADR causality assessment instruments could have been used, potentially yielding different results. The definition of an ADR used in this

evaluation was a categorization of possible or greater and this could be viewed as a limitation because the ADR definition of probable or greater could be applied, potentially changing the variability between instruments.

Conclusions

This study demonstrates that agreement between pharmacovigilance algorithms is at least moderate for ADRs in the ICU. Since possible or greater likelihood rankings by causality instruments are typically the criteria of an ADR, then retrospectively it may be acceptable to use any of the three causality algorithms evaluated in this study without a significant variation in response. However, increased variability is observed between instruments for ADR determination while the patient is hospitalized. A suggestion for a more definitive ADR assessment is to use more than one algorithm. This may be challenging in daily clinical practice; however, it is a reasonable expectation for research or when further review of a significant event is required. Clinicians utilizing these instruments should be familiar with their advantages and limitations.

Acknowledgements

This work was funded in part by an Agency for Healthcare Research and Quality grant (R01-HS017695) and National Institute on Aging grant (K07AG033174) awarded to Dr Steven Handler. The authors have no conflict of interests to disclose.

References

1. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA 1998; 279: 1200-5
2. Kane-Gill SL, Weber RJ. Principles and practices of medication errors in the ICU. Crit Care Clin 2006; 22: 273-90
3. Cullen DJ, Sweitzer BJ, Bates DW, et al. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care units and general care units. Crit Care Med 1997; 25 (8): 1289-97
4. Rothschild JM, Landrigan CP, Cronin JW, et al. The Critical Care Safety Study: the incidence and nature of adverse events and serious medical errors in intensive care. Crit Care Med 2005; 33 (8): 1694-700
5. Weingart SN, Toth M, Sandz DZ, et al. Physician's decisions to override computerized drug alerts in primary care. Arch Intern Med 2003; 163: 2625-31
6. Karch FE, Smith CL, Kerzner B, et al. Adverse drug reactions: a matter of opinion. Clin Pharmacol Ther 1976; 19: 489-92
7. Arimone Y, Begaud B, Miremone-Salame G, et al. Agreement of expert judgment in causality assessment of adverse drug reactions. Eur J Clin Pharmacol 2005; 61: 169-73
8. Benahmed S, Picot MC, Hillaire-Buys D, et al. Comparison of pharmacovigilance algorithms in drug hypersensitivity reactions. Eur J Clin Pharmacol 2005; 61: 537-41
9. Kramer MS, Leventhal JM, Huchinson TA, et al. An algorithm for the operational assessment of adverse drug reactions. JAMA 1979; 242: 623-32
10. Jones JK. Definition of events associated with drugs: regulatory perspectives. J Rheumatol 1988; 15: 14-9
11. Irey NS. Tissue reactions to drugs. Am J Pathol 1976; 82: 613-48
12. Karch FE, Lasanga L. Toward the operational identification of adverse drug reactions. Clin Pharmacol Ther 1997; 21: 247-54
13. Dangomanu J, Evereux JC, Jouglard J. Method for determination of undesirable effects of drugs [in French]. Therapie 1978; 33: 373-81
14. Begaud B, Evreux JC, Jouglard J, et al. Imputation of the unexpected or toxic effects of drugs: actualization of the method used in France. Therapie 1985; 40: 111-8
15. Blanc S, Leuenberger P, Berger JP, et al. Judgments of trained observers on adverse drug ractions. Clin Pharmacol Ther 1979; 25: 493-8
16. Emanueli A, Sacchetti G. An algorithm for the classification of untoward events in large scale clinical trials. Agents Actions Suppl 1980; 7: 318-22
17. Evreux JC, Loupi E, Descotes J, et al. Evaluation of the documentation on adverse effects of drugs: proposition 2. Imputability [in French]. Therapie 1982; 37: 657-70
18. Kitaguchi T, Nojiri T, Suzuki S, et al. Some assessment systems for industry post marketing adverse drug reaction (ADR) information. Iyakuhin Kenkyu 1983; 38: 303-18
19. Lagier G, Vincens M, Castot A. Imputability in drug monitoring: principles of the balanced drug reaction assessment method and principal errors to avoid. Therapie 1983; 38: 303-18
20. Cornelli U. The phase IV monitoring studies: examples of naproxen Na multi-centre Italian trial. In: Crescioni C, James JM, editors. Postmarketing clinical trials: the phase V studies. Paris: Editions de Sante, 1984: 89-105
21. Stephens MD. Assessment of causality in industrial setting. Special workshop: industrial. Drug Inf J 1984; 18: 307-13
22. Castle WM. Assessment of causality in industrial settings. Drug Inf J 1984; 18: 297-302
23. Turner WM. The Food and Drug Administration algorithm. Special workshop: regulatory. Drug Inf J 1984; 18: 259-66
24. Venulet J, Ciucci A, Berneker GC. Standardised assessment of drug-adverse reaction associations: rationale and experience. Int J Clin Pharmacol Ther Toxicol 1980; 18: 381-8
25. Loupi E, Ponchon AC, Ventre JJ, et al. Imputability of a teratogenic effect [in French]. Therapie 1986; 41: 207-10
26. Stricker BHC. Diagnosis and causality assessment of drug-induced hepatic injury. In: Dukes MNG, editor. Drug-induced hepatic injury. Amsterdam: Elsevier, 1985: 1-13

27. Benichou C, Danan G. Causality assessment in the European pharmaceutical industry: presentation of the preliminary results of a new method. Drug Inf J 1992; 26: 589-92
28. Hoskins RE, Mannino S. Causality assessment of adverse drug reactions using decision support and informatics tools. Pharmacoepidemiol Drug Saf 1992; 1: 235-49
29. Danan G, Benichou C. Causality assessment of adverse reactions to drugs: I. A novel method based on the conclusions of international consensus meetings: application to drug-induced liver injuries. J Clin Epidemiol 1993; 46: 1323-30
30. Hsu PH, Stoll RW. Causality assessment of adverse events in clinical trials: II. An algorithm for drug causality assessment. Drug Inf J 1993; 27: 387-94
31. Maria VA, Victorino RM. Development and validation of a clinical scale for the diagnosis of drug-induced hepatitis. Hepatology 1997; 26: 664-9
32. Koh Y, Shu CL. A new algorithm to identify the causality of adverse drug reactions. Drug Saf 2005; 28: 1159-61
33. Horn JR, Hansten PD, Chan LN. Proposal for a new tool to evaluate drug interaction cases. Ann Pharmacother 2007; 41: 674-80
34. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. Clin Pharmacol Ther 1981; 30: 329-45
35. Kelly WN. The quality of adverse drug event reports. Ann Pharmacother 2003; 37: 1774-8
36. Kane-Gill SL, Devlin JW. Adverse drug event reporting in the intensive care unit: a survey of current practices. Ann Pharmacother 2006; 40: 1267-73
37. Steinman MA, Handler SM, Gurwitz JH, et al. Beyond the prescription: medication monitoring and adverse drug events in older adults. J Am Geriatr Soc 2011; 59 (8): 1513-20
38. Kane-Gill SL, Visweswaran S, Saul MI, et al. Computerized detection of adverse drug reactions in the medical intensive care unit. Int J Med Inform 2011; 80 (8): 570-8
39. Michel DJ, Knodel LC. Comparison of three algorithms used to evaluate adverse drug reactions. Am J Hosp Pharm 1986; 43 (7): 1709-14
40. Goh CL. An approach to the evaluation and documentation of adverse drug reaction. Sing Med J 1989; 30: 283-5
41. Kane-Gill SL, Bellamy CJ, Verrico MM, et al. Evaluating the positive predictive values of antidote signals to detect potential adverse drug reactions (ADRs) in the medical intensive care unit (ICU). Pharmacoepidemiol Drug Saf 2009; 18: 1185-91
42. Oderda GM, Said Q, Evans RS, et al. Opioid-related adverse drug events in surgical hospitalizations: impact on costs and length of stay. Ann Pharmacother 2007; 41: 400-7
43. Dormann H, Muth Selbach U, Krebs S, et al. Incidence and costs of adverse reactions during hospitalization. Drug Saf 2000; 22: 161-8
44. Evans RS, Pestotnik SL, Classen DC, et al. Preventing adverse drug events in hospitalized patients. Ann Pharmacother 1994; 28: 523-7
45. Kilbridge PM, Campbell UC, Cozart HB, et al. Automated surveillance for adverse drug events a community hospital and an academic medical center. J Am Med Inform Assoc 2006; 13: 372-7
46. Stokes MB, Davis CS, Knoch GC. Categorical data analysis using the SAS system. Cary (NC): SAS Institute Inc., 1995
47. Kane-Gill SL, Kirisci L, Pathak DS. Are the Naranjo criteria reliable and valid for determination of adverse drug reactions in the intensive care unit? Ann Pharmacother 2005; 39 (11): 1823-7
48. Morimoto T, Gandhi TK, Seger AC, et al. Adverse drug events and medication errors: detection and classification methods. Qual Saf Health Care 2004; 13: 306-14

Correspondence: Dr *Sandra L. Kane-Gill*, PharmD, MSc, FCCP, FCCM, Associate Professor of Pharmacy and Clinical Translational Sciences, University of Pittsburgh, School of Pharmacy, Center for Pharmacoinformatics and Outcomes Research, 918 Salk Hall, 3501 Terrace St, Pittsburgh, PA 15261, USA.
E-mail: SLK54@pitt.edu

