Reliability and accuracy of the screening for adverse events in Brazilian hospitals

ANA LUIZA BRAZ PAVÃO, LUIZ ANTÔNIO BASTOS CAMACHO, MÔNICA MARTINS, WALTER MENDES AND CLÁUDIA TRAVASSOS

Oswaldo Cruz Foundation, Rio de Janeiro, Brazil

Address reprint requests to: Cláudia Travassos, Oswaldo Cruz Foundation, Rio de Janeiro, Av. Brasil, no 4365 sala 214, Manguinhos, 21045-360 Rio de Janeiro-RJ, Brazil. Tel: +5521 3865-3244; Fax: +5521 2270-2668; E-mail: claudia.maria.travassos@gmail.com

Abstract

Objective. To analyze the reliability and accuracy of the screening for adverse events (AEs) conducted by nurses taking the assessment by medical residents as the reference.

Design. A validation study of the screening phase of a previous retrospective cohort study based on the patient record review that estimated the incidence of AEs (base study).

Setting. Three general teaching hospitals in the State of Rio de Janeiro, Brazil.

Participants. A subsample of 242 medical records randomly selected from an original sample of 1103 previously evaluated records.

Main Outcome Measure(s). A two-step approach was used for analysis: the identification of at least one screening criterion (first endpoint) and the validation of each identified criterion (second endpoint), taking the assessment by medical residents as the reference. Kappa coefficient; simple percentage agreement; sensitivity; specificity; positive and negative predictive values were calculated.

Results. The total agreement between medical residents and nurses on the presence of screening criteria was moderate (78.9%, $K = 0.55$). Specificity (81.6%) was higher than sensitivity (74.4%). Nurses detected more screening criteria that were later confirmed as true AEs (179 vs. 171, respectively). Significant differences in the detection of the screening criteria: 'Other complications', 'Hospitalization injury' and 'Prior admission' were observed.

Conclusion. The results suggested a good performance of the nurses in the screening for detection of AE and showed significant differences in relation to detection of specific screening criteria among reviewers. A better understanding of the screening process and the performance of reviewers was provided.

Keywords: agreement, sensitivity, specificity, screening, adverse events

Introduction

Patient safety has gained prominence in the international debate about the quality of health care due to the magnitude of adverse events (AEs) and their presence worldwide [1]. The concept of AE has been essential to the analysis of this issue; an AE is defined as an incident related to health care that results in unnecessary harm to the patient [2].

In recent decades, the studies to detect AEs have used different approaches; the methods based on the retrospective chart review are the most common [3]. The *Harvard Medical Practice Study* was a pioneering study in which the medical records were screened for AE indicators, followed by a review for confirmation [4]. According to this methodology, the screening criteria are markers of AEs and are detected in a screening phase, usually conducted by nurses. Then, the criteria detected will be confirmed or not in a second stage review, conducted by physicians. In subsequent studies in several countries, methodological differences seem to explain, in part, variations in results [5–14]. Patient safety research through the chart review should be reliable and accurate without losing efficiency. The few published studies on the reliability of AE detection indicate reasonable to moderate agreement between physicians and nurses [15, 16]. Similarly, the studies assessing the accuracy of AE screening are scarce and usually show higher sensitivity than specificity [17, 18].

The present study aimed to analyze the reliability and accuracy of the screening for AEs conducted by nurses taking

the assessment by medical residents as the reference, using two endpoints: identification of any screening criterion by the reviewers (at least one criterion detected in the sample) and the same analysis for each criterion detected by nurses and medical residents.

Methods

This is a validation study of the AE screening method used in previous retrospective cohort study ('base study') that estimated the incidence of AEs in a sample of patients admitted to three general teaching hospitals in the state of Rio de Janeiro, Brazil, in 2003 [19]. The reliability and accuracy of AE screening conducted by nurses in the base study were assessed through the analysis of a subsample of patient records selected for review by medical residents in the present study, taken as the reference for the assessment of accuracy ('gold standard'). A flowchart is presented to enlighten the rationale of the present study (Fig. 1).

In the base study, a random sample of 1103 medical records was assessed. Patients younger than 18 years, with lengths of stay under 24 h, and psychiatric patients were excluded. The base study using the methodology of the Canadian study is composed of two phases: the screening phase, conducted by nurses, and the evaluation phase, conducted by experienced physicians [9]. Screening criteria (triggers) were translated and adapted to the Brazilian context [20]. The detection of at least one screening criterion during the screening stage selected patient records for the evaluation phase. In this phase, those physicians conducted an implicit full assessment of the records to state their accordance with the screening criteria detected in the previous stage and verified the occurrence of an AE. All reviewers had >20 years of clinical experience. Four nurses were responsible for

medical record's review in the three hospitals. The reliability of their screening in the base study was good [19].

The present study used a random subsample of 259 medical records from three large general teaching hospitals thought to approximate a broader representation of the average quality standards of teaching hospitals in the state of Rio de Janeiro. The sample size was calculated with an expected agreement of 80%, sensitivity and specificity of 80%, significance level of 5% and 40% of cases meeting at least one criterion. Screening was performed by two internal medicine residents who were familiar with the methods and forms used in the base study. Hence, the residents repeated the screening conducted by nurses in the base study using the same methodology [9, 19].

The validation of the screening phase conducted by nurses in the base study was performed using a two-step approach: first, are shown the results for reliability and accuracy for any screening criterion detected (first endpoint); and then, the same results for the most frequent screening criteria identified in the base study (second endpoint). Medical residents were considered the gold standard in accuracy analysis.

A specific analysis of criteria 9 ('Other complications') and 19 ('Other unwanted events') was performed, as these were generic criteria and the most frequent ones in the screening stage of the base study [19]. Results of this specific analysis will be presented as a part of the second endpoint analysis. Medical records discordant for these criteria were selected. Based on the description given by nurses in the base study, discordant cases for criterion 9 were classified into: (i) 'injury or damage not described in other criteria' (signs, symptoms and disorders), (ii) 'event related to invasive procedure' (including transfusion and hemodialysis reactions) and (iii) 'falls and other accidental trauma'. For criterion 19, four categories were created: (i) 'injury or damage not described in other
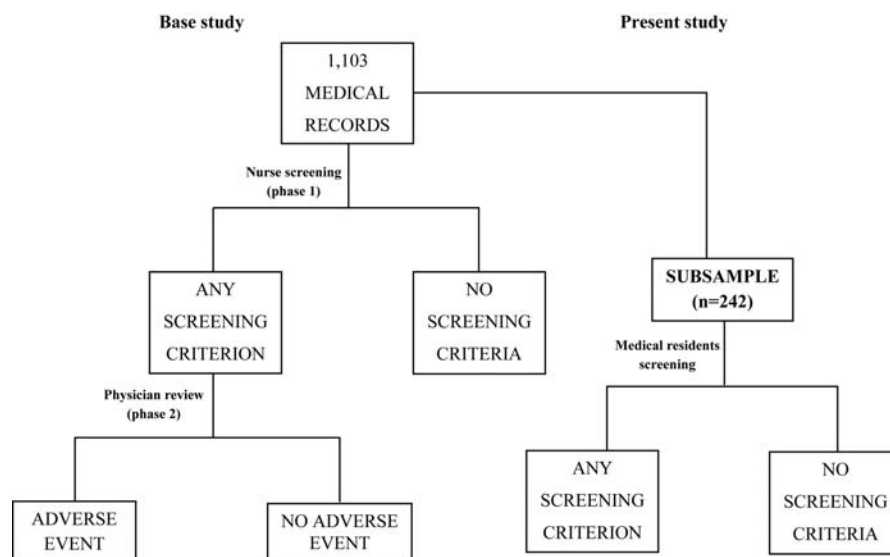

Figure 1 Flowchart of the AE screening method used in previous retrospective cohort study (base study) and in the present study.

criteria', (ii) 'event related to invasive procedure', (iii) 'failure in the administration of supplies' (medication, surgical material, diagnostic tests); (iv) 'missing information in the medical record' ('Missing prescription', 'Missing medical evaluation', 'Another patient's medication chart was included in the medical record').

The frequency of cases meeting the screening criteria in the subsample (screened by medical residents) and in the base study (screened by nurses) was calculated. An agreement between physicians and nurses was measured using the kappa coefficient and simple percentage agreement. Taking the assessment by medical residents as the reference ('gold standard'), the sensitivity, specificity, positive (PPV) and negative (NPV) predictive values of the screening by nurses were calculated.

Sensitivity was estimated by the proportion of records with at least one criterion in nurse screening out of the total with at least one criterion in the review by medical residents. Specificity was estimated by the proportion of records with no screening criteria according to nurses' reviews out of the total of records with no screening criteria according to medical residents' reviews. PPV was estimated by the proportion of cases with at least one screening criterion according to medical residents' screening out of the total of cases with at least one positive criterion according to nurses. NPV was estimated by the proportion of cases with no screening criteria according to the medical residents out of the total of cases with no screening criteria detected by nurses' reviews.

Finally, the PPV of AE screening by medical residents and nurses were also calculated, taking the assessment by experienced physicians in the base study as the reference. At this point, the PPV was estimated by the proportion of cases with confirmed AEs by physicians from the second phase of the base study, out of the total of cases with at least one screening criterion detected by nurses or medical residents.

The proportions and the kappa coefficient were estimated with 95% confidence intervals. McNemar's test was performed to analyse the statistical significance of the agreement between physicians and nurses. Calculations were performed using *Stata* statistical software version 10.0.

Results

Frequency of screening criteria detected by nurses and medical residents

Of the 259 medical records selected, 5 met exclusion criteria (hospital stay under 24 h or inadequate documentation) and 12 medical records showed discordant lengths of stay according to physicians and nurses. Among the 242 medical records in the final sample, nurses identified 95 cases (39.3%) meeting at least one screening criterion, and medical residents identified 90 cases (37.2%). As more than one criterion could be identified in each medical record, nurses identified 179 screening criteria in total, while medical residents identified 171 criteria (Table 1).

The criteria most frequently (>10%) detected by nurses were 'Other unwanted events', 'Prior admission', 'Death', 'Other complications', 'Hospitalization injury' and 'Infection'. These were also the most frequently detected by medical residents, except for 'Hospitalization injury' and 'Other complications'. For these two criteria, nurses detected more cases than the medical residents and differences were statistically significant. Nurses detected 'Hospitalization injury' in 11.2% of cases, while medical residents detected this criterion in 4.7% of cases ($P < 0.001$). The criterion 'Other complications' was detected in 12.3% of cases by nurses and in 7.6% of cases by medical residents ($P = 0.004$). On the other hand, 'Prior admission' and 'Later admission' were identified much more frequently by medical residents than nurses, but the difference was statistically significant only for the former. 'Prior admission' was detected in 18.7% of cases by physicians and in 14.0% of cases by nurses ($P = 0.02$). 'Later admission' was detected in 5.3 and in 2.2% of cases by physicians and nurses, respectively ($P = 0.06$) (Table 1).

At least one screening criterion analysis

A simple agreement between nurses and medical residents regarding cases meeting at least one screening criterion was 78.9% and the chance corrected agreement was moderate across hospitals ($K = 0.55$).

The screening by nurses showed higher specificity (81.6%) than sensitivity (74.4%) for cases meeting at least one screening criterion, taking the assessment of medical residents as the reference (Table 2).

Nurses identified 95 cases meeting the screening criteria, of which 21 (22.1%) were confirmed as AEs by the experienced physicians in the evaluation phase of the base study, whereas medical residents identified 90 cases meeting the screening criteria, of which 18 (20%) were confirmed as AEs in the evaluation phase of the base study. The difference was not statistically significant ($P = 0.08$).

Most frequent screening criteria analysis

Among the criteria most frequently met according to nurses, the agreement was perfect ($K = 1$) for 'Death' and substantial for 'Prior admission' and 'Infection' ($K = 0.66$ and 0.62, respectively). The agreement was reasonable for 'Hospitalization injury' ($K = 0.25$) and low for 'Other complications' and 'Other unwanted events' ($K \leq 0.20$) (Table 3).

All criteria showed higher specificity than sensitivity, except for 'Death' criterion, which showed the highest accuracy, followed by 'Prior admission' (97.6% specificity and 62.5% sensitivity) and 'Infection' (96.4% specificity and 61.9% sensitivity). The criteria 'Other complications' and 'Other unwanted events' had the lowest sensitivity (7.7 and 11.5%, respectively), but showed substantial specificity (90.8 and 89.4%, respectively). The PPV of 'Other complications' criterion was even lower than that of 'Other unwanted events', due to its lower prevalence (Table 3).

Table I The frequency of screening criteria detected by nurses, physicians and both reviewers ($n = 242$)

Screening criteria [19]	Both nurses and physicians, n (%)	Nurses, n (%)	Physicians, n (%)	P-value
1. Unplanned admission (including readmission) as a result of any health care provided during the 12 months prior to the index admission	25 (17.0)	25 (14.0)	32 (18.7)	0.02
2. Unplanned admission to any hospital during the 12 months following discharge from the index admission	4 (2.7)	4 (2.2)	9 (5.3)	0.06
3. Occurrence of injury or harm to patient during hospitalization (including any harm, injury or trauma occurring during index admission)	8 (5.4)	20 (11.2)	8 (4.7)	0.001
4. Adverse drug reaction	6 (4.1)	9 (5.0)	6 (3.5)	0.25
5. Unplanned transfer to intensive or semi-intensive care unit	5 (3.4)	5 (2.8)	5 (2.9)	*
6. Unplanned transfer from or to another acute care hospital (excluding transfers for specialized exams, procedures or care not available in the original hospital)	0 (0.0)	4 (2.2)	0 (0.0)	0.13
7. Unplanned return to surgery room	2 (1.4)	3 (1.7)	2 (1.2)	*
8. Unplanned removal, injury or repair of an organ or structure during surgery, invasive procedure or vaginal delivery	1 (0.7)	1 (0.6)	2 (1.2)	*
9. Other unexpected complications during index admission which are NOT a normal development of the patient's disease or an expected result of the treatment	13 (10.0)	22 (12.3)	13 (7.6)	0.004
10. Development of a neurological alteration absent at admission, but present at the time of discharge from the index admission (includes neurological alterations related to procedures, treatments or investigations)	3 (2.0)	4 (2.2)	3 (1.8)	*
11. Death	24 (16.3)	24 (13.4)	24 (14.0)	*
12. Inappropriate hospital discharge/inadequate discharge plan from index admission (excludes unauthorized discharge)	2 (1.4)	2 (1.1)	5 (2.9)	0.25
13. Reversed cardio-respiratory arrest	4 (2.7)	4 (2.2)	7 (4.1)	0.25
14. Injury related to abortion or labor and delivery	0 (0.0)	1 (0.6)	0 (0.0)	*
15. Hospital infection/septicemia (excludes infections/septicaemia occurring fewer than 72 h after admission)	19 (12.9)	19 (10.6)	21 (12.3)	0.5
16. Dissatisfaction with care received as documented on patient record, or evidence of complaint lodged (includes documents, documented complaint, conflicts between patient/family and health care professionals and unauthorized discharge)	2 (1.4)	2 (1.1)	4 (2.3)	0.5
17. Documentation or correspondence indicating litigation, whether merely intent to sue or actual lawsuit	0 (0.0)	0 (0.0)	1 (0.6)	*
18. Starting with a normal creatinine on admission, creatinine value doubled during the hospital stay	3 (2.0)	4 (2.2)	3 (1.8)	*
19. Any unwanted events not mentioned above	26 (17.7)	26 (14.5)	26 (15.2)	*
Total	147 (100.0)	179 (100.0)	171 (100.0)	—

NS, not significant.

*There was a perfect agreement between reviewers, McNemar's test not applicable.

Among cases discordant for 'Other complications' criterion, the most frequent descriptions were: 'Injury or damage not described in other criteria' (14; 63.6%) and 'Event related to invasive procedure' (6; 27.3%). For criterion 19, the most frequent descriptions were: 'Failure in the administration of supplies' (14; 53.8%), followed by 'Injury or damage not described in other criteria' (9; 34.6%).

Discussion

The present study aimed to validate the screening phase of the base study by analyzing the reliability and accuracy of the screening conducted by nurses. It concluded that nurses were appropriate reviewers for the detection of AEs. The reliability analysis showed a moderate agreement between nurses and

physicians and the accuracy analysis showed specificity higher than sensitivity when taking the assessment of medical residents as the reference.

The result of a moderate agreement between nurses and medical residents for cases meeting at least one screening criterion is consistent with previous studies [15, 16]. Silver *et al.* [15] found an agreement of 78.5% (*K* = 0.57) between physicians and nurses for the presence of at least one AE. An another study found a moderate agreement for medical complications (*K* = 0.59) and reasonable for surgical complications (*K* = 0.36), between physicians and nurses [16].

Specificity was higher than sensitivity; this result is not consistent with previous studies that evaluated screening accuracy, and was possibly caused by methodological differences [17, 18]. Unlike the present study, in the study by Brennan *et al.* [17], the evaluators were not nurses, but professional chart reviewers. In a study by Camacho and Rubin, the assessment tools used by physicians and nurses were not

the same, as the former used a structured form for implicit case assessment [18].

Even though screening should emphasize sensitivity in order to maximize detection, the higher specificity found here favors efficiency. A greater proportion of records meeting a positive criterion that is later confirmed as an AE implies higher cost-effectiveness, especially if confirmation requires costly human resources. Nurses found more records with AEs and their assessment had a greater PPV than that performed by physicians. Moreover, nurses found a higher number of screening criteria than physicians in general and the differences were significant for 'Hospitalization injury' and 'Other complications'.

The screening for quality issues in health care requires simple operational procedures, as the screening relies on the chart review. The screening criteria used in this study have shown to be useful in the detection of AEs. Objective criteria more likely to be identified in patient records, such as criteria 'Prior admission', 'Death' and 'Infection', showed a higher agreement and a greater accuracy [21]. On the other hand, as expected, generic criteria such as 'Other complications' and 'Other unwanted events' showed a lower agreement and accuracy. The remaining criteria showed higher specificity than sensitivity, consistent with the findings of a similar study on the accuracy of screening criteria for AEs [22].

It was found that a large number of cases meeting 'Other complications' and 'Other unwanted events' criteria referred to 'Injury or damage not described in other criteria.' More than half of the cases reported by nurses as meeting 'Other unwanted events' criterion were associated with 'failure in the administration of supplies (medication, surgical material, diagnostic tests)'. It is therefore suggested that other screening criteria should be added such as 'Failure in the administration of supplies', 'Injury or damage not described in other

Table 2 Simple agreement and *Kappa* coefficient between nurses and physicians and sensitivity, specificity, PPV and NPV of AE screening by nurses for cases meeting at least one screening criterion

	%	95% CI
Simple agreement	78.9	65.2–79.7
Kappa	0.55	0.45–0.66
Sensitivity	74.4	65.3–83.6
Specificity	81.6	75.3–87.8
PPV	70.5	58.9–77.9
NPV	84.4	75.3–87.8

Table 3 Simple agreement, *Kappa* coefficient, sensitivity, specificity, PPV and NPV of the most frequently met criteria (>10%) according to nurses

Screening criteria [19]	*n*	Agree (%)	*K*	Sens (%)	Spec (%)	PPV (%)	NPV (%)
1. Unplanned admission (including readmission) as a result of any health care provided during the 12 months prior to the index admission	25	93.0	0.66	62.5	97.6	80.0	94.5
3. Occurrence of injury or harm to patient during hospitalization (including any harm, injury or trauma occurring during index admission)	20	91.7	0.25	50.0	93.2	20.0	98.2
9. Other unexpected complications during index admission which are not a normal development of the patient's disease or an expected result of the treatment	22	86.4	−0.01	7.7	90.8	4.5	94.5
11. Death	24	100.0	1	100.0	100.0	100.0	100.0
15. Hospital infection/septicemia (excludes infections/septicemia occurring fewer than 72 h after admission)	19	94.2	0.62	61.9	96.4	68.4	96.4
19. Any unwanted events not mentioned above	26	81.0	0.01	11.5	89.4	11.5	89.4

Agree, agreement; *K*, kappa; Sens, sensitivity; Spec, specificity; PPV, positive predictive value; NPV, negative predictive value.

criteria' and 'Event related to invasive procedure, in order to improve the quality of the screening process'.

In general, the screening by nurses and medical residents were similar and revealed no significant differences and this result suggests the good performance of the nurses in the screening of AE, as the medical residents was considered the gold standard. Otherwise, the study detected some significant differences in relation to specific screening criteria among reviewers. These results are important to evaluate the implication of different reviewers in the screening process and also show the performance of different screening criteria that compose the screening tool, also giving additional knowledge about the process in order to improve its quality.

One limitation of the accuracy analysis was the choice of medical residents as the reference, which had a lower PPV when assessed by the full implicit assessment. The assumption that the screening by medical residents yields more accurate results, and should be considered the gold standard, can be questioned. However, the implicit assessment performed by physicians, as during the evaluation phase of the base study, is the basis for judging the quality of care; therefore, it seemed appropriate to take it as the reference.

This study explored the potential and limitations of a tool for the detection of AEs, with the advantage of revealing quality problems that would not be noticed otherwise, and showed that nurses had a great performance in the screening for AEs, considering the assessment of medical residents as the reference.

Funding

References

1. Mendes W. Avaliação da ocorrência de eventos adversos em hospital no Brasil [dissertation]. Rio de Janeiro: Fundação Oswaldo Cruz, 2007 (Portuguese).

2. The Conceptual Framework for the International Classification for Patient Safety. Final Technical Report. World Alliance for Patient Safety Taxonomy. Geneva, Switzerland: World Health Organization (WHO), 2009.

3. Murff HJ, Patel VL, Hripcsak G *et al*. Detecting adverse events for patient safety research: a review of current methodologies. *J Am Med Assoc* 2003;**36**:131–43.

4. 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**:370–6.

5. Thomas EJ, Studdert DM, Burstin HR *et al*. Incidence and types of adverse events and negligent care in Utah and Colorado. *Med Care* 2000;**38**:261–71.

6. Wilson RM, Runciman WB, Gibberd RW *et al*. The quality in Australian health care study. *Med J Aust* 1995;**163**:458–71.

7. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ* 2001;**322**:517–9.

8. 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.

9. 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**:1678–86.

10. Zegers M, Bruijne MC, Wagner C *et al*. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. *Qual Saf Health Care* 2009;**18**:297–302.

11. 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**:285–91.

12. 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**:1022–9.

13. Michel P, Quenen JL, Djihoud A *et al*. French national survey of inpatient adverse events prospectively assessed with ward staff. *Qual Saf Health Care* 2007;**16**:369–77.

14. Michel P, Quenon JL, de Sarasqueta AM *et al*. Comparison of three methods for estimating rates of adverse events and rates of preventable adverse events in acute care hospitals. *BMJ* 2004;**328**:199–202.

15. Silver MP, Hougland P, Elder S *et al*. Statewide identification of adverse events using retrospective nurse review: methods and outcomes. *J Nurs Meas* 2007;**15**:220–32.

16. Weingart SN, Davis RB, Palmer RH *et al*. Discrepancies between explicit and implicit review: physician and nurse assessments of complications and quality. *Health Serv Res* 2002;**37**:483–98.

17. Brennan TA, Localio RJ, Laird NL. Reliability and validity of judgements concerning adverse events suffered by hospitalized patients. *Med Care* 1989;**27**:1148–58.

18. Camacho LAB, Rubin HR. Assessment of the validity and reliability of three systems of medical record screening for quality of care assessment. *Med Care* 1998;**36**:748–51.

19. Mendes W, Martins M, Rozenfeld S *et al*. The assessment of adverse events in hospitals in Brazil. *Int J Qual Health Care* 2009;**22**:1–6.

20. Mendes W, Travassos C, Martins M *et al*. Adaptação dos instrumentos de avaliação de eventos adversos para uso em hospitais brasileiros. *Rev Bras Epidemiol* 2008;**11**:55–66.

21. Naessens JM, O'Byrne TJ, Johnson MG *et al*. Measuring hospital adverse events: assessing inter-rater reliability and trigger performance of the Global Trigger Tool. *Int J Qual Health Care* 2010;**22**:266–74.

22. Karson AS, Bates DW. Screening for adverse events. *J Eval Clin Pract* 1999;**5**:23–32.