Consistency of retrospective triage decisions as a standardised instrument for audit

S W Goodacre, M Gillett, R D Harris, K P G Houlihan

Abstract

Objectives—To determine the level of agreement between senior medical staff when asked to perform retrospective case note review of nursing triage decisions, both before and after development of a consensus approach.

Methods—Four medical reviewers independently allocated triage categories to 50 emergency department patients after review of their case notes. They were blind to the identity of the triage nurse and their triage categorisation. The process was repeated twice, firstly after agreement on a consensus approach and then using formal guidelines.

Results—Agreement between reviewers was initially fair to moderate ($\kappa = 0.27$ to 0.53). This failed to improve after development of a consensus approach ($\kappa = 0.29$ to 0.57). There was a trend towards better agreement when guidelines were used but agreement was still only moderate ($\kappa = 0.31$ to 0.63).

Conclusions—Audit of nurse triage categorisation by senior medical staff performing case note review has only fair to moderate consistency between reviewers. Use of this technique will result in frustration among those whose performance is being audited if they recognise inconsistency in the standard they are compared against.

(J Accid Emerg Med 1999;16:322-324)

Keywords: triage; audit; consistency

Appropriate triage is important to the smooth functioning of the emergency department. Failure of effective triage can lead to patient dissatisfaction, staff stress, unnecessary and avoidable morbidity, and even mortality. For these reasons, in 1990 the Australasian College for Emergency Medicine introduced the national triage scale.¹ Patients are categorised by a triage nurse into five categories on the basis of urgency of need (see table 1). This categorisation then determines their subsequent order of assessment by medical staff.

Audit of the triage process is essential if effective triage is to be achieved. Statistics

regarding the proportion of patients assigned to each triage category and the proportion of patients being seen by medical staff within the target time for their category are relatively easy to produce and are routinely collected by emergency departments. However, these data miss a key part of the triage process—the assigning of patients to the correct category by nursing staff. If triage nurses fail to categorise patients appropriately then medical attention may be unacceptably delayed or unnecessarily expedited.

We aimed to assess one potential method of audit—that of retrospective chart review of triage decisions by senior medical staff. This has the advantage of being a relatively quick and inexpensive technique that is available to any emergency department. By reviewing case notes medical staff could retrospectively assign a triage category that would then act as an audit standard against which performance could be judged. Before such a standard is used, it should be demonstrated to be valid and reliable. A valid instrument accurately measures the value of interest, while a reliable instrument produces consistent results on repeated measurement. With regard to audit of triage decisions, a valid standard for audit will accurately describe the “correct” triage decision. A reliable standard will remain consistent when applied by different observers, that is, different medical reviewers presented with the same case will arrive at the same conclusion.

Although validity is clearly important to this process, our study concentrates on establishing the reliability of retrospective review of triage decisions. Previous studies have questioned the ability of physicians to agree on triage decisions.²⁻⁴ We aimed to assess the level of agreement achieved between physicians in our department and whether it can be improved by either development of a common approach or by following a formal set of guidelines.

Methods

Four reviewers, labelled A to D, were selected. All were experienced practitioners in emergency medicine, were familiar with the theoretical and practical use of the national triage scale, and were experienced in chart review and audit. Each was asked to review 50 randomly selected sets of case notes, blind to the assigned triage category and the triage nurse concerned, and assign a triage category to each patient on the basis of a thorough review of the case notes. The reviewers were given the nursing triage notes (with the nurse’s name and triage decision removed), the medical notes up to and including documentation of

Royal North Shore
Hospital, St Leonard’s,
New South Wales,
Australia
S W Goodacre
M Gillett
R D Harris
K P G Houlihan

Correspondence to:
Dr Steve Goodacre, Registrar
in Emergency Medicine, 24
Brookside Crescent, Cuffley,
Hertfordshire EN6 4QN
(e-mail:
steveg10@hotmail.com).

Accepted 20 March 1999

Table 1 The Australasian national triage scale¹

Category	Urgency	Target time (min)
1	Resuscitation	Immediate
2	Emergency	10
3	Urgent	30
4	Semiurgent	60
5	Non-urgent	120

a senior medical opinion, and the results of investigations performed at attendance or shortly after admission.

This procedure was carried out three times. First the reviewers approached the task without any preparation other than their own knowledge and experience of the triage process. The reviewers then met to discuss the cases upon which they had failed to reach a unanimous decision and develop a common theoretical approach to areas of disagreement. After this a second set of 50 randomly selected case notes, prepared in the same way, were reviewed according to the agreed strategy. Finally the second set of notes were immediately reviewed again, this time with each reviewer making their decision in strict accordance with protocols laid out in the book *Emergency Triage* by the Manchester Triage Group.⁵ These protocols cover most presentations to the emergency department and produce triage categorisation by means of a series of discriminatory questions presented in a flow diagram.

Concordance was determined by calculating the κ statistic between pairs of reviewers and the proportion of cases in which the reviewers achieved three out of four and four out of four agreement. The κ statistic is a measure of the degree of agreement between two observers after adjustment to take into account agreement purely due to chance. It is calculated as the actual degree of agreement achieved beyond that due to chance divided by the maximum potential degree of agreement beyond that due to chance. Its range varies from 0 (the degree of agreement achieved is entirely due to chance) to 1 (perfect agreement). By convention, a score of 0.8 to 1.0 indicates excellent agreement, 0.6 to 0.8 is substantial, 0.4 to 0.6 is moderate, 0.2 to 0.4 is fair, and less than 0.2 is slight.⁶

Results

In the first review the four assessors achieved complete agreement in their decisions for 15 out of the 50 cases reviewed. For a further 19 cases, the assessors achieved a majority three out of four agreement. No overall agreement was achieved for 16 cases. In the second review complete agreement was achieved for 12 cases, a majority three out of four agreement was achieved for 24 cases, and no overall agreement was achieved for 14 cases. In the third review complete agreement was achieved for 22 cases, a majority three out of four agreement was achieved for 15 cases, and no overall agreement was achieved for 13 cases.

Table 2 outlines the κ values achieved between the four reviewers for the three reviews performed. By way of comparison table 3 shows κ values for agreement between each of the reviewers and the triage nurses' original categorisation on the two sets of case notes.

Discussion

If nurse triage is to be audited the process by which this is done must be consistent. Frustration will arise if those whose performance is being audited recognise that the standard against which they are being compared is sub-

Table 2 κ Values for agreement between reviewers A to D in the three reviews undertaken

Comparison	1st review	2nd review	3rd review
A with B	0.45	0.29	0.46
A with C	0.27	0.57	0.44
A with D	0.32	0.40	0.31
B with C	0.53	0.32	0.63
B with D	0.40	0.34	0.55
C with D	0.41	0.34	0.62

Table 3 κ Values for agreement between each of the reviewers and the triage nurse

Comparison	1st review	2nd review
A with nurse	0.23	0.43
B with nurse	0.45	0.41
C with nurse	0.37	0.38
D with nurse	0.34	0.60

ject to significant variability. This was the case when our reviewers first approached the task without prior preparation. Levels of agreement between the reviewers were only fair to moderate. Indeed agreement between the reviewers was no better than between each reviewer and the triage nurse decisions they were supposed to be auditing.

Discussion among the reviewers after this initial exercise revealed several areas of inconsistency in their approach. The most significant of these was the degree to which different reviewers used information in the cases notes to make their decision which could not have been available to the triage nurse. The commonest examples of this were the results of blood or radiological investigations. As a result it was agreed that decision making should be based purely on information that could reasonably be elicited at triage. Other inconsistencies regarding the importance of pain assessment in the triage process and perceptions of categorisation for "inappropriate" attendance were also addressed and a common approach agreed.

When the process was repeated after this discussion, however, there was no evidence of any improvement. Agreement between reviewers remained fair to moderate and again was no better than agreement between each reviewer and the triage nurse decisions they were auditing. It appears that the inconsistencies revealed in the discussion were not primarily responsible for the poor level of agreement and that agreement reached in the discussion regarding cases analysed from the first review could not be generalised to the second.

The aim of the final review was to determine whether following a formal protocol could result in improved agreement. This had the disadvantage of removing the element of experience and expertise it was hoped that the reviewers could bring to the audit process, but would replace this with a consistent approach. The results show a trend towards greater consistency but agreement is still only moderate. There are several possible reasons for this disappointing result. The protocols are designed to be used prospectively and do not lend themselves to retrospective review that involves searching the case notes for prompts which would be easy to elicit when the patient is present. There is also a degree of subjectivity in

the protocols, such as in assessing the degree of pain, which may produce greater variability when these symptoms are assessed retrospectively. Although they were instructed to follow the protocols strictly, the reviewers may have been tempted to impose their own judgment when they disagreed with the categorisation given by the protocol. Finally, patients with more than one presenting symptom (for example, abdominal pain and vomiting) may be triaged by more than one pathway, each leading to a different categorisation. This last reason was cited specifically by physician A and it is noticeable in the third review that physicians B, C, and D showed substantial agreement while physician A did not. It may be that this difficulty with the guidelines was a particular problem for physician A.

Previous studies have demonstrated disagreement between physicians as to the degree of urgency with which cases should be managed. Retrospective medical records review carried out by two physicians revealed divergence regarding the designation of visits as "emergencies" when studied by Foldes *et al.*² When Gill *et al* asked two emergency physicians to categorise 266 cases as urgent or non-urgent by chart review, the agreement was moderate ($\kappa = 0.41$).³ Read *et al* asked two physicians to prospectively and retrospectively assign degrees of urgency and produced κ values from 0.41 to 0.62.⁴ Similar levels of agreement were achieved by our reviewers and could not be improved upon by attempting to develop a consensus approach.

Although our reviewers had little difficulty following the Manchester guidelines, it is possible that greater familiarity with them, and even a period of training in their use, could lead to a more consistent approach. It is likely that a more didactic interpretation of the protocols will produce greater consistency. However, even if consistency is achieved this does not ensure validity in the audit of triage. To be valid, the reference standard (in our case, the physician's decision) must accurately reflect an external gold standard. It is one of the major shortcomings of the triage process that measures of outcome by which the performance of triage can be judged have proved difficult to define. In short, formal methods of triage have not yet been demonstrated to convincingly improve outcomes such as mortality or morbidity. Without this, we risk developing an audit process that is consistent but lacks validity.

Senior medical staff were chosen to review triage because they will ultimately be responsi-

ble for the patient's treatment in the emergency department and therefore have a special interest in ensuring effective triage. There is no reason, however, why senior nursing staff should not carry out audit in the same way. Previous studies have also demonstrated disagreement between nurses when retrospectively assessing urgency³ and between nurses and physicians in the same process.^{3 7-9} Our study could be repeated with four senior nurses as reviewers to identify if the outcome is similar.

It may be argued that the retrospective nature of our audit can lead to inaccuracy and that contemporaneous assessment would be more effective. Indeed video assessment of triage has been suggested as a useful audit tool.¹⁰ However, there is no indication in the literature that prospective assessment of triage improves agreement and it may even make agreement worse.⁸ The consistency of video audit of triage should be demonstrated before it is recommended as a routine method of audit.

Conclusion

Audit of nurse triage categorisation by senior medical staff performing retrospective chart review has only fair to moderate consistency between reviewers. Use of this technique will result in frustration among those whose performance is being audited if they recognise inconsistency in the standard they are compared against.

We thank Jane McMullen and Alison Simpson for their help with this project.

Conflict of interest: none.

Funding: none.

- 1 Australasian College for Emergency Medicine Policy Document. National triage scale. *Emergency Medicine* 1994;6:145-6.
- 2 Foldes SS, Fischer LR, Kaminsky K. What is an emergency? The judgements of two physicians. *Ann Emerg Med* 1994;23:833-40.
- 3 Gill J M, Reese CL, Diamond JJ. Disagreement among health care professionals about the urgent care needs of emergency department patients. *Ann Emerg Med* 1997;30:474-9.
- 4 Read S, George S, Westlake L. Piloting an evaluation of triage. *Int J Nurs Stud* 1992;29:275-88.
- 5 Manchester Triage Group. *Emergency triage*. London: BMJ Publishing Group, 1997.
- 6 Sackett DL, Haynes RB, Guyatt GH, et al. *Clinical epidemiology: a basic science for medicine*. 2nd Ed. Boston: Little Brown, 1991.
- 7 Brillman JC, Doezeema D, Tandberg D, et al. Triage: limitations in predicting need for emergent care and hospital admission. *Ann Emerg Med* 1996;27:493-500.
- 8 Brillman JC, Doezeema D, Tandberg D, et al. Does a physician visual assessment change triage? *Am J Emerg Med* 1997;15:29-33.
- 9 George S, Read S, Westlake L, et al. Differences in priorities assigned to patients by triage nurses and by consultant physicians in accident and emergency departments. *J Epidemiol Community Health* 1993;47:312-5.
- 10 Williams JC, Jones NL, Richardson FJ, et al. The nursing triage process: a video review and proposed audit tool. *J Accid Emerg Med* 1996;13:398-9.