

Close only counts in horseshoes and . . . triage?

Eric Grafstein, MD

SEE ALSO PAGE 240.

The ability to reliably triage patients has become front and centre in the collective consciousness of physicians and nurses who want to ensure that the sickest patients get timely care, among administrators who want to measure emergency department (ED) performance, among researchers who want to describe ED populations, and among politicians who would like to base physician remuneration models on triage levels. But all of these potential uses mandate that our triage systems and processes are reliable — in other words, that similar patients receive similar triage levels when presenting to different hospitals in different regions.

The article by Worster and colleagues¹ in this issue of the Journal (see page 240) compares the ability of nurses to learn and apply Version 3 of the Emergency Severity Index (ESI[®] v.3) and the *Canadian Emergency Department Triage and Acuity Scale* (CTAS). Most triage scales, including CTAS, assign triage levels based on how quickly patients require care. Like CTAS, the ESI uses an “urgency” assessment to define the sickest patients (level 1 and 2); unlike CTAS, it uses the perceived need for diagnostic testing to differentiate patients in levels 3, 4 and 5. The potential benefit is that ESI has a simple algorithm based on expected resource utilization to help triage nurses categorize lower acuity patients. Readers might therefore infer that the ESI represents an improvement over CTAS, but this is not the case.

In this study, the authors used written triage scenarios and quadratically weighted kappa values to assess triage reliability. They concluded that agreement was high, with weighted kappa values of 0.91 and 0.89 for the ESI and CTAS groups respectively. Unfortunately, they did not report the raw agreement between triage nurses, and this lim-

its our ability to compare the 2 tools. The notion of using a single score to describe agreement between 5 nurses (i.e., 5 nurses in each group) and 200 scenarios is somewhat simplistic. It does not, for example, tell us whether there was good agreement for level 5 patients and poor agreement in levels 3 and 4. Moreover, when interpreting reliability statistics, it is important we understand the difference between the unweighted kappa statistic and the weighted kappa statistic, both measures of agreement. In most situations, agreement is agreement; a disease or characteristic is either present or absent. But when there are many options to choose from, it sometimes makes sense to use a weighted kappa statistic, which provides partial credit for being close.² For example, if we studied emergency physician reliability in documenting ICD-10 codes,³ it quickly becomes apparent that there are more than 10 000 possible codes, so the likelihood of exact agreement is low. In this situation, it makes sense to use a weighed kappa statistic and give partial credit if 2 physicians apply codes that match the same system (e.g., cardiovascular). But the situation is different for triage scores, where there are only 5 possible triage levels and where 99% of patients fall in the lowest 4 levels. If 1 triage nurse codes a patient or scenario as level 3 (the most common level), quadratically weighted kappa values will give the second nurse credit if he or she scores the patient as anything from 2 to 4. This means that weighted kappa scores tend to overestimate the level of agreement between observers.² Given the importance of accurately describing ED patient populations, quadratically weighted kappa scores are not sufficient. At the very least, an assessment of triage reliability requires authors to report an unweighted kappa value and the raw agreement on exact triage level. Preferably, they should also report un-

Research Director, Informatics, Department of Emergency Medicine, Providence Health Care, and St. Paul's Hospital, Vancouver, BC

Received: May 14, 2004; final submission: May 17, 2004; accepted: May 17, 2004

Can J Emerg Med 2004;6(4):288-9

weighted kappa values between adjacent triage levels (e.g., between levels 3 and 4 or between levels 2 and 3).

Another key limitation of the ESI method is that basing triage assignment on expected resource utilization introduces a process of circular logic. If the triage nurse believes that diagnostic and therapeutic resources will be required and assigns a treatment location based on this, there is much greater likelihood those resources will indeed be utilized. To illustrate, patients with abdominal pain who are triaged to “acute” stretchers tend to undergo more investigation than those triaged to “fast track,” based on triage location alone. This self-fulfilling prophecy will artificially enhance apparent triage reliability.

The ESI attempts to mix acuity with resource utilization, but there is a fundamental problem in using 1 tool for distinctly different purposes. Triage scales were meant to determine how quickly patients need care. If they also predict resource utilization and complexity, this is a bonus but, often they do not, because acuity is not the same as complexity. Patients with anaphylactic reactions, body fluid exposures and caustic splashes to the eye require rapid treatment but minimal resource utilization; they are of high acuity and low complexity. Conversely, a weak and dizzy nursing home patient requires higher resource utilization (high complexity) but is of low acuity. If we confuse the purposes of a triage system and try to develop a tool that stratifies treatment urgency, predicts resource utilization and defines physician remuneration categories, we may end up with a tool that does nothing well. The Swiss Army knife is a dandy knick-knack, but no one uses it to do meaningful work. A system that mixes acuity with resource utilization does not, on the surface, offer any less potential confusion than the current CTAS system that is widely used across Canada.

Worster and colleagues refer to “the 2” previously published studies, based on written summaries, that measured CTAS reliability. Although the authors suggest that these sorts of trials can only occur using paper scenarios, previously published work has, in fact, assessed the real-time reliability of on-duty nurses using CTAS to triage actual patients arriving in the ED.⁴ This prior study reported good unweighted kappa values for exact agreement on triage levels (raw agreement = 0.75, kappa = 0.66). Moreover, using live patients more accurately reflects the real triage environment and helps to overcome potential prevalence biases associated with paper scenarios.⁵

The Canadian Association of Emergency Physicians’ CTAS working group is now developing an algorithm to help increase triage reliability by linking CTAS levels to a nationally adopted Presenting Complaint List.⁶ This im-

provement, along with explicit triage level modifiers (e.g., vital signs and pain scores), will improve overall CTAS reliability and facilitate the creation of ED case-mix groupings based on acuity and presenting complaint. Of interest, Canadian and other international studies suggest that, like ESI, the CTAS also correlates well with resource utilization.⁷⁻¹⁰

Given the similar performance of the 2 scales for paper-based scenarios, the lack of ESI information from real patient encounters, and the absence of key (ESI) data such as actual triage level agreement, the advantages of this system over CTAS remain unclear and do not suggest the need for a change in Canadian triage systems.

Competing interests: None declared.

References

1. Worster A, Gilboy N, Fernandes CM, Eitel D, Eva K, Geisler R, Tanabe P. Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *Can J Emerg Med* 2004;6(4):240-5.
2. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol* 1993;46(5):423-9.
3. International Statistical Classification of Diseases and Related Health Problems, 10th rev. Geneva: World Health Organization; 1992.
4. Grafstein E, Innes G, Westman J, Christenson J, Thorne A. Inter-rater reliability of a computerized presenting-complaint-linked triage system in an urban emergency department. *Can J Emerg Med* 2003;5(5):323-9.
5. Beveridge R, Ducharme J, Janes L, Beaulieu S, Walter S. Reliability of the Canadian Emergency Department Triage and Acuity Scale: interrater agreement. *Ann Emerg Med* 1999;34:155-9.
6. Grafstein E, Unger B, Bullard M, Innes G; for the Canadian Emergency Department Information System (CEDIS) Working Group. Canadian Emergency Department Information System (CEDIS) Presenting Complaint List (Version 1.0). *Can J Emerg Med* 2003;5(1):27-34.
7. Jiménez JG, Murray MJ, Beveridge R, Pons JP, Cortés EA, Ferrando Garrigós JB, et al. Implementation of the *Canadian Emergency Department Triage and Acuity Scale* (CTAS) in the Principality of Andorra: Can triage parameters serve as emergency department quality indicators? *Can J Emerg Med* 2003;5(5):315-22.
8. Yoon P, Steiner I, Reinhardt G. Analysis of factors influencing length of stay in the emergency department. *Can J Emerg Med* 2003;5(3):155-61.
9. Stenstrom R, Grafstein E, Innes G, Christenson J. Real-time predictive validity of the Canadian Triage and Acuity Scale (CTAS) [abstract]. *Acad Emerg Med* 2003;5:512.
10. Murray MJ, Levis G. Does triage level (Canadian Triage and Acuity Scale) correlate with resource utilization for emergency department visits? [abstract]. *Can J Emerg Med* 2004;6(3):180.

Correspondence to: egrafstein@providencehealth.bc.ca