Inter-rater reliability of a computerized presenting-complaint–linked triage system in an urban emergency department

Eric Grafstein, MD;[*†] Grant Innes, MD;[*†] Julie Westman, RN;[*] James Christenson, MD;[*†] Anona Thorne, MSc[†]

ABSTRACT

Background: Triage reliability studies typically use hypothetical scenarios and weighted kappa scores where agreement within one level is considered satisfactory. But if triage category is used to help define ED case-mix groups for comparative or benchmarking processes, agreement on exact triage level and major system involved is important. Our hypothesis was that a computerized menu that links presenting complaints to preferred triage levels (PC-linked triage) would provide high triage reliability.

Objectives: Our objective was to assess inter-rater reliability of PC-linked triage using the *Canadian Emergency Department Triage and Acuity Scale* (CTAS) in a real-time clinical setting, considering agreement on exact triage level and primary body system involved.

Methods: On duty triage nurses entered patient presenting complaint and PC-linked triage level as per standard procedure. In a convenience sample of patients, a second nurse, blinded to triage assignment, observed the triage interaction and independently entered presenting complaint and triage level on a dummy terminal.

Results: During the study, 15 nurse pairs triaged 266 patients. Study patients matched actual emergency department case mix closely. Triage nurses agreed exactly in 74% of cases and within one level in 94% of cases. The unweighted kappa value was 0.66 (95% confidence interval [CI], 0.60–0.73) and the quadratic weighted kappa value was 0.75 (95% CI, 0.68–0.81). Kappa for agreement on major system involved was 0.80 (95% CI, 0.69–0.91).

Conclusion: PC-linked triage has high inter-rater reliability in a real-time clinical setting. PC-linked triage may be useful as one factor in defining case-mix groups for benchmarking and comparative purposes.

Key words: *Canadian Emergency Department Triage and Acuity Scale*; CTAS; presenting complaints; reliability

Introduction

Emergency department (ED) triage scales are used to prioritize patient care, enhance department efficiency, define ED case-mix groups and predict hospital resource utilization.[1–3] The *Canadian Emergency Department Triage and Acuity Scale* (CTAS)[1] is a 5-level scale (Resuscitation, Emergent, Urgent, Less Urgent and Non Urgent) based on Australia's National Triage Scale,[4] now called the Australasian Triage Scale (ATS). CTAS was implemented in

*Department of Emergency Medicine, St. Paul's Hospital, Vancouver, BC
†Centre for Health Evaluation and Outcome Sciences, St. Paul's Hospital, Vancouver, BC

Received: July 8, 2003; final submission: Aug. 1, 2003; accepted: Aug. 8, 2003

This article has been peer reviewed.

Can J Emerg Med 2003;5(5):323-9

RÉSUMÉ

Contexte : Les études de fiabilité du triage utilisent d'habitude des scénarios hypothétiques et des scores kappa pondérés où la concordance au sein d'un même niveau est jugée satisfaisante. Mais si la catégorie de triage est utilisée pour aider à définir le profil de clientèle au DU à des fins de comparaison ou de référence, la concordance entre le niveau de triage exact et le système d'organes touché est importante. Notre hypothèse était qu'un menu informatisé reliant les raisons de consultation aux niveaux de triage de prédilection (triage par ordinateur) garantirait une fiabilité de triage élevée.

Objectifs : Notre objectif était de déterminer la fiabilité inter-évaluateurs du triage par ordinateur utilisant *L'échelle canadienne de triage et de gravité pour les départements d'urgence* (ÉTG) dans un cadre clinique en temps réel, en tenant compte de la concordance entre le niveau de triage exact et le système d'organes principalement touché.

Méthodes : Des infirmières de triage en poste entrèrent la raison de consultation du patient et déterminèrent le niveau de triage par ordinateur selon les consignes habituelles. Une seconde infirmière, travaillant à l'insu quant à l'assignation de triage, observa l'interaction de triage parmi un échantillon de convenance de patients et entra séparément la raison de consultation et le niveau de triage sur un terminal factice.

Résultats : Au cours de l'étude, 15 paires d'infirmières procédèrent au triage de 266 patients. Le profil des patients de l'étude s'apparentait étroitement au profil de clientèle réel du département d'urgence. Les décisions des infirmières concordaient exactement dans 74 % des cas et dans 94 % à l'intérieur d'un même niveau de triage. La valeur kappa non pondérée était de 0,66 (IC 95 %, 0,60–0,73) et la valeur kappa quadratique pondérée était de 0,75 (IC 95 %, 0,68–0,81). La valeur kappa pour la concordance quant au système d'organes principalement touché était de 0,80 (IC 95 %, 0,69–0,91).

Conclusion : Le triage par ordinateur produit une fiabilité inter-évaluateurs élevée dans un cadre clinique en temps réel. Le triage par ordinateur pourrait se révéler utile au nombre des facteurs de détermination des groupes de cas à des fins de référence ou de comparaison.

the mid-1990s and has become a national standard. It is now mandated in all Ontario hospitals and is being widely implemented in many Canadian EDs.

Previous studies[5–13] suggest that reliable triage scales exist, but these studies have used hypothetical written triage scenarios rather than actual patients, have tested fundamentally different (i.e., 3-level) scales, or have reported weighted kappa scores that provide partial credit for near agreement on triage level, a methodology that may inflate triage reliability estimates. If triage is used to define ED case-mix groups for comparative, benchmarking or remunerative purposes, a high level of agreement on exact triage level and major body system involved is important, and researchers must demonstrate that this agreement is possible in actual ED patients.

Triage nurses assign patients to acuity levels based on presenting problem, perceived illness severity, vital signs, pain severity and intuition;[1] consequently, triage can be a subjective process that varies with experience and setting. Our hypothesis was that the use of an explicit presenting complaint list, in which presenting complaints link to specific triage levels (PC-linked triage), would provide high inter-rater reliability by "forcing" triage nurses toward the preferred triage levels for any given presenting complaint. Our primary objective was to assess inter-rater reliability between triage nurses using PC-linked triage in a real-time clinical setting where only exact agreement on triage level is considered correct (unweighted kappa). Our secondary objective was to determine inter-rater agreement for the major presenting complaint category.

Methods

Setting

This prospective observational study took place at St. Paul's Hospital in Vancouver, BC, an urban academic centre with 47 000 annual ED visits. It enrolled a convenience sample of patients who were triaged in the ED between August 2000 and December 2002.

Intervention (PC-linked triage)

As a pilot site participating in the development of a common national International Classification of Diseases (ICD)-10–based presenting complaint list to characterize patients who present to EDs,[3] we developed an electronic system that links patient presenting complaints to preferred triage acuity levels. This system uses explicit CEDIS (Canadian Emergency Department Information System) presenting complaints linked to triage acuity levels based on descriptors published in the CTAS Implementation Guidelines.[1] In cases

https://doi.org/10.1017/S1481803500015839 Published online by Cambridge University Press

where the CTAS guidelines did not specify an acuity level for a presenting complaint (e.g., epistaxis), an expert internal panel of 2 triage nurses, the clinical nurse triage instructor and 2 emergency physicians assigned the most appropriate level based on existing ED practice.

At the time the study began, PC-linked triage had been a standard department process for approximately 2 years, and all ED nurses assigned to perform triage were trained in its use. These nurses had also all undertaken standard CTAS triage training provided by an approved instructor. When patients arrived and reported the main reason for their visit, the on-duty triage nurse selected the most appropriate presenting complaint from an explicit presenting complaint list[3] (available from the authors on request). The presenting complaint and its electronically-linked triage level were entered into the hospital's admission, discharge and transfer (ADT) system and subsequently downloaded into the ED's administrative and research database. Figure 1 illustrates several presenting complaints and their linked triage levels.

Interobserver assessments

During shifts when ED coverage was adequate and an additional "float" nurse trained in PC-linked triage was available (typically between 1200 and 1800), this nurse was asked to perform interobserver triage assessments for the reliability study. The interobserver nurse sat at a "dummy" terminal adjacent to the primary triage nurse, listened to the triage interaction, and coded the patient's PC-linked triage level exactly as she would if she was the primary triage nurse. The interobserver nurse was not permitted to directly question the patient, and both nurses were blinded to the other's triage assessment. The interobserver triage assessment was printed as a paper copy and later keyed to an Excel spreadsheet, where it was linked to the primary triage assessment from the ED administrative database.

Presenting Complaint	Triage code	CTAS acuity level
Unspecified abdominal pain	GI003	III
Localized swelling/redness	SK033	IV
Lower extremity injury	OC063	IV
Chest pain (non-specific)	CV023	III
Chest pain, possible cardiac	CV022	II
Medication request	MC041	V
Traumatic arrest	TR001	I

Fig. 1. Sample presenting complaints and PC-linked triage levels. PC-linked = a computerized menu that links presenting complaints to preferred triage levels (PC-linked triage)

One of the investigators observed the paired triage assessments to assure blinding and protocol compliance. Nurse initials were recorded and no nurse pairs triaged together for more than one triage session.

Data management and statistics

Paired triage assessments from all study subjects were collated in the Excel database and analyzed using S-PLUS (Insightful Corp., Seattle). Raw agreement, as well as a prevalence and bias kappa statistic (PABAK), were calculated for the major complaint category and the linked triage acuity level. To enable comparison to previous studies, linear and quadratic-weighted kappa values were also determined. This study was reviewed and met the ethical standards of the University of British Columbia/Providence Health Care Research Ethics Board.

Results

During the study period, 15 different nurse pairs co-triaged 266 distinct patients. Each pair evaluated an average of 17 patients (range, 2 to 25). Table 1 shows the proportion of patients in CTAS Levels I to V, demonstrating that the study sample was representative of the department's annualized case mix, and Figure 2 illustrates the major complaint categories represented in the study sample. This grouping of complaints into major/main systems predates the creation of the CEDIS presenting complaint list,[3] and there are only 13 instead of the 18 CEDIS major presenting complaint categories. Raw interobserver agreement for main system involved (e.g., cardiovascular, gastrointestinal, respiratory) was 0.80 (95% confidence interval [CI], 0.69–0.91) and, because of the large number of possible presenting complaint categories, this approximates the unweighted kappa value for agreement on presenting complaint category.

Figure 3 summarizes the reliability statistics assessed in this study, enabling comparison with previous studies. Raw interobserver agreement for exact triage level was 0.74 (95% CI, 0.68–0.80). The unweighted PABAK was 0.66

Table 1. Patient acuity distribution in study sample v. actual emergency department (ED) population

CTAS triage level	ED census data (2002), % *n* = 46 505	Study sample, % *n* = 266
I	0.75	1.1
II	11.7	12.9
III	32.7	35.9
IV	36.7	32.8
V	18.1	18.0

https://doi.org/10.1017/S1481803500015839 Published online by Cambridge University Press

(95% CI, 0.60–0.73), and the quadratic-weighted kappa value (most commonly used in other triage studies) was 0.75 (95% CI, 0.68–0.81), indicating good agreement.[14] Figure 4 shows that triage nurses agreed on exact triage level in 196 cases (73.7%) and were within one triage level in 249 cases (93.6%). Of the 53 patients where there was one-level triage discordance (20%), the disagreement was between Levels II and III in 14 cases (5%), between Levels III and IV in 23 cases (9%), and between Levels IV and V in 16 patients (6%).

Discussion

This prospective study of real-time patient encounters shows that triage nurses using a presenting complaint linked system, where triage menus explicitly link complaints to preferred acuity levels, can achieve high levels of agreement for triage category and exact CTAS acuity level. These data suggest that PC-linked triage is a reliable tool, and that the combination of presenting complaint and acuity level may be useful as one factor to define ED case-mix groups for comparative and benchmarking purposes.

Previous studies

Previous triage reliability studies have usually been based on written case scenarios with mock patients. These studies typically describe inter-rater reliability using quadratically weighted kappa values, which do not require exact agreement, but which give "partial credit" for near agreement within 1 or 2 triage levels. In 1998, Beveridge and colleagues[11] assessed CTAS reliability by having 10 nurses

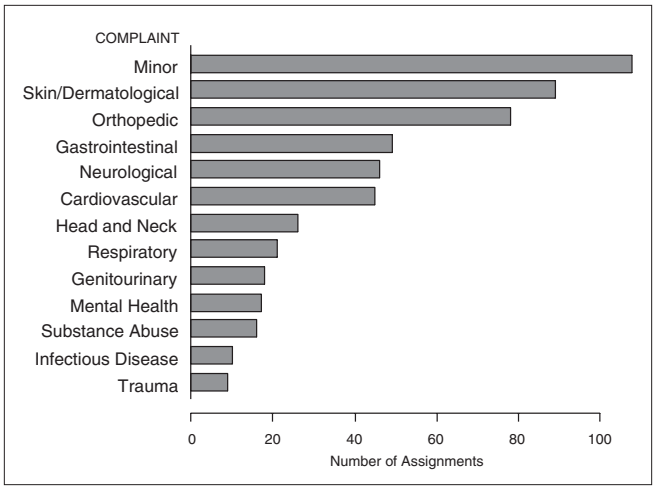

Fig. 2. Distribution of major complaint categories for 532 assignments

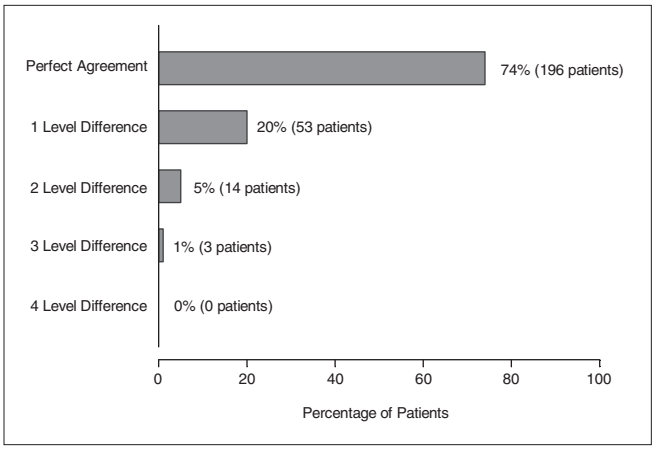

Fig. 4. CTAS level — Overall results (15 pairs of nurses evaluating a total of 266 patients)

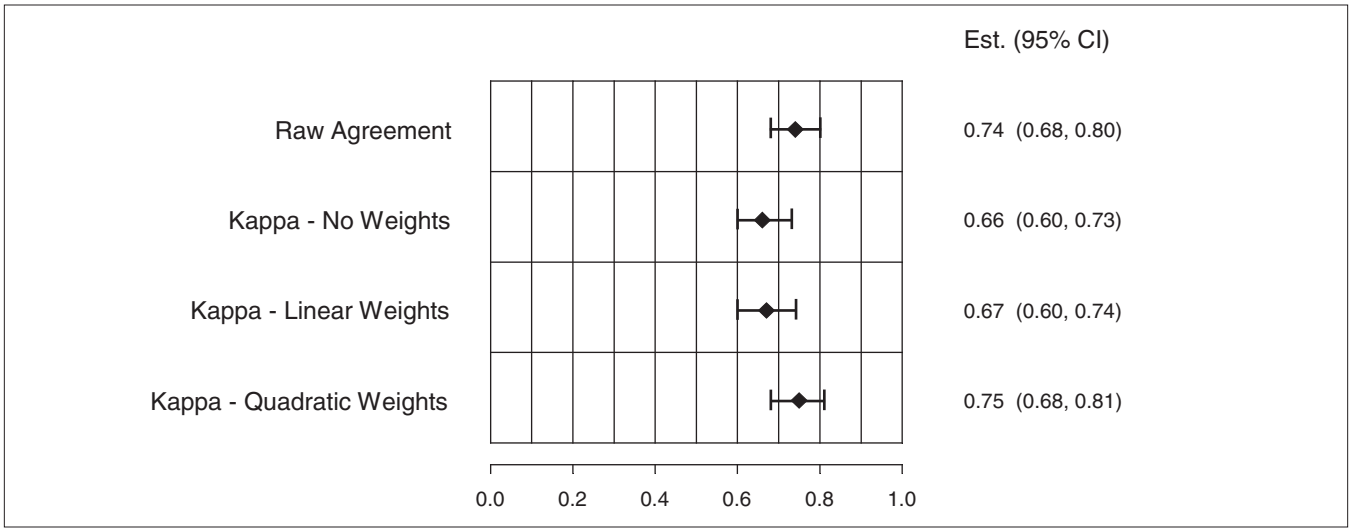

Fig. 3. CTAS triage reliability statistics with confidence intervals

https://doi.org/10.1017/S1481803500015839 Published online by Cambridge University Press

and 10 physicians review written case scenarios and "triage" the patients described. In this study, the nurses and physicians achieved a raw agreement of 54% with a weighted kappa value of 0.80 (95% CI, 0.79–0.81). In a recent similar study, Manos and coworkers[8] studied triage agreement by having 5 physicians, 5 nurses, 5 Basic Life Support paramedics and 5 Advanced Life Support paramedics review 42 written case scenarios abstracted from actual ED encounters. As in the Beveridge study, each mock patient was assigned a triage level from I to V based on the CTAS. In the Manos study, exact agreement on triage level was 63.4%, agreement within one level was 94.9%, and the quadratic-weighted kappa value was 0.77 (95% CI, 0.76–0.78).[8]

Other studies have been reported that involve "real" patients. In the ESI (Emergency Severity Index) reliability study,[10] one of the investigators completed a triage score retrospectively based on the triage nurses' initial assessment. The investigator score versus the triage nurses' scores formed the agreement matrix. In this study raw agreement was 77% but only the weighted kappa of 0.80 (95% CI 0.76–0.84) is reported. Another study comparing CTAS in the prehospital setting using the triage nurse as the gold standard compared to paramedics in the field had a raw agreement of about 60%.[5]

In the current study, exact agreement on triage level was 73.7%, agreement within one level was 93.6%, and the quadratic-weighted kappa value was 0.75 (95% CI, 0.68–0.81). However, there are important differences. Our study was conducted on actual ED patients and reflected a "real" CTAS acuity distribution, while prior studies used paper triage scenarios that distributed case acuity more evenly across all CTAS levels. This is a critical distinction, because the prevalence and bias related to paper triage scenario studies can make it difficult to generalize their findings to actual clinical practice. The reason for this is that two observers are more likely to agree exactly when patients are in Levels I or V, and less likely to agree exactly when patients are in Levels III or IV;[11] therefore, it may be more difficult to demonstrate agreement in "real" acuity distributions, where most of the patients are in these intermediate levels. When assessing triage reliability data, it is important to consider the acuity spectrum, the type of kappa statistic reported, and the raw agreement, since quadratic kappa values reported in isolation may be misleading.[15]

Improving triage reliability

When presenting complaint and triage level are not explicitly linked, nurses have more freedom to subjectively place patients in the triage level of their preference. This may be more or less valid, depending on the experience of the triage nurse and the criteria used to assign triage level, but it is a threat to reliability. A system that links presenting complaints to a limited range of triage levels offers enhanced reliability while preserving some nurse judgement. To illustrate, in the PC-linked system described, the default level for abdominal pain is Level III, but triage nurses may upgrade this to Level II (abdominal pain with abnormal vital signs) or Level I (abdominal pain with shock); alternatively, they may downgrade it to Level IV (chronic abdominal pain). Many other presenting complaints, such as traumatic arrest or medication request, map to only one triage level (Levels I and V respectively). We believe that PC-linking improves reliability by "forcing" nurses towards preferred triage levels for any given presenting complaint. We are now modifying the system to allow nurses to override the preferred level in appropriate circumstances (e.g., severe pain or frequent visits with the same complaint).

PC-linked triage is a promising means of improving triage reliability, but only if EDs use common or very similar presenting complaint systems. Several presenting complaint lists exist, but some are relevant to family practice or clinic settings, some mix presenting complaints and diagnoses (e.g., "asthma"), and one that was designed for the ED is too general to be used to define case-mix groups.[16] In the hopes of standardizing the way that Canadian EDs characterize and stratify their patients, the CEDIS Working Group has published a set of 158 presenting complaints based on ICD-10 codes.[3] These complaints were chosen based on their frequency in 3 Canadian EDs, modified by pediatric emergency specialists, and endorsed by the Canadian Association of Emergency Physicians and the National Emergency Nurses Affiliation. There is approximately 95% overlap between the presenting complaint system used in this study and the recently proposed CEDIS system.[3] Our data therefore suggest that PC-linked triage using the CEDIS presenting complaint system will enhance triage reliability and facilitate the identification of ED case-mix groups.

Defining emergency department case-mix groups

To facilitate benchmarking and comparison of EDs in a community, region or country, it is important to develop ED case-mix groupings (CMGs) and reliable ways to define them. Hospital-based CMGs are typically related to final diagnosis, but this may be problematic for EDs. The diagnosis may be uncertain at the end of the ED visit, and 80% to 90% of patients are discharged from the ED and

https://doi.org/10.1017/S1481803500015839 Published online by Cambridge University Press

therefore do not undergo the same level of investigation and diagnostic confirmation as hospitalized patients. In addition, the development of ED short-stay units and decreasing access to inpatient hospital beds has changed admitting practices, which may vary greatly between departments, making inpatient diagnosis-related case-mix groups potentially quite different in different hospitals.

Using the National Triage Scale (NTS), Cleary and colleagues proposed basing case-mix groups on triage acuity, discharge diagnosis and disposition.[17] This is reasonable for conditions that have reliable diagnostic criteria and are treated in relatively constant fashion (e.g., acute myocardial infarction [AMI], fractured femoral neck); however, for syndromes that lack uniform diagnostic criteria and those that are managed differently in different settings (e.g., abdominal pain, pneumonia, minor trauma), presenting complaint may be a better parameter on which to base CMGs because it captures all patients — not just those admitted — and it eliminates potential biases related to divergent diagnostic and disposition practices. To illustrate, it may be important to evaluate the health outcomes and utilization patterns for all patients with Level II chest pain — not just the small proportion who had AMI. In addition to providing a more relevant CMG denominator, presenting complaint may be more reliable than ED diagnosis for the many patients with ill-defined conditions, therefore more suited to the assessment of emergency health services delivery.

Threats to reliability

Triage training and experience differ between settings, but even if these key determinants are similar, nurses may be motivated to "game" the system and up- or down-triage patients. In overcrowded departments and those lacking on-site physicians, nurses may "down-triage" because they cannot provide the rapid access to stretchers and care that is specified in higher triage levels. In other situations, nurses may down-triage common acute problems. For example, in a department that treats a large volume of heroin overdoses, nurses may view an apneic patient as "just another overdose" rather than a Level I emergency. Likewise, nurses may up-triage low-acuity patients to avoid placing them in categories where they face longer waiting times. In the future, if payers link triage level to remuneration, triage nurses may experience pressure to up-triage.

Limitations

We enrolled a convenience sample of patients, and it is possible this introduced a sampling bias. The nurses who provided interobserver triage assessments did so during normal working shifts, therefore were more likely to be available when the ED was quiet and less likely to participate during busy periods. This also might have influenced our results. Interobserver nurses only observed the primary triage interaction and were not permitted to directly question the patients. This may have limited their triage assessment and altered the level of agreement seen. Nevertheless, the ability to observe and listen to a real patient is clearly much closer to reality than is reading about that patient in a written scenario. Many patients in this study presented with more than one complaint. In these cases, the primary complaint recorded by the triage nurse often matched the secondary complaint recorded by the interobserver nurse (and vice versa), but we compared only the first presenting complaint recorded by each nurse; therefore this type of agreement was lost in our analysis.

Although we showed that 2 nurses can apply the PC-linked triage tool reliably, we did not show that nurses in 2 different hospitals can do so, and inter-site reliability is a significant concern, particularly if triage acuity is used for comparative or remunerative processes. Future research should address this important topic.

With respect to presenting complaint and CMG, we measured only the agreement on major systems, not on specific presenting complaints. There are 158 presenting complaints in the CEDIS list, and to assess agreement for all of these would require a huge sample size. Future research should focus on assessing interobserver agreement for key CMGs such as Level II chest pain, Level III abdominal pain, and Level IV extremity injury.

Conclusions

Presenting-complaint–linked triage can increase triage reliability and is a promising method of defining reliable ED case-mix groups. EDs that develop PC-linked triage systems are encouraged to adopt a common ED presenting complaint list to maximize comparability with other sites.

Competing interests: None declared.

References

1. Beveridge R, Clarke B, Janes L, Savage N, Thompson J, Dodd G, et al. Canadian emergency department triage and acuity scale: implementation guidelines. Can J Emerg Med 1999;1 (suppl 3):S1-24. Available: www.caep.ca/002.policies/002-02.ctas.htm (accessed 2003 Aug 13).
2. Stenstrom R, Grafstein E, Innes G, Christenson J. The predictive validity of the Canadian triage and acuity scale (CTAS) [abstract]. Can J Emerg Med 2003;5(3):184.
3. 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.
4. Australasian College of Emergency Medicine. A National Triage Scale [policy]. Emerg Med 1994;6:145-6.
5. Murray MJ, Bondy S. The reliability of the Canadian emergency department triage and acuity scale in the prehospital setting: interrater agreement between paramedics and nurses [abstract]. Can J Emerg Med 2002;4(2):128.
6. Jelinek G, Little M. Interrater reliability of the National Triage Scale. Emerg Med 1996;8:226-30.
7. Grafstein E, Innes G, Christenson J, Clarke L. Canadian Emergency Department Triage and Acuity Scale triage: reliability for high acuity patients [abstract]. Can J Emerg Med 2001;3(2):143.
8. Manos D, Petrie DA, Beveridge RC, Walter S, Ducharme J. Interobserver agreement using the Canadian Emergency Department Triage and Acuity Scale. Can J Emerg Med 2002;4(1):16-22.
9. Fernandes C, Wuerz R, Clark S, Djurdjev O. How reliable is emergency department triage? Ann Emerg Med 1999;34(2):141-59.
10. Wuerz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and validity of a new five-level triage instrument. Acad Emerg Med 2000;7(3):236-42.
11. 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.
12. Doherty S. Application of the National Triage Scale is not uniform. Aust Emerg Nurs J 1996;1(1):26.
13. Dilley S, Standen P. Victorian triage nurses demonstrate concordance in the application of the National Triage Scale. Emerg Med 1998;10:12-8.
14. Landis J, Koch G. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.
15. Byrt T, Bishop J, Carlin J. Bias. Prevalence and Kappa. J Clin Epidemiol 1993;46(5)423-9.
16. Aronsky D, Kendall D, Merkley K, James BC, Haug PJ. A comprehensive set of coded chief complaints for the emergency department. Acad Emerg Med 2001; 8(10): 980-9.
17. Cleary MI, Ashby RH, Jelinek GA, Lagaida R. The future of casemix in emergency medicine and ambulatory care. Med J Aust 1994;161 (suppl):S30-3.

Correspondence to: Dr. Eric Grafstein, Department of Emergency Medicine, St. Paul's Hospital, 1085 Burrard St., Vancouver BC V6Z 1Y6; egrafstein@providencehealth.bc.ca

https://doi.org/10.1017/S1481803500015839 Published online by Cambridge University Press