



Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study

M N Storm-Versloot, D T Ubbink, V Chin a Choi and J S K Luitse

Emerg. Med. J. 2009;26:556-560
doi:10.1136/emj.2008.059378

Updated information and services can be found at:
<http://emj.bmj.com/cgi/content/full/26/8/556>

These include:

References

This article cites 18 articles, 2 of which can be accessed free at:
<http://emj.bmj.com/cgi/content/full/26/8/556#BIBL>

Rapid responses

You can respond to this article at:
<http://emj.bmj.com/cgi/eletter-submit/26/8/556>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Notes

To order reprints of this article go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Emergency Medicine Journal* go to:
<http://journals.bmj.com/subscriptions/>

Observer agreement of the Manchester Triage System and the Emergency Severity Index: a simulation study

M N Storm-Versloot,¹ D T Ubbink,² V Chin a Choi,¹ J S K Luitse¹

¹ Departments of Emergency Medicine, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands; ² Departments of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, Amsterdam, The Netherlands

Correspondence to:
Dr M Storm-Versloot,
Department of Surgery, G4-233,
Academic Medical Center, PO
Box 22700, 1100 DE
Amsterdam, The Netherlands;
m.n.storm@amc.uva.nl

Accepted 27 June 2008

ABSTRACT

Objectives: To compare inter and intra-observer agreement of the Manchester Triage System (MTS) and the Emergency Severity Index (ESI).

Methods: 50 representative emergency department (ED) scenarios derived from actual cases were presented to 18 ED nurses from three different hospitals. Eight of them were familiar with MTS, six with ESI and four were not familiar but trained in both systems. They independently assigned triage scores to each scenario according to the triage system(s) they were familiar with. After 4–6 weeks the same nurses again judged the scenarios in a different order. Unanimity in judgement and unweighted and quadratic-weighted kappas were calculated.

Results: Unanimity in judgement for MTS was 90% and for ESI 73%. One-level disagreement was found in 8% and 23% of the cases, respectively. Interobserver unweighted kappas were 0.76 (95% CI 0.68 to 0.83) for MTS and 0.46 (95% CI 0.37 to 0.55) for ESI. Quadratic-weighted kappas were 0.82 (95% CI 0.74 to 0.89) and 0.73 (95% CI 0.64 to 0.83), respectively. At 4–6 weeks, one-level intra-observer disagreements were 10% and 22% and 2-level disagreement 1% and 2%, respectively. Intra-observer unweighted kappas were 0.84 (95% CI 0.73 to 0.94) for MTS and 0.65 (95% CI 0.59 to 0.72) for ESI.

Conclusion: Using paper-based clinical scenarios, MTS was found to have a greater inter and intra-observer agreement than ESI.

Because the demand for emergency services outpaces available resources, emergency department (ED) triage systems face increasing scrutiny. Longer waits for care make the use of reliable, valid triage systems imperative to patient safety. Triage is defined as the initial clinical sorting process in hospital ED. ED generally use some form of triage, either formal or informal, in order to assess the patient's clinical needs and priority of care.¹ Informal triage systems rely on intuition and clinical experience of the ED nurse and decisions made cannot be tested afterwards. Formal triage systems offer more transparency,^{2,3} but depend on its reliability. This is usually expressed by means of a weighted kappa statistic. However, unanimity in judgement is rarely mentioned, which best illustrates triage uniformity.

Worldwide, four formal, five-level triage systems exist. In The Netherlands, these four systems have been critically appraised by the Dutch National Institute of Quality Control in Healthcare (CBO). Only the Manchester Triage System (MTS) and the Emergency Severity Index (ESI) were found to

be applicable in our country.² Therefore, both other systems are not in use in The Netherlands.

MTS comprises 52 flowcharts based on patient complaints. The presenting complaint is indicative of the severity and defines which flowchart is to be followed. Each flowchart is based on a five-step decision process that uses discriminators at each step to assign patients to one of the five triage levels.^{4,5} Although existing literature does not allow evaluation of validity, under and overtriage, and interobserver agreement of this system,^{6–8} the CBO has adopted MTS in the current guideline on triage at the ED.²

ESI uses one algorithm, with ratings ranging from level 1 (the most acutely ill patients) to level 5 (the least resource-intensive patients). For patients not meeting ESI level 1 or 2 criteria, the triage nurse estimates the number of resources needed to discharge the patient from the ED.⁹ The resource usage was shown to correlate well with the different triage levels,^{10–12} and showed a high interobserver agreement with quadratic-weighted kappas of 0.68–0.89.^{10,11,13–17}

Both MTS and ESI seem to be useful,² but no studies have compared the diagnostic validity and reliability of these systems. In this study, we focused on the reliability of both systems. We determined and compared the inter and intra-observer agreement by investigating judgement unanimity and the agreement of both triage systems with a reference standard.

METHODS

Study design, setting and population

This comparative clinical survey was performed at the Academic Medical Center (AMC) and Onze Lieve Vrouwe Gasthuis (OLVG) in Amsterdam and the Medical Spectrum Twente (MeSpTw) in Enschede, The Netherlands. The AMC is an urban tertiary care university hospital with a level 1 trauma centre. The ED sees almost 31 000 patients annually. The overall admission rate is approximately 18%. Approximately 16% of the patients are younger than 16 years. For triaging patients, ED nurses use an informal system.

The OLVG is an urban teaching hospital with a level 2 trauma centre and approximately 42 000 ED visits a year. The admission rate is approximately 10% and about 16% of the patients are younger than 16 years.¹⁸ Since 2003, ED nurses trained in ESI use this system for triaging their patients.

The MeSpTw is an urban teaching hospital with a level 1 trauma centre in the east of The Netherlands with almost 32 000 ED visits a year

at two locations. The admission rate is approximately 20%, but the number of patients younger than 16 years is not known. Since 2003, ED nurses trained in MTS use this system for triaging their patients.

Methods of measurement

Between November and December 2005, a total of 900 ED cases were prospectively collected at the AMC. For the purpose of triaging the patients according to MTS and ESI, six ED nurses received a 6-h combination of didactic and practical training for each system according to national standards. At random days of the week, between 12:00 and 22:00 hours, all consecutive patients entering the ED were triaged. Patients who were triaged by the ambulance staff before presentation at the ED and who met the criteria for treatment in the shock room according to current guidelines were not triaged again, but were classified as "red" (MTS) or "level 1" (ESI) patients. All patients gave oral informed consent for the study as the ethics review board waived the requirement for written informed consent.

Based on the 900 triaged patients, the distribution of the urgency levels was assessed. According to this distribution a representative sample of 50 cases was chosen. These cases were converted into written patient scenarios, using the documented triage notes and ED forms and checked by three nurses from the other contributing centres for comprehensible interpretation, missing data and feasible judgement. Scenarios included age and gender, chief complaint, patient's appearance, pain as expressed by the patient and scored by the nurse, history of presenting illness and vital signs such as pulse rate, blood pressure, temperature and oxygen saturation if appropriate. An example is given in box 1.

For MTS, the scenario writer assigned an urgency level to each scenario. Two independent expert nurses from the MeSpTw did the same. Disagreement existed in three out of 50 scenarios. These were adjudicated by another nurse. The final judgement was regarded as the reference standard. For ESI levels 1 or 2, the reference standard was determined by signs of a critical condition of the patient using the flowchart. For the remaining levels the actual ED resources used (laboratory testing, ECG, radiology, speciality consultations, intravenous fluids or hydration, intravenous or intramuscular medication, simple and complex procedures).

Eight ED nurses of the MeSpTw assigned urgency levels to the 50 patient scenarios using MTS, six ED nurses of the OLVG using ESI and four ED nurses of the AMC using both systems. No discussion was allowed during the assignment. The same judging procedure was repeated with the same scenarios in a different order after an interval of 4–6 weeks. The nurses were kept unaware of their own original assignment and were not allowed to discuss their assignments. This was achieved by supervising the nurses during the assignment sessions.

Box 1 Example of an ED patient scenario

A 76-year-old man is transported by ambulance to the ED. Yesterday evening he collapsed at home and according to his wife he did not want to get up. The whole night she spent with him on the floor. His consciousness was diminished. He was aphasic and had a paresis of his right arm. His vital signs were heart rate 116 beats/minute, blood pressure 111/73 mm Hg, body temperature 38.4°C.

Data analysis

Inter and intra-observer agreement was calculated using AGREE version 7 (Scienceplus Group, Groningen, The Netherlands), a software program dedicated to calculate kappa values. Kappa values lie between -1 and 1 . A kappa value of above 0.8 is called "very good", between 0.8 and 0.6 "good", between 0.6 and 0.4 "moderate" and below 0.4 "poor".¹⁹ To assess inter and intraobserver agreement among ED nurses and the agreement with the reference standard, pairwise kappa values were calculated, computing an unweighted and quadratic-weighted group kappa for several fixed observers.^{20–21} The group kappa gives a measure of average agreement between all the pairs in excess of chance.²² Differences in group kappa values between the hospitals as well as between less or more than 5 years ED experience in the MeSpTw hospital were calculated with Agree.

Because of possible shortcomings of kappa statistics,^{23–24} we also assessed the unanimity of judgements for the first judgement. This was defined as the percentage of scenarios given the same urgency level by all observers and the total number of judgements by all observers given the same urgency level.

Differences in nurse characteristics were analysed using SPSS version 12.

RESULTS

The mean age of all nurses was 39 years (SD 6.7). They had a mean of 9 years (SD 6.1) of ED experience. Nurse characteristics did not differ significantly among the hospitals, except for triage experience: Nurses from the OLVG and MeSpTw had 3 years of triage experience, whereas those of the AMC had none.

MTS scores as given by all observers were unanimous in 23 (46%) of all 50 scenarios; ESI scores in only five (10%) of all scenarios. A one-level urgency disagreement with respect to the triage classification according to MTS occurred in 22 (44%) scenarios and in 34 (68%) scenarios judged by ESI. Two-level disagreement occurred in five (10%) and nine (18%) scenarios, respectively. Three-level disagreement occurred in two (4%) scenarios, but only when using ESI.

In total, 594 (99%) of the 600 (50 scenarios judged by 12 nurses) urgency levels were obtained according to MTS and 498 (99%) of the 500 (50 scenarios judged by 10 nurses) according to ESI. A total of 534 (90%) of MTS-judged urgency levels were unanimous and 363 (73%) of ESI judgements. One-level disagreement occurred in 49 (8%) and 113 (23%) of the judgements, respectively, and two-level disagreements in 10 (2%) and 20 (4%).

Interobserver agreement by using the unweighted kappa was better with MTS and with more triage experience using MTS, but disappeared using the weighted kappa (table 1). For MTS no differences were found in years of ED experience.

Because one of the AMC nurses did not perform a second judgement, intra-observer analysis was calculated for 11 and nine nurses, respectively. Overall agreement between the first and second judgement was 89% for MTS (table 2) and 75% for ESI (table 3). Nearly all disagreements occurred within one level for both systems. Intra-observer agreement followed the same trend as interobserver agreement (table 4).

Compared with the reference standard, a 5% undertriage rate was found using MTS and 13% using ESI. For MTS (dis)agreement between the ED nurses' judgements and the reference standard followed the same trend as the intra and interobserver agreement (table 5). For ESI, agreement was lower than the intra-observer agreement, but the majority of disagreements was still within one level (table 6).

Table 1 Interobserver agreement for MTS and ESI, for each hospital and experience level of ED nurses at the MeSpTw, based on the first judgement

	Unweighted kappa (95% CI)	Quadratic-weighted kappa (95% CI)
MTS (n = 12)	0.76* (0.68 to 0.83)	0.82 (0.74 to 0.89)
MeSpTw (n = 8)	0.85† (0.77 to 0.93)	0.87 (0.79 to 0.95)
<5 years experience	0.87 (0.78 to 0.96)	0.90 (0.81 to 0.99)
>5 years experience	0.81 (0.71 to 0.91)	0.84 (0.73 to 0.95)
AMC (n = 4)	0.60† (0.48 to 0.73)	0.74 (0.63 to 0.85)
ESI (n = 10)	0.46* (0.37 to 0.55)	0.73 (0.64 to 0.83)
OLVG (n = 6)	0.48 (0.38 to 0.57)	0.75 (0.65 to 0.84)
AMC (n = 4)	0.41 (0.30 to 0.53)	0.72 (0.61 to 0.83)

*Significance level $p < 0.05$, between Manchester Triage System (MTS) and Emergency Severity Index (ESI); †significance level $p < 0.05$, between Medical Spectrum Twente, Enschede (MeSpTw) and Academic Medical Centre, Amsterdam (AMC). ED, emergency department; OLVG, Onze Lieve Vrouwe Gasthuis, Amsterdam.

Limitations

The limitations of our study are in the first place the use of standardised and abstracted case scenarios. Case scenarios are artificial, because they do not show the non-verbal clues from the live interview. Therefore, we used prospectively triaged patients and kept subjective information and physical appearance in the scenarios. Intra-observer reliability can only be determined if the first and second judgements are based on identical information, as only scenarios can provide.

Second, the strategy for selecting the reference standard was different for both triage systems. Both systems have a different conceptual foundation. We therefore used the judgement of expert ED nurses as reference standard for MTS in order to determine if the system was used correctly. In contrast, ESI not only scores patient urgency, but also required resources. Therefore, we determined the resources actually used. The number of resources needed depends on local hospital standards. Because of the close collaboration and similarity (patient population, emergency physician training, protocol use and available facilities) between AMC and OLVG hospitals this is not a likely confounder. For the purpose of this study we did not want to compare the diagnostic validity of both systems when the same reference standard should be used.

Third, we did not include level 1 urgency patients. The majority of these patients are already triaged before arrival at the hospital by ambulance staff and are transported directly to the shock room. Moreover, the inclusion of these level 1 cases would have overestimated the kappa values found in each triage group without altering differences between the two groups.

Fourth, the number of observers judging the scenarios in the AMC was rather small. Therefore, the difference in MTS scores found between the hospitals for inter and intra-observer agreement may have occurred due to outliers in the AMC. We

did not exclude these outliers, because it reflects actual clinical practice. For ESI, we did not find these effects.

Finally, we used the first version of the MTS and the third version of the ESI, although presently an updated version of both systems exists. At the time of the study, these versions were not available for Dutch hospitals.

DISCUSSION

By means of written case scenarios, we found the MTS to show a high degree of triage unanimity and a good agreement among ED nurses and when compared with a reference standard. For the ESI, the degree of unanimity was lower, but differences were usually not larger than one urgency level.

Our study is the first to compare triage agreement by means of MTS and ESI while distinguishing ED nurses with and without triage experience. Most studies on agreement only report the quadratic-weighted kappa, if at all specified, but rarely exact agreement or unanimity. Unanimity results are more conservative, while weighted kappa values appreciate near disagreement. This seems right for the lower urgency levels, but a one-level difference in the higher urgency levels can delay treatment, which is potentially dangerous to the patient. We did report weighted kappa values to allow comparison between our results and those from other studies.

Few studies report the agreement between triage judgements by ED nurses and a reference standard, mostly based on case scenarios. Our results for ESI are comparable with existing literature, showing good weighted kappa values, ranging from 0.68 to 0.89.^{11 13-17} Few studies focus on MTS, but an unweighted kappa of 0.60 for the inexperienced hospital in our study matches a value of 0.63 as reported in the guideline on triage.² The remaining kappa values we found were much better.

Table 2 MTS: comparison of ED nurses' triage judgements between the first and second judgement

ED nurse: second judgement	ED nurse: first judgement					
	Red	Orange	Yellow	Green	Blue	Total
Red	1	2	0	0	0	3
Orange	0	107	5	1	0	113
Yellow	0	5	173	19	2	199
Green	0	3	10	195	5	213
Blue	0	0	1	6	9	16
Total	1	117	189	221	16	544

Overall agreement Manchester Triage System (MTS): 485/544 (89.2%); one-level disagreement: 52/544 (9.6%); two-level disagreement: 7/544 (1.3%). ED, emergency department.

Table 3 ESI: Comparison of ED nurses' triage judgements between the first and second judgement

ED nurse: second judgement	ED nurse: first judgement					
	Level 1	Level 2	Level 3	Level 4	Level 5	Total
Level 1	0	0	0	0	0	0
Level 2	0	90	9	2	1	102
Level 3	0	11	132	10	4	157
Level 4	0	1	26	31	22	80
Level 5	0	1	2	21	84	108
Total	0	103	169	64	111	447

Overall agreement Emergency Severity Index (ESI): 337/447 (75.4%); one-level disagreement: 99/447 (22.1%); two-level disagreement: 9/447 (2.0%); three-level disagreement: 2/447 (0.4%). ED, emergency department.

Table 4 Intra-observer agreement for MTS and ESI, for each hospital and experience level of ED nurses at the MeSpTw

	Unweighted kappa (95% CI)	Quadratic-weighted kappa (95% CI)
MTS (n = 11)	0.84* (0.73 to 0.94)	0.90 (0.83 to 0.96)
MeSpTw (n = 8)	0.91 (0.83 to 0.98)	0.94 (0.88 to 0.99)
<5 years experience	0.91 (0.74 to 1.00)	0.94 (0.83 to 1.00)
>5 years experience	0.91 (0.78 to 1.00)	0.93 (0.83 to 1.00)
AMC (n = 3)	0.65 (0.30 to 1.00)	0.79 (0.59 to 0.99)
ESI (n = 9)	0.65* (0.59 to 0.72)	0.85 (0.80 to 0.90)
OLVG (n = 6)	0.65 (0.56 to 0.74)	0.84 (0.78 to 0.90)
AMC (n = 3)	0.66 (0.38 to 0.95)	0.88 (0.70 to 1.00)

*Significance level $p < 0.05$, between Manchester Triage System (MTS) and Emergency Severity Index (ESI). AMC, Academic Medical Centre, Amsterdam; ED, emergency department; MeSpTw, Medical Spectrum Twente, Enschede; OLVG, Onze Lieve Vrouwe Gasthuis, Amsterdam.

We found a significant difference using the unweighted kappa, which disappeared by using the weighted kappa, both for inter and intra-observer agreement and for the agreement with the reference standard. This is because the weighted kappa accepts one-level disagreement as some form of agreement, which mitigates the differences. A one-level disagreement was also most common in previous studies.^{11–16} Unfortunately, we cannot compare the magnitude of disagreement for the inter and intra-observer agreement. We therefore recommend that in future studies both exact agreement and weighted kappa values are presented.

Some studies report undertriage rates varying from 9% to 12%. In these studies the judgements were compared with the “true” urgency ESI triage level, determined by an expert panel.^{14–16} If undertriage occurs, potentially seriously ill patients may be triaged as non-urgent, resulting in an increasing risk of adverse outcomes for these patients. We found an undertriage rate of 13% (67 judgements) for ESI. Of these, only 14 judgements spread over five scenarios showed a two-level disagreement. This may seem serious, but to determine the actual consequences of these undertriage judgements the diagnostic validity of the system also has to be assessed. The relatively low unanimity and high disagreement of ESI might be because the determination of urgency depends on implicit knowledge rather than explicit flowcharts.

For MTS only two judgements in two scenarios showed a two-level disagreement. Cooke and Jinks⁶ reported that almost 20% of incorrect classifications in critically ill patients were due to non-adherence to the MTS guideline; most errors were because of training problems rather than the triage system. We found less non-adherence, but we did not restrict our cases to

critically ill patients. All nurses were trained according to standard procedures before using MTS. In a computer-aided environment, adherence is easier to achieve. Unanimity of scoring with MTS can reach 68% without and 96% with computerised decision support.² In our study judgement was performed without computer support, but the flow charts could be consulted. We found a fairly high number of unanimous judgements (90%), although some difference occurred between the AMC (84%) and the MeSpTw (94%). Apparently, nurses need to learn how to use the system correctly. Using a computer aid could help overcome the nurses’ tendency to follow their own line of reasoning in interpreting MTS flowcharts.

CONCLUSIONS

We conclude that MTS has very good agreement and a high unanimous classification rate, whereas ESI has only moderate to good results. For MTS, agreement was not influenced by the ED nurses’ experience, but appeared to be affected by the level of experience with the system. Determination of triage system reliability is a necessary step in establishing its usefulness and is pivotal in any attempt to measure performance in emergency medicine. Beyond triage reliability, as was investigated here, diagnostic validity should be determined of both systems by comparing the triage classifications with an identical reference standard.

Acknowledgements: The authors are grateful to the following members of the study group: ER Schinkel (PhD), R Köhlinger (RN) and DC Schutte (RN) for their enthusiastic and critical cooperation. Also thanks to the heads of the ED departments of the AMC for supporting and facilitating this study at their department: W de Graaf and NF ten Grotenhuis. Furthermore, the authors would also like to thank MW van de Kamp for

Table 5 MTS: Comparison between ED nurses’ triage judgement and the reference standard based on the first judgement

ED nurse judgement	Reference standard					Total
	Red	Orange	Yellow	Green	Blue	
Red	0	3	0	0	0	3
Orange	0	111	5	8	0	124
Yellow	0	5	199	12	2	218
Green	0	1	22	208	2	233
Blue	0	0	1	7	8	16
Total	0	120	227	235	12	594

Agreement Manchester Triage System (MTS): 526/594 (88.6%); one-level disagreement: 56/594 (9.4%); two-level disagreement: 12/594 (2.0%); undertriage: 32/594 (5.4%); overtriage: 36/594 (6.1%); unweighted kappa: 0.84 (0.80 to 0.87); quadratic-weighted kappa: 0.87 (0.70 to 1.00). ED, emergency department.

Table 6 ESI: Comparison between ED nurses’ triage judgement and the reference standard based on the first judgement

ED nurse judgement	Reference standard					Total
	Level 1	Level 2	Level 3	Level 4	Level 5	
Level 1	0	0	0	0	0	0
Level 2	0	95	6	7	3	111
Level 3	0	29	60	72	20	181
Level 4	0	4	16	31	39	90
Level 5	0	2	8	8	98	116
Total	0	130	90	118	160	498

Agreement Emergency Severity Index (ESI): 284/498 (57.0%); one-level disagreement: 170/498 (34.1%); two-level disagreement: 39/498 (7.8%); three-level disagreement: 5/498 (1.0%); undertriage: 67/498 (13.5%); overtriage: 147/498 (29.5%); unweighted kappa: 0.43 (0.38 to 0.49); quadratic-weighted kappa: 0.71 (0.61 to 0.82). ED, emergency department.

her support during the second part of data analysis. They also wish to thank the heads of the ED departments at the Medical Spectrum Twente and the Onze Lieve Vrouwe Gasthuis: C Schenkeveld, J Huiskes, and BPM Huybrechts for their enthusiasm and assistance to make this study possible. The authors also thank all the nurses who participated.

Competing interests: None.

Patient consent: Obtained.

REFERENCES

- Fernandes CM**, Wuerz R, Clark S, *et al.* How reliable is emergency department triage? *Annals of emergency medicine* 1999;**34**:141–7.
- Coenen IGA**, Hagemeyer A, De Caluwé R, *et al.* Richtlijn triage op de Spoedeisende Hulp. Kwaliteitsinstituut voor de Gezondheidszorg CBO, 2004. http://www.cbo.nl/product/richtlijnen/folder20021023121843/triage_rl_2005.pdf/ (accessed May 2009).
- Cronin JG**. The introduction of the Manchester triage scale to an emergency department in the Republic of Ireland. *Accid Emerg Nurs* 2003;**11**:121–5.
- Manchester Triage Group**. *Emergency triage*. London: BMJ Publishing, 1997.
- De Caluwé R**, De Voeght F, Breuer G, *et al.* *Triage voor de spoedeisende hulp*. Maarssen: Elsevier Gezondheidszorg, 2002. ISBN 90 352 2545 7.
- Cooke MW**, Jinks S. Does the Manchester Triage System detect the critically ill? *J Accident Emerg Med* 1999;**16**:179–81.
- Evans A**, O'Connor N, St John A, *et al.* Accident and emergency triage audit report. *J Clin Governance* 2002;**10**:195–8.
- Speake D**, Teece S, Mackway-Jones K. Detecting high-risk patients with chest pain. *Emerg Nurse* 2003;**11**:19–21.
- Agency for Healthcare Research and Quality**. *Emergency severity index, version 4: Implementation handbook*, May 2005. <http://www.ahrq.gov/research/esi> (accessed May 2009).
- Tanabe P**, Gimbel R, Yarnold PR, *et al.* The Emergency Severity Index (version 3) 5-level triage system scores predict ED resource consumption. *J Emerg Nurs* 2004;**30**:22–9.
- Eitel DR**, Travers DA, Rosenau AM, *et al.* The Emergency Severity Index triage algorithm version 2 is reliable and valid. *Acad Emerg Med* 2003;**10**:1070–80.
- Elshove-Bolk J**, Mencl F, van Rijswijk BTF, *et al.* Validation of the Emergency Severity Index (ESI) in self-referred patients in a European emergency department. *Eur J Emerg Med* 2006;**13**:325–9.
- Wuerz RC**, Milne LW, Eitel DR, *et al.* Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000;**7**:236–42.
- Tanabe P**, Gimbel R, Yarnold PR, *et al.* Reliability and validity of scores on the Emergency Severity Index version 3. *Acad Emerg Med* 2004;**11**:59–65.
- Worster A**, Gilboy N, Fernandes CM, *et al.* Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *Can J Emerg Med* 2004;**6**:240–5.
- Travers DA**, Waller AE, Bowling JM, *et al.* Five-level triage system more effective than three-level in tertiary emergency department. *J Emerg Nurs* 2002;**28**:395–400.
- Wuerz RC**, Travers D, Gilboy N, *et al.* Implementation and refinement of the Emergency Severity Index. *Acad Emerg Med* 2001;**8**:170–6.
- Elshove-Bolk J**, Mencl F, van Rijswijk BTF, *et al.* Emergency department patient characteristics: potential impact on emergency medicine residency programs in the Netherlands. *Eur J Emerg Med* 2006;**13**:325–9.
- Landis JR**, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- Cohen J**. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213–20.
- Altman DG**. *Practical statistics for medical research*. London: Chapman and Hall, 1991:407–9.
- Fleiss JL**. The measurement of interrater agreement. In: *Statistical methods for rates and proportions*, 2nd ed. New York: John Wiley, 1981:212–36.
- Brenner H**, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996;**7**:199–202.
- Sim J**, Wright CC. The kappa statistic in reliability studies: use, interpretation, and the sample size requirements. *Phys Ther* 2005;**85**:257–68.