



OPEN ACCESS

# Inter-rater and intrarater reliability of the South African Triage Scale in low-resource settings of Haiti and Afghanistan

Mohammed Dalwai,<sup>1,2</sup> Katie Tayler-Smith,<sup>3</sup> Michèle Twomey,<sup>1</sup> Masood Nasim,<sup>4</sup> Abdul Qayum Popal,<sup>4</sup> Waliul Haq Haqdot,<sup>5</sup> Olivia Gayraud,<sup>6</sup> Sophia Chéréal,<sup>6</sup> Lee Wallis, Pola Valles<sup>2</sup>

<sup>1</sup>Division of Emergency Medicine, University of Cape Town, Cape Town, South Africa  
<sup>2</sup>Medical Department, Médecins Sans Frontières, Operational Centre Brussels, Brussels, Belgium  
<sup>3</sup>Operational Research Unit Luxembourg, Médecins Sans Frontières, Luxembourg City, Luxembourg  
<sup>4</sup>Medical Department, Médecins Sans Frontières, Kabul, Afghanistan  
<sup>5</sup>Ministry of Health, Kabul, Afghanistan  
<sup>6</sup>Medical Department, Médecins Sans Frontières, Port au Prince, Haiti

## Correspondence to

Dr Mohammed Dalwai, Division of Emergency Medicine, University of Cape Town, Cape Town 7708, South Africa; mkdalwai@gmail.com

Received 18 July 2017

Revised 6 February 2018

Accepted 13 February 2018

## ABSTRACT

**Objective** The South African Triage Scale (SATS) has demonstrated good validity in the EDs of Médecins Sans Frontières (MSF)-supported sites in Afghanistan and Haiti; however, corresponding reliability in these settings has not yet been reported on. This study set out to assess the inter-rater and intrarater reliability of the SATS in four MSF-supported EDs in Afghanistan and Haiti (two trauma-only EDs and two mixed (including both medical and trauma cases) EDs).

**Methods** Under classroom conditions between December 2013 and February 2014, ED nurses at each site assigned triage ratings to a set of context-specific vignettes (written case reports of ED patients). Inter-rater reliability was assessed by comparing triage ratings among nurses; intrarater reliability was assessed by asking the nurses to retriage 10 random vignettes from the original set and comparing these duplicate ratings. Inter-rater reliability was calculated using the unweighted kappa, linearly weighted kappa and quadratically weighted kappa (QWK) statistics, and the intraclass correlation coefficient (ICC). Intrarater reliability was calculated according to the percentage of exact agreement and the percentage of agreement allowing for one level of discrepancy in triage ratings. The correlation between years of nursing experience and reliability of the SATS was assessed based on comparison of ICCs and the respective 95% CIs.

**Results** A total of 67 nurses agreed to participate in the study: In Afghanistan there were 19 nurses from Kunduz Trauma Centre and nine from Ahmed Shah Baba; in Haiti, there were 20 nurses from Martissant Emergency Centre and 19 from Tabarre Surgical and Trauma Centre. Inter-rater agreement was moderate across all sites (ICC range: 0.50–0.60; QWK range: 0.50–0.59) apart from the trauma ED in Haiti where it was moderate to substantial (ICC: 0.58; QWK: 0.61). Intrarater agreement was similar across the four sites (68%–74% exact agreement); when allowing for a one-level discrepancy in triage ratings, intrarater reliability was near perfect across all sites (96%–99%). No significant correlation was found between years of nursing experience and reliability.

**Conclusion** The SATS has moderate reliability in different EDs in Afghanistan and Haiti. These findings, together with concurrent findings showing that the SATS has good validity in the same settings, provide evidence to suggest that SATS is suitable in trauma-only and mixed EDs in low-resource settings.

## Key messages

### What is already known on this subject

► There are few triage scales designed specifically for use in low/middle-income countries (LMIC); the South African Triage Scale (SATS) is one of them and has been shown to have good validity in such settings. The inter-rater reliability of SATS in South Africa has been reported as moderate to substantial, with intrarater reliability ranging from 80% to 86%. Its performance across a spectrum of different LMIC settings, mainly non-sub-Saharan African and trauma-only settings, has not been adequately evaluated.

### What this study adds

► In this cross-sectional study using case vignettes, ED nurses in Afghanistan and Haiti assigned triage ratings using the SATS. Inter-rater reliability was moderate and intrarater reliability for exact agreement ranged from 68% to 74%. Added to evidence showing good validity of this scale, this suggests the SATS could be suitable for low-resource settings.

## INTRODUCTION

Triage has a central role in emergency care systems: prioritising patients based on acuity improves effective use of resources, and ultimately patient outcomes.<sup>1</sup> A number of different scales exist for in-hospital use, but most of these have been developed for and evaluated in high-resource settings.<sup>2,3</sup> Context-appropriate triage tools for low/middle-income countries (LMIC) are very uncommon.<sup>4</sup> Among the few tools that have been contextually modified, validated and implemented in various settings is the South African Triage Scale (SATS), which was developed for in-hospital EDs.<sup>5</sup> The SATS has been assessed extensively in South Africa and implemented in several settings,<sup>6–8</sup> but further assessment of its performance in low-resource settings, particularly non-sub-Saharan settings, is still needed.<sup>4,9</sup>

For a triage scale to be effective, it needs to demonstrate good validity (ie, an acuity rating assigned using the scale must closely reflect a patient's true acuity) and a high degree of reliability



**To cite:** Dalwai M, Tayler-Smith K, Twomey M, *et al.* *Emerg Med J* Epub ahead of print: [please include Day Month Year]. doi:10.1136/emered-2017-207062

**Table 1** Characteristics of the study sites in Afghanistan and Haiti

|                                | Ahmad Shah Baba  | Kunduz Trauma Centre  | Martissant Emergency Centre         | Tabarre Surgical and Trauma Centre   |
|--------------------------------|--|---|-------------------------------------|--|
| Country                        | Afghanistan  | Afghanistan   | Haiti                               | Haiti  |
| Location                       | Kabul City, district 12  | Kunduz Province   | Port-au-Prince, Martissant district | Port-au-Prince, Tabarre district   |
| Estimated catchment population | 219 000  | 1 000 000   | 1 200 000                           | 1 000 000  |
| Level of MSF support           | Partnership with Ministry of Health                                | MSF only  | MSF only                            | MSF only   |
| Services offered               | OPD<br>ED<br>Maternity IPD: surgery, internal medicine, paediatric | OPD<br>ED<br>IPD trauma care: surgery, orthopaedic ICU, physiotherapy | ED only                             | OPD<br>ED<br>IPD visceral and trauma care: surgery, orthopaedic ICU, physiotherapy |
| Type of ED cases               | Mixed  | Trauma only   | Mixed                               | Visceral surgery and trauma  |
| ED caseload (per month)*       | 4715   | 1848  | 4919                                | 793  |
| Introduction of the SATS       | 2011   | 2011  | 2013                                | 2012   |

\*Mean cases seen per month during 2014.

ICU, intensive care unit; IPD, inpatient department; MSF, Médecins Sans Frontières; OPD, outpatient department; SATS, South African Triage Scale.

(ie, it must yield the same triage rating on repeated assessments of the same patient). For any given patient, tools should have high inter-rater (the degree of variability among different nurses) and intrarater (the variability of retriage ratings for one nurse) reliability.

Médecins Sans Frontières (MSF), an international medical humanitarian organisation, provides free medical care to vulnerable populations in many LMIC settings. It operates within constrained resources and serves populations with little health-care access. Since 2011, MSF-Operational Centre Brussels has implemented the SATS in projects where it provides emergency care. The validity of the SATS was recently assessed in the EDs of MSF-supported sites in Afghanistan and Haiti<sup>10</sup> with good results, but corresponding reliability in these sites has yet to be reported on. This is the basis of the current study.

## METHODS

### Study design

This was a cross-sectional study using a set of ED vignettes (short written clinical case reports of actual ED patients) as a proxy for live patients, in which ED nurses assigned triage ratings using the SATS.

### Study setting

The study was conducted at four active MSF project sites between December 2013 and February 2014: two hospitals in Afghanistan (Ahmad Shah Baba (ASB) and Kunduz Trauma Centre (KTC)) and two facilities in Haiti (Martissant Emergency Centre (MT) and Tabarre Surgical and Trauma Centre (TB)). Specific details on these four sites are summarised in [table 1](#).

### SATS and its use in the ED

Described in detail elsewhere,<sup>10</sup> the SATS is a four-tiered triage tool which depicts a patient's urgency for care using the following colour codes: priority 1: red—'emergency' (to be seen immediately); priority 2: orange—'very urgent' (to be seen within 10 min); priority 3: yellow—'urgent' (to be seen within 60 min); priority 4: green—'routine' (to be seen within 240 min). The SATS also allocates the colour blue (black was used in the study countries for cultural purposes) to 'dead on arrival' cases.

### Study population

The study included all ED nurses at the four study sites who fulfilled the following inclusion criteria: (1) had received training in use of the SATS and (2) agreed to participate in the study. All nurses employed by MSF have a basic nursing degree and are registered with the country nursing authority.

### Study protocol

Under classroom conditions, all nurses who agreed to participate in the study were asked to use the SATS to triage a set of vignettes and assign one of the following four categories to each vignette: 'emergency', 'very urgent', 'urgent' and 'routine'. Each set comprised between 28 and 30 vignettes generated from information extracted from randomly selected patient files of real ED cases who had presented at the study centres between June and December 2013. Each vignette included information on patient gender, age, presenting complaint, mode of arrival to the ED and vital signs. All clinical information in the triage paperwork was copied into the vignettes including information from additional investigations such as blood glucose and haemoglobin levels (see [box 1](#) for an example of a vignette).

Professionals translated the vignettes from English into the relevant local languages. Local bilingual doctors ratified the translations to ensure correct medical terminology.

Under classroom conditions, all nurses who agreed to participate in the study assigned one of four SATS categories to the set of reference vignettes.

### Data analysis

Inter-rater reliability was measured by comparing the triage ratings assigned for each of the vignettes by different nurses at each

#### Box 1 Example of a vignette used to assess the South African Triage Scale (SATS) in Afghanistan and Haiti, 2013

A 17-year-old boy presents with abdominal pain, loose motion and vomiting since this morning. He says he ate something last night that did not agree with his stomach and since this morning has not been feeling well. At triage, you find an alert boy with moderate abdominal pain. No signs of dehydration are present.

BP: 120/80; HR: 109; RR: 16; temperature: 36°C

**Table 2** Nurses' response rate at each study site in Afghanistan and Haiti, 2013

| Study site        | Nurses invited to participate (n) | Nurses agreeing to participate (n) | Response rate (%) |
|-------------------|-----------------------------------|------------------------------------|-------------------|
| KTC (Afghanistan) | 21                                | 19                                 | 90                |
| ASB (Afghanistan) | 9                                 | 9                                  | 100               |
| MT (Haiti)        | 21                                | 20                                 | 95                |
| TB (Haiti)        | 20                                | 19                                 | 95                |

ASB, Ahmad Shah Baba; KTC, Kunduz Trauma Centre; MT, Martissant Emergency Centre; TB, Tabarre Surgical and Trauma Centre.

study site. Intrarater reliability was measured by asking nurses to retriage 10 random vignettes from the original set 1–10 days later (depending on their availability), and comparing these duplicate ratings.

In accordance with the Guidelines for Reporting Reliability and Agreement Studies (GRRAS), inter-rater reliability was assessed using the unweighted kappa (UWK), linearly weighted kappa (LWK) and quadratically weighted kappa (QWK) statistics, as well as the intraclass correlation coefficient (ICC).<sup>11</sup> UWK and LWK point estimates were assessed and included as per GRRAS guidelines, but in keeping with triage literature we only interpreted QWK and ICC point estimates using the Landis and Koch classification system: 0.0–0.20—slight agreement; 0.21–0.40—fair agreement; 0.41–0.60—moderate agreement; 0.61–0.80—substantial agreement; 0.81–1.00—almost perfect agreement.<sup>12</sup> In triage reliability studies, UWK and LWK can be ignored. QWK and ICC yield almost identical results hence either one could be used based on ease of calculation.<sup>12</sup>

Intrarater reliability was assessed by calculating both the percentage of exact agreement and the percentage of agreement allowing for one level of discrepancy in triage ratings. 95% CIs were calculated for all measures.

In addition, we assessed whether there was any correlation between years of nursing experience (ie, years of being a qualified nurse) and the ICC based on comparison of the 95% CIs and use of bootstrapping.

## RESULTS

### Study population

Table 2 shows the sample size at each study site. The response rate ranged from 90% in KTC to 100% in ASB.

### Reliability of nurse triage ratings

Table 3 summarises the different reliability measures calculated to assess inter-rater and intrarater reliability across the four study sites. Inter-rater agreement was moderate across all sites, apart from TB where it was moderate to substantial. Trauma-only facilities (KTC and TB) yielded very similar results (ICC: 0.60 and 0.58 and QWK: 0.59 and 0.61, respectively) whereas among the mixed settings (ASB and MT), there was a wider variability in results (ICC: 0.50 and 0.59 and QWK: 0.50 and 0.59, respectively).

Intrarater agreement was similar across the four sites, ranging from 68% exact agreement in ASB to 74% in MT. When allowing for a one-level discrepancy in triage ratings, intrarater reliability was near perfect across all sites ranging from 96% in TB and ASB to 99% in KTC.

Table 4 shows the correlation between years of nursing experience and ICC across the four sites. The mean years of nursing experience were similar across all sites ranging (6.3–7.1 years). The ICC for nurses with 5 or more years of nursing experience appeared to be higher than for those with less than 5 years of experience, but 95% CIs overlapped (even after applying a bootstrapping technique) indicating no statistical significant difference.

## DISCUSSION

Our study shows that the SATS has moderate inter-rater and intrarater reliability when used by nurses in trauma-only and mixed ED settings in Afghanistan and Haiti. This is evidenced to suggest that the SATS could be suitable for use in low-resource settings. Further reliability studies in low-resource settings are needed to confirm these findings.

The main strengths of this study are its multisite nature, the high response rate of participants and the fact that the vignettes reflected real ED cases seen in each specific setting. In previous studies assessing the SATS in contexts outside of South Africa, the vignettes used were based on South African ED cases, not ED cases specific to the study setting.<sup>8,9</sup>

### Limitations

There were a number of study limitations. First, using paper-based vignettes as a proxy for real ED cases has the inherent limitation of not mimicking real life.<sup>9</sup> Although conducting consecutive live triage assessments on a single patient at one point in time and at multiple points in time is not feasible or practical,<sup>13</sup> use of paper-based vignettes assessed under

**Table 3** Inter-rater and intrarater reliability measures for the SATS in Afghanistan and Haiti, 2013

| Inter-rater reliability measures             | Ahmad Shah Baba     | Kunduz Trauma Centre | Martissant Emergency Centre | Tabarre Surgical and Trauma Centre |
|--|---------------------|----------------------|-----------------------------|------------------------------------|
| Point estimates (95% CI)                     |                     |                      |                             |                                    |
| ICC  | 0.50 (0.37 to 0.66) | 0.60 (0.48 to 0.74)  | 0.59 (0.46 to 0.73)         | 0.58 (0.44 to 0.73)                |
| QWK  | 0.50 (0.37 to 0.66) | 0.59 (0.22 to 0.77)  | 0.59 (0.32 to 0.78)         | 0.61 (0.32 to 0.79)                |
| LWK  | 0.44 (0.16 to 0.69) | 0.46 (0.21 to 0.66)  | 0.50 (0.27 to 0.70)         | 0.48 (0.23 to 0.68)                |
| UWK  | 0.40 (0.26 to 0.53) | 0.33 (0.23 to 0.41)  | 0.41 (0.31 to 0.48)         | 0.35 (0.23 to 0.47)                |
| Level of agreement*                          |                     |                      |                             |                                    |
| ICC  | Moderate            | Moderate             | Moderate                    | Moderate                           |
| QWK  | Moderate            | Moderate             | Moderate                    | Substantial                        |
| Intrarater reliability measures, % (min-max) |                     |                      |                             |                                    |
| Mean exact agreement                         | 68 (40–90)          | 71 (40–100)          | 74 (40–90)                  | 73 (50–90)                         |
| Mean agreement with one-degree discrepancy   | 96 (80–100)         | 99 (90–100)          | 97 (60–100)                 | 96 (70–100)                        |

\*According to the Landis and Koch criteria.<sup>11</sup>

ICC, intraclass correlation coefficient; LWK, linearly weighted kappa; QWK, quadratically weighted kappa; SATS, South African Triage Scale; UWK, unweighted kappa.

**Table 4** Effect of nurse experience on inter-rater reliability of the SATS

| Study site        | Nurses agreeing to participate (n) | Mean years of experience | Nurses with <5 years of experience |                  | Nurses with ≥5 years of experience |                  |
|-------------------|------------------------------------|--------------------------|------------------------------------|------------------|------------------------------------|------------------|
|                   |                                    |                          | n                                  | ICC              | n                                  | ICC              |
| KTC (Afghanistan) | 19                                 | 6.6                      | 11                                 | 0.53 (0.38–0.68) | 9                                  | 0.67 (0.54–0.79) |
| ASB (Afghanistan) | 8*                                 | 7.1                      | 5                                  | 0.46 (0.29–0.64) | 3                                  | 0.55 (0.35–0.74) |
| MT (Haiti)        | 20                                 | 6.3                      | 11                                 | 0.53 (0.38–0.68) | 9                                  | 0.64 (0.50–0.78) |
| TB (Haiti)        | 19                                 | 6.3                      | 6                                  | 0.70 (0.57–0.83) | 13                                 | 0.57 (0.43–0.71) |

\*Information available on eight out of the nine nurses.

ASB, Ahmad Shah Baba; ICC, intraclass correlation coefficient; KTC, Kunduz Trauma Centre; MT, Martissant Emergency Centre; SATS, South African Triage Scale; TB, Tabarre Surgical and Trauma Centre.

classroom conditions may have influenced the relative degree of reliability that was observed. For example, the wording of the vignettes may have been interpreted differently by different nurses. That said, a previous study has shown that there is little difference between the inter-rater reliability measures generated using paper-based cases compared with live cases.<sup>13</sup> Second, translation of the vignettes from English into the local language may have slightly distorted some of the original information. We tried to limit this by recruiting professional translators with some medical background to carry out the translations in each setting, and having local medical staff back translate.

Originally developed for use in South Africa, the SATS has been assessed extensively in South Africa, and also in Botswana, Malawi and Pakistan with good results.<sup>4 7 8 14</sup> But the degree to which these findings are applicable to other LMIC settings—particularly those outside of sub-Saharan Africa and those that deal with trauma-only caseloads—has remained unclear. This is what prompted a recent study assessing the validity of the SATS in different EDs in Afghanistan and Haiti.<sup>10</sup> The results of this study were good, but reliability in these settings was still unknown. Reliability of triage across both high-resource and low-resource settings varies greatly. Two articles assessing reliability in South Africa report moderate to substantial reliability with QWK of 0.57 and 0.66, respectively.<sup>14 15</sup> In Ghana, the SATS showed moderate reliability with QWK of 0.59 and 0.60<sup>16</sup> while studies in Pakistan and Botswana reported substantial to near-perfect results with QWK of 0.77 and 0.87, respectively.<sup>8 9</sup> In high-resource settings, the Canadian Emergency Department Triage and Acuity Scale (CTAS), a 5-level triage scale, reported a chance corrected kappa of 0.80 and a weighted kappa of 0.77.<sup>17 18</sup> The Emergency Severity Index (ESI) has reported inter-rater reliability ranging from 0.76 to 0.8 with the Manchester Triage System (MTS) showing a weighted kappa from 0.62 to 0.82.<sup>2 19 20</sup> No studies were found in low-resource settings for either the CTAS or MTS. The ESI was implemented in Iran but according to Mirhaghi *et al* may not reveal optimal outcomes for LMICs.<sup>21</sup> Standardisation of reporting reliability is poor with some studies not identifying which weighted kappa statistic was used to calculate reliability, making comparisons between studies difficult.<sup>11</sup> The one-two-triage scale, the only other new scale developed in 2015 for low-resource settings, reported a kappa of 0.308 among nurses in Cambodia.<sup>22</sup>

The results of our study confirm that the SATS is valid in Haiti and Afghanistan and demonstrates moderate reliability. This latter finding is most certainly a reflection of the relative simplicity of the SATS, both in terms of its construct and application, and supports its value in resource-constrained settings where highly skilled staff are often in short supply.

Reliable use of the SATS did appear to be higher among nurses with 5 or more years of nursing experience, although our results were not statistically significant. The latter however may be related to our relatively small sample size, and thus low statistical power.

This finding is similar to previous research by Göransson *et al* that found no significant difference between nursing experience and reliability of triage when using the Canadian Triage and Acuity Scale.<sup>18</sup>

In addition, there may be other factors that influence reliability and which confounded the relationship between years of experience and reliable use of the SATS, for example, how regularly the nurses were working in triage (all the nurses were working on a rotational basis and therefore were not permanently based in the ED).

It would be useful to explore these sorts of factors further in order to establish how they affect reliability and ultimately what could be done to optimise the reliable use of the SATS.

## CONCLUSION

In conclusion, our study shows that the SATS is a moderately reliable tool for use in different EDs in Afghanistan and Haiti. These findings, together with concurrent findings showing that the SATS has good validity in the same settings, provide evidence to suggest that SATS is suitable in trauma-only and mixed EDs in low-resource settings.

**Acknowledgements** We thank all the staff who participated in the study both nationally and internationally for going above and beyond and always trying to improve healthcare for vulnerable populations. We also thank the Centre for Evidence-based Health Care at the University of Stellenbosch for help with the statistics.

**Contributors** MD, PV, MT, LW and KTS designed, analysed and interpreted the study and data. AQP, WHH and MN were the project leads in Afghanistan. OG and SC were the project leads in Haiti. All authors contributed to the revision of the final article.

**Funding** The degree from which this study emanated was funded by the South African Medical Research Council under the SAMRC Clinicians Researcher Development Scholarship PhD programme.

**Competing interests** None declared.

**Patient consent** Not required.

**Ethics approval** Ethics approval was obtained from the National Ethics Committees in Afghanistan and Haiti, from the MSF Ethics Review Board and from the University of Cape Town (UCT).

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Data sharing is available on request.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.



## REFERENCES

- Robison JA, Ahmad ZP, Nosek CA, *et al.* Decreased pediatric hospital mortality after an intervention to improve emergency care in Lilongwe, Malawi. *Pediatrics* 2012;130:e676–82.
- 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.
- van der Wulp I, van Baar ME, Schrijvers AJ. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. *Emerg Med J* 2008;25:431–4.
- Dalwai MK, Tayler-Smith K, Trelles M, *et al.* Implementation of a triage score system in an emergency room in Timergara, Pakistan [Short communication]. *Public Health Action* 2013;3:43–5.
- Gottschalk SB, Wood D, DeVries S, *et al.* The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006;23:149–53.
- Twomey M, Wallis LA, Myers JE. Evaluating the construct of triage acuity against a set of reference vignettes developed via modified Delphi method. *Emerg Med J* 2014;31:562–6.
- Harrison HL, Raghunath N, Twomey M. Emergency triage, assessment and treatment at a district hospital in Malawi. *Emerg Med J* 2012;29:924–5.
- Twomey M, Mullan PC, Torrey SB, *et al.* The Princess Marina Hospital accident and emergency triage scale provides highly reliable triage acuity ratings. *Emerg Med J* 2012;29:650–3.
- Dalwai MK, Twomey M, Maikere J, *et al.* Reliability and accuracy of the South African Triage Scale when used by nurses in the emergency department of Timergara Hospital, Pakistan. *S Afr Med J* 2014;104:372.
- Dalwai M, Valles P, Twomey M, *et al.* Is the South African Triage Scale valid for use in Afghanistan, Haiti and Sierra Leone? *BMJ Glob Health* 2017;2:e000160.
- Kottner J, Audigé L, Brorson S, *et al.* Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.
- Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their development and use*. 5th ed. New York: Oxford University Press, 2015. doi.
- Worster A, Sardo A, Eva K, *et al.* Triage tool inter-rater reliability: a comparison of live versus paper case scenarios. *J Emerg Nurs* 2007;33:319–23.
- Twomey M, Wallis LA, Thompson ML, *et al.* The South African Triage Scale (adult version) provides reliable acuity ratings. *Int Emerg Nurs* 2012;20:142–50.
- Twomey M, de Sá A, Wallis LA, *et al.* Inter-rater reliability of the South African Triage Scale: Assessing two different cadres of health care workers in a real time environment. *African J Emerg Med* 2011;1:113–8.
- Gyedu A, Agbedinu K, Dalwai M, *et al.* Triage capabilities of medical trainees in Ghana using the South African triage scale: an opportunity to improve emergency care. *Pan Afr Med J* 2016;24.
- Manos D, Petrie DA, Beveridge RC, *et al.* Inter-observer agreement using the Canadian Emergency Department Triage and Acuity Scale. *CJEM* 2002;4:16–22.
- Beveridge R, Ducharme J, Janes L, *et al.* Reliability of the Canadian emergency department triage and acuity scale: interrater agreement. *Ann Emerg Med* 1999;34:155–9.
- Parenti N, Reggiani ML, Iannone P, *et al.* A systematic review on the validity and reliability of an emergency department triage scale, the Manchester Triage System. *Int J Nurs Stud* 2014;51:1062–9.
- 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.
- Mirhaghi A, Kooshar H, Esmaeili H, *et al.* Outcomes for emergency severity index triage implementation in the emergency department. *J Clin Diagn Res* 2015;9:OC04–OC07.
- Khan A, Mahadevan SV, Dreyfuss A, *et al.* One-two-triage: validation and reliability of a novel triage system for low-resource settings. *Emerg Med J* 2016;33:709–15.