

ORIGINAL ARTICLE

Limitations in validating emergency department triage scales

Michele Twomey, Lee A Wallis, Jonathan E Myers

Emerg Med J 2007;24:477-479. doi: 10.1136/emj.2007.046383

See end of article for authors' affiliations

Correspondence to: Michele Twomey, School of Public Health, University of Cape Town, Cape Town, South Africa; satriage@webmail.co.za

Accepted 29 April 2007

Objective: To examine whether current validation methods of emergency department triage scales actually assess the instrument's validity.**Methods:** Optimal methods of emergency department triage scale validation are examined in developed countries and their application to developing countries is considered.**Results and conclusion:** Numerous limitations are embedded in the process of validating triage scales. Methods of triage scale validation in developed countries may not be appropriate and repeatable in developing countries. Even in developed countries there are problems in conceptualising validation methods. A new consensus building validation approach has been constructed and recommended for a developing country setting. The Delphi method, a consensual validation process, is advanced as a more appropriate alternative for validating triage scales in developing countries.

Emergency department (ED) triage is the process of sorting and filtering patients based on medical priority. It aims to determine a patient's acuity level in order to facilitate timely and effective care before their condition worsens. A patient's acuity level is defined as the urgency for effective care. In the ED triage setting effective care is defined as the provision of an intervention or treatment that reduces the patient's urgency for care or prevents clinical deterioration.¹ If patients receive timely and effective care, triage has achieved its purpose (as seen at point A in fig 1).

This illustration of triage is a highly simplified approach to a complex set of interrelationships. It is acknowledged that additional variables may influence optimal time to care and effectiveness of care significantly (such as variability in triage nurse decisions).

RELIABILITY

The evaluation of a triage tool involves assessing reliability and validity.² *Reliability* refers to the degree to which repeated

assessments of the same patient with a triage instrument will deliver the same acuity level. *Inter-rater* reliability determines whether there is significant variability between different triage officers rating the same patient, and *intra-rater* reliability assesses the variability within a single triage officer re-rating the same patient. Reliability makes no reference to a criterion, and so only illustrates consistency with triage repetition. It shows nothing about its validity (whether it is a reflection of the truth). A measure can therefore be highly reliable without being valid.³

Reliability can be estimated by evaluating different types of agreement. Percentage agreement, the κ coefficient and the weighted κ coefficient are three common ways of measuring agreement between raters,⁴ but these measures can generate quite different values. Measuring only the percentage agreement is not recommended because it does not take into account agreement expected on chance alone.⁵ The κ coefficient considers both percentage agreement between raters and percentage agreement expected by chance; unfortunately, it does not take into account the magnitude of disagreement, which may become significant in ordinal data. As a result, the weighted κ coefficient has become the instrument of choice as it assigns different weights of agreement according to the magnitude of disagreement, and enables more explicit comparisons between different studies.⁴ While the majority of research in triage has focused on *inter-rater* and *intra-rater* reliability, which has its uses, it is of greater importance to determine whether a triage tool is in fact valid. We will therefore be focusing on the validity of a triage tool rather than its reliability.

VALIDITY

Validity refers to the degree with which the measured acuity level reflects the patient's true acuity at the time of triage. The term valid implies that there is some sort of external reference or "gold standard" which by definition has absolute accuracy.³ Studies that aim to see how closely an instrument approximates the truth, test *criterion validity*. Unfortunately it is not possible to measure the truth for patient acuity,⁶ as there are myriad events that can occur from the time that a patient presents to the ED to the time of discharge (including the length of time to initiation of care, the quality of that care, and

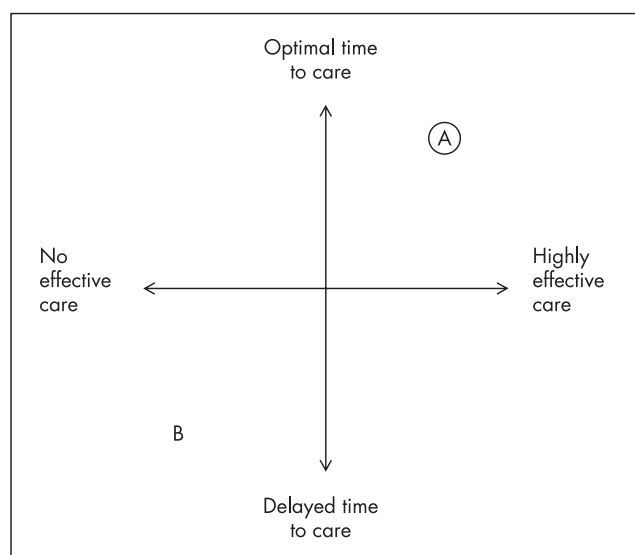


Figure 1 Triage scales are most valuable at (A), where triage facilitates optimal time to highly effective care, and least valuable at (B), where triage facilitates delayed time and ineffective care.

Abbreviations: ATS, Australasian Triage Scale; CTAS, Canadian Triage Acuity Scale; ED, emergency department; ESI, Emergency Severity Index; ETAT, Emergency Triage Assessment and Treatment; MEWS, Modified Early Warning Score

non-medical factors influencing disposal—for example, social factors). As a result, surrogate outcome markers have been used as criteria to assess validity. This has led to other ways of assessing validity for ED triage tools. The two most commonly found in the literature are tests of predictive or consensual validity. These have been approached in a unifying manner by Streiner and Norman, who reconceptualise a variety of notions of validity commonly used in the literature as *construct validity*.⁷

There is a hierarchy of validity testing in which criterion is the best (table 1). Streiner and Norman have shown that unlike the traditional classification of validity, predictive, consensual and other types of validity are all seen as variants of construct validity.⁷ Typically in developed countries, criterion validity methods are used.

We will use Streiner and Norman’s conceptual framework to answer the following questions:

- Do current methods of triage tool validation actually assess the validity and what are the limitations underlying these methods?
- How can these limitations be overcome with special reference to developing countries?

CURRENT METHODS OF TRIAGE TOOL ASSESSMENT AND THEIR LIMITATIONS

A number of different triage systems are used in developed countries. To date, four reliable ordinal ED triage scales have been researched and published: the Australasian Triage Scale (ATS),⁸ the Canadian Triage Acuity Scale (CTAS),⁹ the Emergency Triage Scale (aka Manchester Triage Scale)¹⁰ and the Emergency Severity Index (ESI).¹¹ While there has been some focus on the reliability of triage tools, not much is published on their validity. *Predictive validity* (a type of construct validity) is the most frequently used method of assessing tools.¹² It considers the degree to which the triage acuity level is able to predict true acuity. Particular outcomes, or events with time-ordering, are selected as surrogate markers (such as mortality rates, hospital admission rates, resource utilisation, and length of stay in hospital). There are methodological problems with the use of this type of validity as it does not always answer the core question: “Is the triage instrument able to measure what is supposed to be measured?” In patients it does not measure acuity at the time of assessment (and is inherently confounded by the effectiveness of the health care intervention).

Examples of predictive validity abound in the triage literature, as surrogate outcome markers are practical to measure and are claimed to be closely associated with true acuity.³ This has compelled clinicians and researchers to utilise triage instruments as prediction tools. However, our ability to identify and measure the relationship between patient acuity level and outcome depends not only on the measurement of the surrogate outcome marker and the patient’s acuity level, but also very importantly on confounding variables such as variability in triage nurse decisions, and delayed and ineffective treatment. These may affect the surrogate outcome marker.

Table 1 Traditional validity testing versus Streiner and Norman’s framework

Traditional	Streiner and Norman
Criterion	Criterion
Construct	
Predictive	Construct
Consensual	

HOW CAN THESE LIMITATIONS BE OVERCOME?

A detailed literature review revealed that very little has been published on triage in developing countries. The World Health Organization reports that triage research is not a priority in low-to middle-income countries.¹³ They have accordingly developed the Emergency Triage Assessment and Treatment (ETAT)¹⁴ for application to developing countries. While this subjective system has been successfully implemented in Malawi, countries like India, Brazil and South Africa have sought a more objective triage instrument based on physiology. They have either adopted the triage instrument from a developed country or modified it to their own local context and needs (Patriacia Neto, Quinta D’or Hospital, Rio de Janeiro, May 2007, personal communication). South Africa has adapted the Modified Early Warning Score (MEWS) as the South African Triage Scale after validating it on the local national population.¹⁵ Some areas of Brazil have adopted the CTAS, others the ESI.

During any validity testing an important distinction needs to be made between internal validity (which refers to inferences about the source population), and external validity (whether inferences may be generalised to people outside the source population).¹⁶ A triage tool designed for a developed country may be valid in that context, leading to favourable results that are meaningful and have implications for action. If, however, the same triage tools were applied in a developing country, results may vary due to different resources and skills. Similarly results may vary when applying surrogate markers from developed countries to undertake validity testing in developing countries. This variability may increase the random error in both triage acuity level and outcome category; it would therefore be more appropriate to apply a locally developed tool that is meaningful in the local context (has internal validity), but that may not be applicable in a developed country (lack of external validity).

Whichever tool is used, an assessment of its usefulness in these settings is required. When selecting surrogate outcome markers (such as mortality rates, hospital admission rates, resource utilisation, and length of stay in hospital), it is assumed that there is systematic record keeping, and that the care given is effective. While this may often be the case in developed countries, it is typically not the case in developing countries. Poor record keeping and ineffective care may have significant effects on surrogate outcome markers and patients’ final dispositions. Markers such as these are imperfect measures of patient acuity in the developing world. It is thus important to identify and measure all confounding variables that may be affecting the surrogate outcome marker: given the poor record keeping and lack of efficiency, this is unlikely to be feasible in developing countries.

Delphi methodology

The Delphi method was developed in the 1950s by the RAND Corporation in California, USA.¹⁷ The technique has diversified and is being applied to more mainstream social sciences, in business and, in the last two decades, within the healthcare arena.¹⁸ It is a consensus building technique designed to gain insight into a particular field to enable decision making in areas where published information is inadequate or non-existent.¹⁹ The approach of the Delphi technique is to establish a panel of appropriate experts that have agreed to complete an iterative process on a particular issue, with the key objective being to reach consensus.²⁰ Panellist anonymity is maintained throughout the process and controlled feedback is provided from each iterative round, resulting in a statistical aggregation of the group response.¹⁸

The Delphi method is another form of construct validity that may be useful when assessing triage scales in developing countries. It allows the development of a surrogate “gold standard” determined by specialist panel consensus. The triage

tool's validity may then be tested against this construct of true underlying acuity that is consensually arrived at. There appear to be only very few examples in the world literature that elaborate on the use of this form of construct validity.

Wallis *et al*²¹ used consensus from Delphi methodology to establish triage acuity levels against which to test pre-hospital mass casualty triage tools: such methodology may be used in ED triage tool assessment.

There are several reasons why the Delphi methodology is best suited to assessing ED triage tools in developing countries. The Delphi technique eliminates potential bias due to individual group dynamics and is financially feasible.

Limitations of the Delphi technique are mostly a result of poorly conducted studies rather than fundamental problems. One of the weaknesses cited is that the response rates can be low and often decrease as the rounds progress. However, non-response is typically very low in practice, since most researchers have personally obtained assurance of participation. Similarly attrition tends to be low and the researcher can easily ascertain the cause by talking with the dropouts.²² Selection of the Delphi panel depends on the research question. Problems may arise with a lack of representativeness in that only experts with an interest and involvement will become participants. Another potential weakness of the Delphi as a consensus method is that it overlooks important minority issues because it tries to obtain consensus.²³ However, despite these limitations we believe that the Delphi process is the most appropriate form with which to test the validity of triage tools in the developing world.

CONCLUSION

In developing countries a form of construct validity derived from a consensual process appears to be the most appropriate form of validation of triage tools. This is due to lack of criteria for true acuity, confounding variables that relate to differential health care resources by level of development, and lack of external validity of other triage scales.²⁴ We propose the Delphi method when testing the South African Triage Scale. This is an example of construct validity testing in the developing world.

Authors' affiliations

Michele Twomey, School of Public Health, University of Cape Town, Cape Town, South Africa

Lee A Wallis, Division of Emergency Medicine, University of Cape Town & Stellenbosch University, Capetown, South Africa

Jonathan E Myers, Occupational and Environmental Health Research Unit, University of Cape Town, Capetown, South Africa

Funding: None

Competing interests: None

Contributions: LW had the original idea; MT wrote the first draft; both authors contributed to the final article.

REFERENCES

- 1 **Bindman AB**. Triage in accident and emergency departments. *BMJ* 1995;**311**:404.
- 2 **Rutschmann OT**, Kossovsky M, Geissbühler A, *et al*. Interactive triage simulator revealed important variability in both process and outcome of emergency triage. *J Clin Epid* 2006;**59**:615–21.
- 3 **Lewis RJ**. Reliability and validity: meaning and measurement. Presented at 1999 Annual Meeting of the Society for Academic Emergency Medicine (SAEM) in Boston, Massachusetts. <http://www.ambpeds.org/ReliabilityandValidity.pdf>.
- 4 **Jakobsson U**, Westergren A. Statistical methods for assessing agreement for ordinal data. *Scand J Caring Sci* 2005;**19**:427–31.
- 5 **Fernandes CB**, Groth SJ, Johnson LA, *et al*. A uniform triage scale in emergency medicine. American College of Emergency Physicians, 1999. <http://www.acep.org/NR/rdonlyres/EE2B10F7-51BE-42A8-94B7-779726144017/0/triagescaleip.pdf>.
- 6 **Fernandes CM**, Tanabe P, Gilboy N, *et al*. Five-level triage: a report from the ACEP/ENA five-level triage task force. *J Emerg Nurs* 2005;**31**:39–50.
- 7 **Streiner DL**, Norman GR. "Precision" and "accuracy": Two terms that are neither. *J Clin Epidemiol* 2006;**59**:327–30.
- 8 **Australasian College of Emergency Medicine**. Guidelines for the Implementation of the Australasian Triage Scale in Emergency Departments. Australasian College of Emergency Medicine, 1998.
- 9 **Canadian Association of Emergency Physicians**. Implementation guidelines for the Canadian Emergency Department Triage and Acuity Scale (CTAS), Canadian Association of Emergency Physicians, 1998.
- 10 **Manchester Triage Group**. *Emergency triage*. London: BMJ Publishing Group, 1997.
- 11 **Gilboy N**, Tanabe P, Travers DA, *et al*. The Emergency Severity Index implementation handbook: a five-level triage system. Des Plaines, Illinois: Emergency Nurses Association, 2003.
- 12 **Fan J**, Al Darrab A, Eva K, *et al*. Triage scales in the emergency department: a systematic review. *Ann Emerg Med* 2005;**46**:S41.
- 13 **Razzak, Junaid A, Kellermann, Arthur L**. Emergency medical care in developing countries: is it worthwhile? *Bull World Health Organ* 2002;**80**:900–5.
- 14 **Robertson MA**, Molyneux EM. Triage in the developing world – can it be done? *Arch Dis Child* 2001;**85**:208–13.
- 15 **Gottschalk SB**, Wood D, de Vries 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.
- 16 **Rothman KJ**, Greenland S. *Modern epidemiology*. Philadelphia: Lippincott-Raven Publishers, 1998:115–47.
- 17 **Dalkey N**, Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science* 1963;**9**:458.
- 18 **Jones J**, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;**311**:376–80.
- 19 **Pill J**. The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Econ Plan Sci* 1971;**5**:57–71.
- 20 **Hasson F**, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;**32**:1008–15.
- 21 **Wallis LA**, Carley S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. *Emerg Med J* 2006;**23**:291–5.
- 22 **Okoli C**, Pawlowski S. The Delphi method as a research tool: an example, design considerations and applications. *Information and Management* 2004;**42**:15–29.
- 23 **Van Teijlingen E**, Pitchforth E, Bishop C, *et al*. Delphi method and nominal group techniques in family planning and reproductive health research. *J Fam Plann Reprod Health Care* 2006;**32**:249252.
- 24 **Fernandes CB**, Groth SJ, Johnson LA, *et al*. A uniform triage scale in emergency medicine. American College of Emergency Physicians, 1999; <http://www.acep.org/NR/rdonlyres/EE2B10F7-51BE-42A8-94B7-779726144017/0/triagescaleip.pdf>.