SHORT REPORT

Are bedside features of shock reproducible between different observers?

H Otieno, E Were, I Ahmed, E Charo, A Brent, K Maitland

Arch Dis Child 2004;**89**:977–979. doi: 10.1136/adc.2003.043901

Shock is often under-reported in children attending hospitals in developing countries. Readily obtainable features of shock (capillary refill time, temperature gradient, pulse volume, and signs of dehydration) are widely used to help prioritise management in the emergency assessment of critically ill or injured children. However, data are lacking on their validity, including, importantly, reproducibility between observers. Agreement of these signs was examined in 100 consecutive children admitted to a paediatric ward on the coast of Kenya. After an initial training of clinical sign recognition, there was moderate agreement for most features of cardiovascular compromise (delayed capillary refill ⩾4 s, κ = 0.49; and weak pulse volume, κ = 0.4) and only substantial agreement for temperature gradient (κ = 0.62). For hydration status, only in the assessment of skin turgor was there a moderate level of agreement (κ = 0.55). Capillary refill times and assessment of pulse volume recommended by the recent American consensus guidelines achieved only a "low" moderate to poor interrater agreement, questioning the reliability of such parameters.

The utility of any clinical sign is a function of both its ability to detect or predict a particular condition, and its reproducibility between different observers. In emergency paediatrics an informed consensus has been reached on the clinical parameters defining shock.[1] These include the presence of tachycardia, hypothermia, or hyperthermia, decreased mental status, signs of impaired perfusion (flash capillary refill and bounding pulse (warm shock) or delayed capillary refill time >2 s, cold peripheries and decreased pulse volume (cold shock), and urine output <1 ml/kg/h in a child with suspected infection.[1] Readily identifiable features of shock (capillary refill time, temperature gradient, pulse volume, and signs of dehydration) have been incorporated into advanced life support guidelines as part of the rapid, structured cardiopulmonary assessment of critically ill or injured children.[2 3] As such they are used to prioritise initial management. However, information supporting the reliability of these clinical signs to define shock is lacking, in particular data examining the robustness of these signs between different observers. The interpretation of these features therefore requires more detailed assessment, since their presence is often the sole basis on which emergency interventions are initiated. The current study sought to examine prospectively the inter-observer reproducibility of bedside clinical features of shock. It did not, however, seek to validate the ability of any sign to define shock.

PATIENTS AND METHODS

Study site and population

The study was conducted at Kilifi District Hospital (KDH) on the coast of Kenya. Detailed descriptions of the facilities and routine clinical assessment of children admitted to KDH have been published previously.[4] During weekdays from 1 June to 1 July 2003, four clinicians independently assessed consecutive morning admissions to the general paediatric ward. Each clinician had 2–3 years postgraduate clinical experience. All assessments were conducted within one hour of each other. The study clinicians were unaware of each child's clinical details and admission diagnosis, and categorical definitions and standard methods for eliciting each clinical feature were agreed initially (see table 1). Capillary refill time (CRT) was assessed by applying pressure to a finger pulp for three seconds and counting the time required for the blanched finger to fully re-perfuse. Temperature gradient was assessed by running the back of the palm of the hand down the limb and reported for both the upper and lower limbs. The radial pulse was used to assess pulse volume. Reduced skin turgor was assessed by pinching a longitudinal skin fold midway between the umbilicus and the flank (as recommended by the WHO Integrated Management of Childhood Illness (IMCI) guidelines) and observing whether the skin pinch goes back slowly.[5] A WHO IMCI training video was used as an aid to training in both decreased skin turgor and in the recognition of sunken eyes. For the examination of dry mucous membranes, each doctor made a subjective assessment of whether mucous membranes were "drier than normal". In the context of a busy admission suite on the general paediatric ward it was not possible to assess urinary output.

Data analysis

Cohen's kappa statistic (κ) was used as a measure of agreement,[6] which represents the chance corrected proportional agreement. Agreement is scaled between zero and one, where 0 indicates no agreement (that is, no better than would be observed by chance) and 1 signifies perfect agreement (see table 3). The measure was calculated for the four raters (or observers), and a uniform number of outcomes for each feature assessed (STATA, version 8). Results are given in kappa values (κ); 95% confidence intervals were calculated from the standard error of κ.

RESULTS

One hundred consecutive paediatric admissions were assessed independently by each of the four clinicians. The study group age ranged from 2 days to 10 years 11 months. Presenting complaints included fever (n = 78), cough (n = 43), respiratory distress (n = 25), diarrhoea and/or vomiting (n = 26), and convulsions (n = 25). Many had poor nutritional status: undernutrition (WAZ score −2 to −3 SD)

Abbreviations: APLS, acute paediatric life support; CRT, capillary refill time; IMCI, integrated management of childhood illness; κ, kappa-statistic; KDH, Kilifi District Hospital; LRTI, lower respiratory tract infection; SD, standard deviation; WAZ, weight for age Z score; WHO, World Health Organisation

Table 1 Categorical definitions of the features assessed by the clinicians

Feature	Values
Capillary refill time (seconds)	1, 2 ,3, 4 or more
Temperature gradient	Yes, no
Pulse volume	Weak (or absent), normal, strong/bounding
Decreased skin turgor	Yes, no
Sunken eyes	Yes, no
Dry mucous membranes	Yes, no

and severe malnutrition (WAZ score <−3SD, plus visible severe wasting) were present in 22% and 18% respectively, and seven children had oedematous malnutrition (kwashiorkor). Final diagnoses are shown in table 2, which were typical of the usual diagnostic spectrum in children admitted to KDH. Three children subsequently died (a 1 month old infant with gastroenteritis and LRTI, and two children with malnutrition, aged 16 and 32 months, respectively), one child absconded (LRTI), and two were transferred for surgical intervention.

Table 3 shows the level of agreement for all features assessed by the study team. Overall agreement for CRT was moderate ($\kappa = 0.42$), and was better for normal values ($\leqslant 1$ second) ($\kappa = 0.48$) and clearly abnormal values ($\geqslant 4$ seconds) ($\kappa = 0.49$). There was moderate to substantial agreement between observers for temperature gradient, being slightly better for the lower limb ($\kappa = 0.62$) than the upper limb ($\kappa = 0.57$). There was moderate agreement in the assessment of weak pulse volume ($\kappa = 0.40$); however, there was little to no agreement for bounding pulse volume ($\kappa = -0.01$). In the assessment of hydration status the level of agreement was substantially better for a decreased skin turgor ($\kappa = 0.55$) than either sunken eyes or dry mucous membranes, for which agreement was only fair (0.34 and 0.39 respectively). There was no significant difference in these findings after stratification for the presence or absence of malnutrition.

Table 2 Final primary diagnosis of children

Condition	Frequency
Abdominal obstruction	1
Anaemia	7
Asthma	1
Bronchiolitis	1
Burkitt's lymphoma	1
Cellulitis or abscess	3
Encephalopathy—unknown	1
Epilepsy	1
Febrile convulsions	2
Fever of unknown origin	3
Gastroenteritis	15
Hydrocephalus	1
LRTI	11
Malaria	24
Malnutrition	10
Meningitis	1
Neonatal jaundice	3
Neonatal sepsis	5
Other skin disease	1
Rash (mumps)	1
Septicaemia/sepsis	1
Sickle cell disease (plus anaemia)	1
Surgical abdomen	1
Testicular torsion	1
Trauma/fracture	1
URTI (with febrile convulsion)	2
Total	100

Table 3 Inter-observer agreement between four clinicians in the signs of shock

Feature	Kappa (κ)	95% CI
Capillary refill time		
1	0.48	0.34 to 0.62
2	0.37	0.25 to 0.49
3	0.35	0.23 to 0.47
4	0.49	0.35 to 0.63
Combined	0.42	0.29 to 0.55
Temperature gradient		
Upper limb	0.57	0.42 to 0.72
Lower limb	0.62	0.47 to 0.77
Pulse volume		
Weak	0.40	0.28 to 0.52
Normal	0.30	0.19 to 0.41
Strong/bounding	−0.01	
Dehydration		
Dry mucous membranes	0.39	0.27 to 0.51
Decreased skin turgor	0.55	0.40 to 0.70
Sunken eyes	0.34	0.23 to 0.45

Interpretation of kappa statistic:[16]
Below 0, poor agreement
0–0.2, slight
0.2–0.4, fair
0.41–0.6, moderate
0.61–0.8, substantial
0.81–1.0, almost perfect agreement.

DISCUSSION

We have examined the between-observer agreement in medically qualified workers for the assessment of clinical signs of cardiovascular compromise, that are currently recommended by current guidelines for the assessment of critically ill and injured children.[2 3] We found that agreement was substantial for the assessment of temperature gradient (in the lower limbs): high moderate agreement for upper limb temperature gradient and low moderate agreement for the assessment of weak pulse volume. Agreement for capillary refill time was greatest for normal values (1 second) and significantly delayed refill time ($\geqslant 4$ seconds), but only fair for other values. In the assessment of hydration status the best agreement between observers was for reduced skin turgor. Sunken eyes and decreased skin turgor are also non-acute features of severe malnutrition, and in the presence of severe malnutrition the assessment of hydration status is difficult.[7 8] In this study however, the interrater agreement for signs of dehydration was similar in children both with and without anthropometric features of severe malnutrition. The more important question would have been, with regard to sign recognition in the malnourished group, an appraisal of the validity of the assessments, but this was beyond the scope of this study.

Basic and advanced paediatric life support (APLS) and American College of Critical Care guidelines recommend the use of readily obtainable features of cardiovascular compromise (delayed capillary refill time, poor pulse volume, and temperature gradient) to determine initial management, in particular the use of resuscitation fluids.[1–3] Furthermore, both clinical and biochemical features of compensated shock (hypovolaemia) are also recommended to define the severe inflammatory response syndrome (SIRS).[9] In settings where resources allow, invasive monitoring and quantification of the biochemical derangements (blood gas analysis, urea, creatinine, and lactate estimation) provide supportive evidence for hypovolaemia. However, in resource-poor areas of the world clinical assessment is often the sole determinant of a child's initial management. It is important to provide information on just how reliable bedside features of shock really are. Capillary refill times and assessment of pulse volume recommended by the recent American consensus guidelines[1] achieved only a low moderate to poor interrater agreement, questioning the reliability of such parameters.

Nolan and colleagues have shown that initial triage and emergency treatment are poor in most district hospitals and in many teaching hospitals in developing countries. Inadequate staff training and the absence of guidelines for standard management were identified as contributing factors.[10] We found, despite a period of training in sign recognition, that inter-observer agreement over a wide range of clinical conditions was moderate for most features assessed and only substantial and high moderate for the assessment of temperature gradient, indicating that, for clinical signs with poor interrater agreement, either training was inadequate or more plausibly that these features are too subjective to be amenable to accurate definition.

Emergency triage assessment and treatment (ETAT) guidelines were designed with the aid of the WHO for use in developing countries.[11] The emergency signs used to define shock are delayed capillary refill time (⩾4 s), cold hands, and weak, fast pulse, and indicate the urgent need for volume expansion. Evaluation in 3837 Brazilian children attending the emergency department, identified shock in only four children (0.1%), all of whom died.[12] In a separate study in 2281 Malawian children attending casualty, delayed capillary refill time (>3 s) and cold hands were ascertained by both a doctor and nurse in only five and 13 patients respectively.[13] These data suggest that these criteria are insensitive, since in both studies shock was rarely diagnosed (<0.5%), although the case fatality in those defined as shocked approached 100%. In our hospital on the coast of Kenya, clinical assessment for shock has been part of our routine admission practice for a number of years. A review of data from children admitted in the year 2002 found delayed CRT (>3 s) was diagnosed in 8% (case fatality = 24%), a temperature gradient described in 4% (case fatality = 12%), and weak pulse volume in 3% (case fatality = 23%). Case fatality in children without any of these features was 2.6%; suggesting that even when routinely used by medical personnel (with varying degrees of clinical experience), these signs have some value in being able to identify a group with poor prognosis in whom volume resuscitation should be considered.[14] Furthermore, these data also suggest that shock is much more common among routine admissions to the general paediatric ward than previous studies in the developing world have indicated.

Published literature regarding the inter-observer variation in triage signs is limited,[13 15] and is especially lacking for clinical signs of shock. In the current climate of clinical audit, it is surprising that inter-observer variation studies have not been more widely used to validate clinical practice. A general appreciation of the need for training of basic sign recognition is fundamental to improving triage and initial management. In developing countries a culture of early recognition of compensated shock and the institution of simple, potentially life saving treatments may improve outcome. This study provides reasonable evidence of the reproducibility of some of the bedside clinical features currently used in the assessment of shock, in particular, for lower limb temperature gradient and delayed capillary refill ⩾4 seconds in the assessment of shock and for skin turgor in the assessment of hydration status. Poor inter-observer agreement in the assessment of pulse volume and other features of dehydration questions the reliability of these signs. Further studies should be conducted to examine the validity of these signs and other less subjective features of shock (such as tachycardia and tachypnoea), but these are unlikely to be within the scope of a busy tropical paediatric unit.

ACKNOWLEDGEMENTS

The authors would like to thank the medical and nursing staff of Kilifi District Hospital and KEMRI/Wellcome Trust, Kilifi, and Charles Newton for reading an early version of the paper. K Maitland is supported by the Wellcome Trust, UK. This manuscript is published with the permission of the Director of KEMRI.

Authors' affiliations

H Otieno, E Were, I Ahmed, E Charo, A Brent, K Maitland, The Centre for Geographic Medicine Research, Coast, KEMRI, Kenya, PO Box 230, Kilifi, Kenya

Correspondence to: Dr K Maitland, KEMRI/Wellcome Trust Unit, PO Box 230, Kilifi, Kenya; kmaitland@kilifi.mimcom.net

Accepted 5 February 2004

REFERENCES

1 **Carcillo JA**, Fields AI. Clinical practice parameters for hemodynamic support of pediatric and neonatal patients in septic shock. *Crit Care Med* 2002;**30**:1365–78.
2 *Advanced paediatric life support: the practical approach*, 1st edn. London: BMJ Publishing Group, 1993.
3 **American Heart Association**. *Pediatric Advanced Life Support 1997–1999. Emergency cardiovascular care programs.* American Heart Association, 1997.
4 **Berkley JA**, Ross A, Mwangi I, *et al*. Prognostic indicators of early and late death in children admitted to district hospital in Kenya: cohort study. *BMJ* 2002;**326**:361–7.
5 **Gove S**. Integrated management of childhood illness by outpatient health workers: technical basis and overview. The WHO Working Group on Guidelines for Integrated Management of the Sick Child. *Bull World Health Organ* 1997;**75**(suppl 1):7–24.
6 **Posner KL**, Sampson PD, Caplan RA, *et al*. Measuring interrater reliability among multiple raters: an example of methods for nominal data. *Stat Med* 1990;**9**:1103–15.
7 **Waterlow J**. *Protein-energy malnutrition*, 2nd edn. Edward Arnold, 1992.
8 **WHO**. *Management of the child with a serious infection or severe malnutrition*. Geneva: World Health Organisation, 2000.
9 **Saez-Llorens X**, Vargas S, Guerra F, *et al*. Application of new sepsis definitions to evaluate outcome of pediatric patients with severe systemic infections. *Pediatr Infect Dis J* 1995;**14**:557–61.
10 **Nolan T**, Angos P, Cunha ALA, *et al*. Quality of hospital care for seriously ill children in less-developed countries. *Lancet* 2001;**357**:106–10.
11 **Gove S**, Tamburlini G, Molyneux E, *et al*. Development and technical basis of simplified guidelines for emergency triage assessment and treatment in developing countries. WHO Integrated Management of Childhood Illness (IMCI) Referral Care Project. *Arch Dis Child* 1999;**81**:473–7.
12 **Tamburlini G**, Di Mario S, Maggi R, *et al*. Evaluation of guidelines for emergency triage assessment and treatment in developing countries. *Arch Dis Child* 1999;**81**:478–82.
13 **Robertson MMEM**. Triage in the developing world—can it be done? *Arch Dis Child* 2001;**85**:208–13.
14 **Maitland K**, Newton C, English M. Intravenous fluids for seriously ill children. *Lancet* 2004;**363**:242–3.
15 **Wuerz R**, Fernandez C, Alarcon J. Inconsistency of emergency department triage. *Ann Emerg Med* 1998;**32**:431–5.
16 **Landis J**, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.