BMJ Open

Interrater reliability between nurses for a new paediatric triage system based primarily on vital parameters: the Paediatric Triage Instrument (PETI)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012748
Article Type:	Research
Date Submitted by the Author:	20-May-2016
Complete List of Authors:	Karjala, Jaana; Department of pediatrics, Mälarsjukhuset Hospital; Centre for Clinical research Sörmland, Uppsala university Eriksson, Staffan; Centre for clinical research Sörmland, Uppsala University; Department of Neuroscience, Physiotherapy, Uppsala University
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Emergency medicine
Keywords:	PAEDIATRICS, ACCIDENT & EMERGENCY MEDICINE, HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Interrater reliability between nurses for a new paediatric triage system based primarily on vital parameters: the Paediatric Triage Instrument (PETI)

Jaana Karjala[1,2], Staffan Eriksson[2, 3, 4]

[1]Department of Pediatrics, Mälarsjukhuset Hospital, Eskilstuna, Sweden. E-mail: jaanakarjala@gmail.com

[2]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

[3]Department of Neuroscience, Physiotherapy, Uppsala University, Uppsala, Sweden

[4]Department of Community Medicine and Rehabilitation, Physiotherapy, Umeå University, Umeå, Sweden

Corresponding author: Staffan Eriksson, Centre for Clinical Research Sörmland, Uppsala University, 631 88 Eskilstuna, Sweden. Telephone number: +46 16 105251, +46 73 949 99 33. E-mail: staffan.eriksson@germed.umu.se

Key words: paediatrics, triage, reliability, and emergency department.

Word count: 3547 words including competing interests, funding, contributors and ethics approval. One figure, 3 tables, 27 references, and one supplementary file for online only publication.

ABSTRACT

Introduction The major paediatric triage systems are primarily based on flow charts involving signs and symptoms for orientation and subjective estimates of the patient's condition. In contrast, the four level Paediatric Triage Instrument (PETI) is primarily based on vital parameters and was developed exclusively for paediatric triage in patients with medical complaints. The aim of this study was to assess the interrater reliability of this triage system in children when used by nurses.

Methods A design was employed in which triage was performed simultaneously and independently by a research nurse and an emergency department nurse using the PETI. All patients aged ≤12 years who presented at the emergency department (ED) with a medical complaint were considered eligible for participation.

Results The 89 participants exhibited a median age of 2 years and were triaged by 27 different nurses. The interrater reliability between nurses calculated with the quadratic-weighted kappa, was 0.78 (95% CI 0.67–0.89); the linear-weighted kappa was 0.67 (95% CI 0.56–0.80); and the unweighted kappa was 0.59 (95% CI 0.44–0.73). For the patients < 1 year old, 1–3 years old, and > 3 years old, the quadratic-weighted kappa values were 0.67 (CI 0.39–0.94), 0.86 (CI 0.75–0.97) and 0.73 (CI 0.49–0.97), respectively. The median triage duration was 6 minutes.

Conclusions The PETI exhibited substantial reliability when used in children ≤ 12 years old and almost perfect reliability among 1–3-year-old children. Moreover, rapid application of the PETI was demonstrated.

INTRODUCTION

Since the early 1990s, there has been a dramatic increase in the number of emergency department (ED) visits.[1, 2] In addition to the increase in emergency visits, several other circumstances have contributed to the overcrowding of EDs, including an inadequate inpatient capacity, the increasing complexity of paediatric patients, the lack of medical staff, and the lack of easy access to primary care.[1] With overcrowding comes greater risks of medical errors and adverse events.[1, 2] The overcrowding of EDs has made triage systems important, and several such systems, such as the Australasian Triage Scale (ATS), the Manchester Triage System (MTS), the Canadian Triage and Acuity Scale (CTAS) and the Emergency Severity Index (ESI) emerged in the 1990s. These four systems are the most established triage systems for adults, and they are also used for paediatric patient populations with some adaptations.[3-9]

The triage of children in an ED setting offers several challenges that differ from adult triage. First, infants and smaller children depend almost entirely on their parents and medical professionals for correct judgements of their status. Second, substantial physiologic variations and immaturity of organ development make small children more susceptible to sudden deterioration, which necessitates the continuous reassessment of children.[10]

Some of the currently used paediatric triage systems have reached a substantial level of interrater reliability, although there is still room for improvement. In simultaneous "live" triage conducted in an independent manner, weighted kappa values of 0.57, 0.65 and 0.74 have been reported for ESI version 4, MTS and CTAS, respectively.[5, 7, 9] One factor that may contribute to errors in triage is that triage decisions are based to a large extent on informed but subjective estimates of the patient's presenting condition, such as estimates of pain and future resource utilization in the ATS and ESI, respectively.[11, 12] Another negative factor may be the complexities of triage systems with large numbers of different presenting complaints.[8, 13, 14] To determine acuity levels, these complaints are accompanied by general and complaint-specific discriminating questions in the MTS and sets of general and complaint-specific criteria in the CTAS.

In contrast to the major triage systems, the Paediatric Triage Instrument (PETI) relies primarily on measurements of vital parameters (VPs) that are acquired irrespective of the presenting complaints**.** The use of VPs is accepted as important in triage because VPs offer objective measurements on which decisions can be based, and such objective measurements are expected to be especially important in children.[8, 14] Moreover, a triage system based on VPs should be easy and quick to use. An additional possible advantage is increased control of the deterioration of patients because a base-line is established during the first triage, and a rapidly applied triage system makes continuous reassessments more achievable.

The PETI is a four-level triage system that is exclusively applied for paediatric triage and is based primarily on the VPs of patients with medical complaints. In creating this system, the main focus was placed on achieving an initial assessment that is quick and objective.

The aim of this study was to assess the interrater reliability of the PETI in children with medical complaints when used by nurses. The secondary aims were to assess the interrater reliability of the PETI for three different age groups and to assess the duration of the triage procedure associated with the PETI.

METHODS

Study design

This study of interrater reliability applied a design in which each patient was simultaneously and independently triaged by a research nurse and an ED nurse who were blinded to each other's collection of the data and triage assignments. The participants were included prospectively and consecutively.

Study setting and population

The study was conducted at a county hospital in the centre of Sweden. The department of paediatrics provides care for a population of 60,000 individuals aged 0-18 years with a rich ethnic diversity. The ED at the hospital receives 45,000 patients-visits annually, and in 2011 13% of these visits were made by paediatric patients 12 years and younger with medical complaints.

All patients aged 12 years or younger who presented to the ED with a medical complaint were considered eligible for inclusion in the study**.** Only children between the ages of 0 to 12 years were included because a different triage system has been introduced in the ED for children older than 12 years. The number of participants was decided upon based on the pre-planned time frame of data collection. Written informed consent was obtained from the parents.

The triage system: PETI

The PETI is a four-level triage system primarily based on measurements of the following five VPs: respiratory rate, heart rate, capillary saturations, capillary refill time, and core temperature (appendix 1). The measurement of each of these VPs is compared with an age-specific reference interval. Depending on the degree of deviation, the VP is assigned 1, 2 or 4 points. The final acuity level is given by the sum of the points assigned to each of the five VPs. Summed scores of 0-1, 2-5, 6-9 and ≥10 corresponds to acuity levels of 'non-urgent' (green), 'urgent' (yellow), 'very urgent' (orange) and 'emergent' (red), respectively. Hence, to limit over-triage, a minimum of 2 points is necessary for triage into the 'urgent' acuity level and a minimum of 6 points is necessary for triage into the 'very urgent' acuity level. In addition, to emphasize severe cases, extra weight is added to large deviations in the VPs (assigned 4 vs. 2 points). The normal reference intervals for the VPs of respiratory rate and heart rate were set according to the Advanced Paediatric Life Support (APLS) system.[10] The normal reference value for the capillary refill time was adjusted based on the APLS value with the intention of increasing reliability. The normal reference intervals for saturation and temperature were set according to established experience. The reference intervals for deviations corresponding to 4 points for the VPs of temperature, capillary saturations, heart rate, and respiratory rate were set according to the cut-off values for danger zone vitals in the ESI along with clinical experience.[15] The cut-off value for deviations in capillary refill time corresponding to 4 points was set according to the APLS.[10] The reference values for deviations corresponding to 1-2 points were evenly distributed between normal and 4 points.

Some signs and symptoms included in the PETI are related to the airway and neurology and were selected from the ABCDE- model, including the alert, voice-, pain-, and unresponsive scale (AVPU-scale), which individually creates a "force majeure" that complements triage based on VPs (appendix 1).[10] Triage based on a "force majeure" is independent from triage based on VPs, and the patient is assigned the highest acuity level between these two methods.

These signs and symptoms are assessed prior to or during the collection of VP data. The sign and symptom of mild recession results in assignment of the patient to the 'urgent' acuity level. Any of the following signs and symptoms result in assignment of the patient to the 'very urgent' acuity level: compromised airway, severe recession, a sloppy or irritable infant, or assessment of the child as voice responsive. Any of the following signs and symptoms result in assignment of the patient to the 'emergent' acuity level: airway obstruction, stridor, convulsions, or assessment of the child as either pain responsive or unresponsive.

The development of the PETI was influenced both by the major triage systems and, more importantly, by paediatric early warning systems, which rely heavily on VPs.[16, 17] During the development of the PETI, feedback was given by groups of paediatricians and other emergency staff.

Data collection

The ED nurses were trained in use of the PETI when the system was introduced at the ED one year prior to the study. This training was implemented via a two-hour lecture and through the opportunity to ask questions for 30 minutes the day the instrument was introduced, or via e-mail, or when the first author was serving at the ED. The research nurse had no previous experience with the PETI and was trained in the use of the system through two one-hour training sessions prior to the study. The research nurse performed 29 shifts of 6 hours each to recruit and triage patients for the study. All but one shift lasted from 4 p.m. to 10 p.m. on normal weekdays. The ED nurses who were working the same 29 shifts during which the research nurse was at the ED participated in the study. According to the established routine in the ED, all triage of children should be performed by an ED nurse.

Triage was performed simultaneously by an ED nurse and the research nurse and included measurements of five VPs and assessments of signs and symptoms related to a "force majeure". The measurements used to calculate the acuity levels of the PETI were performed via the application of two separate sets of instruments. The nurses concealed their data collection from each other by distancing themselves in the room, with the research nurse angling the instrument in use to shield it from the ED nurse. Both the ED nurse and the research nurse calculated acuity levels blindly and separately in different rooms, or separated by distance when in the same room. They were informed not to discuss their data collection or the assignment of acuity levels. Only the ED nurse's triage results were used in patient care. The characteristics of the study participants were documented by the research nurse.

Statistical analysis

Interrater reliability was calculated for the whole group (primary analysis) and for the following post hoc subgroups: <1-, 1-3-, and 4-12 year-olds. The choice of subgroups was based on the purposes of analysing a group of patients <1 year old, in whom difficulties in triage have previously been reported, and creating groups with a sufficient number of participants for the analyses.[7] The primary test of interrater reliability that was calculated for the primary and subgroup analyses was Cohen's kappa with quadratic weights. The quadratic-weighted kappa was chosen because it accounts for the degree of disagreement and the severity of disagreement at higher acuity levels.[18] Additionally, to enable comparison with other studies, Cohen's kappa with linear weights or no weights was also calculated for the whole group. The kappa values were interpreted according to the following categories: <0.40 poor-fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81-1.0 almost perfect.[19] The duration of triage from the beginning of the collection of the triage data to the assignment

of the acuity level was determined for the research nurse. The kappa values and 95% confidence intervals (CIs) were calculated with MedCalc® 12.4.[18]

RESULTS

Data collection was performed from November 3, 2011, to January 11, 2012. Twenty-seven ED nurses participated in the study, six of whom began their employment at the ED after training on the PETI took place and were trained solely by their colleagues while working. The median amount of experience in emergency medicine was 4 years (interquartile range [IQR] 2–15) for the ED nurses and 1.5 years for the research nurse.

The ED nurses triaged a median of 2 participants each (IQR 1–5). One hundred and four patients agreed to participate in the study, fifteen of whom were excluded, and 89 participants were included in the analysis (figure 1). The median age of the included patients was 2 years (IQR 0–11) and 48% were girls (table 1). Overall the characteristics of the study participants corresponded rather well to the characteristics of the patient population (table 1). In nine of the 89 participants, acuity levels were assigned by "force majeure". The blindness and independency of the triage procedure between the ED and research nurses was preserved for 75 of 89 participants (84%). The reasons for the failure to preserve blindness included use of the same measurement for temperature due to parental discomfort (n=11) and the need for acute medical procedures in the emergency room (the 3 emergent participants).

Table 1 Characteristics of the participants and the patient population

	Study participants, n=89*	All patients, n=651†
Gender female, n (%)	43 (48)	286 (44)
Gender male, n (%)	46 (52)	365 (56)
Age		
<1years, n (%)	29 (33)	169 (26)
1-3 years, n (%)	33 (37)	306 (47)
4-12 years, n (%)	27 (30)	176 (27)
Chief complaints		
Asthma and allergy, n (%)	10 (11)	55 (9)
Fever, n (%)	4 (5)	24 (4)
GI- and urinary tract, n (%)	20 (23)	96 (15)
Neurological, n (%)	7 (8)	10 (2)
Observation, n (%)	2 (2)	32 (5)
Respiratory tract, n (%)	31 (35)	280 (43)
Other, n (%)	15 (17)	154 (24)

*Participants with a medical complaint, non-surgical, non-orthopaedic
†All patients 0-12 years old; period 3rd November 2010 to 11th January 2011 (i.e. the corresponding period a year prior to the study)
‡Interquartile range

The agreement in the acuity level of the PETI between the research nurse and the ED nurses was 73% (table 2). There was no evident systematic disagreement, as either the research nurse or the ED nurse triaged a participant to a higher acuity level than the other nurse on approximately the same number of occasions: 11 (7 + 3 + 1) and 13 (7 + 6) occasions, respectively (table 2). The agreement by age was 76%, 76% and 67% for participants with ages of <1-, 1–3-, and 4–12 years, respectively (table 3). The mean (± SD) duration of the triage procedure was 6 minutes and 6 seconds ± 135 seconds (n=81).

Table 2 Agreement of acuity levels between the research nurse and the ED nurses, n = 89

ED nurse	Research nurse				
	Non-urgent	Urgent	Very urgent	Emergent	Total
Non-urgent (Green)	29*	7	1	0	37
Urgent (Yellow)	7	22*	3	0	32
Very urgent (Orange)	0	6	11*	0	17
Emergent (Red)	0	0	0	3*	3
Total	36	35	15	3	89

*Cases showing agreement between the ED nurse and the research nurse

Table 3 Agreement by age, n = 89 participants

	< 1 yr	1-3 yr	4-12 yr	Total
Agreement	22	25	18	65
Disagreement by one level	6	8	9	23
Disagreement by two levels	1	0	0	1
Total	29	33	27	89

The interrater reliability values for the nurses were 0.78 (95% CI 0.67–0.89) based on the quadratic-weighted kappa, 0.67 (95% CI 0.56–0.80) based on the linear-weighted kappa, and 0.59 (95% CI 0.44–0.73) based on the unweighted kappa. (The corresponding kappa values, including cases with unauthorized triage decisions by nurse aids, were 0.78, 0.68 and 0.58, n=94).

The quadratic-weighted kappa (95% CI) values were 0.67 (0.39–0.94), 0.86 (0.75–0.97), and 0.73 (0.49–0.97) for patients with ages of <1, 1–3, and 4–12 years, respectively.

DISCUSSION

This study demonstrated substantial interrater reliability of the new PETI triage system for paediatric patients ≤ 12 years old. Additionally, the time required for the completion of the PETI was quite short. The PETI therefore exhibited promise, particularly considering that this study was conducted in a clinical setting in which the ED nurses received no extra training or practice in the triage of case scenarios prior to the study of live triage, which is common in other studies.[3, 7, 9]

The level of reliability observed in our study for the PETI triage instrument is comparable to the best kappa values for simultaneous live triage that have previously been published. Quadratic-weighted kappa values of 0.74 and 0.65 have been reported for the CTAS and MTS, respectively, whereas an unspecified-weight kappa for ESI version 4 was reported as 0.57.[5, 7, 9] Some studies have obtained higher kappa values in the range of 0.8–0.9 for ESI versions 3 and 4 and the MTS.[3, 4, 6, 7, 9] In two of these studies, the authors used a similar design with “live” triage and a design that approached “live” triage, whereas triage of paper

case scenarios was performed in the other studies. The study of "live" triage and the study approaching "live" triage exhibited drawbacks including a limited sample size and dependency of the measurements used for the calculation of triage by the nurses, which makes the interpretation of their findings questionable.[3, 6, 20] The other studies, reporting higher kappa values, employed paper case scenarios instead of "live" triage of patients.[4, 7, 9] One could argue that case scenarios do not reflect the real clinical setting in which the interactions between the nurse, patient, environment and triage system may contribute to mistriage.[13, 20] Indeed, in studies in which both paediatric paper case scenarios and "live" paediatric patients were triaged within the same study, the kappa values were approximately 0.1–0.2 units higher for case scenarios than for "live" triage.[3, 7, 9] In contrast, a study that compared "live" triage with case scenarios based on the use of the CTAS in a mixed population of adults and children found that the kappa value was higher for the "live" use of the triage algorithm.[21] However, the use of a mixed population in this study makes the comparison of results between studies difficult.

The PETI exhibited a tendency towards showing the best reliability in children aged 1-3 years. A possible explanation for this finding is that the VPs provided a clearly defined framework for the triage of children who lack the ability to efficiently communicate. It has previously been demonstrated that complementing subjective triage decisions with VP data often results in changes in triage decisions in children ≤2 years old and in children whose parents have communications difficulties.[22] The PETI exhibited a tendency towards inferior reliability for children <1 year old, which agrees with previous results for the ESI.[7] In general, triage in infants is particularly difficult because the severity of illness is expressed in multiple and subtle manners and can change rapidly.[10] However, this observation should be regarded with caution because it stemmed from a discrepancy of two levels in a singly participant.

As assessment using the PETI was shown to be rapid, this tool will facilitate retriage and thereby facilitate the control of patient deterioration. It will also potentially decrease the strain on staff and contribute to resource effectiveness.

It has previously been shown that paediatric triage systems that rely to a large extent on VPs are prone to over-triage (low sensitivity).[23] However, in developing the PETI, the risk of over-triage was compensated through the levels in the scoring system, such that a minimum of 2 points was required for triage into the `urgent` acuity level, and a minimum of 6 points was required for triage into the `very urgent` acuity level. As this was not a validation study, there were no available data on the participants' "true" acuity levels, and it is not possible to answer the question of whether triage with the PETI is prone to over- or under-triage. Nevertheless, it is notable that approximately 40% of the participants were triaged to each of the 2 lowest acuity levels (non-urgent and urgent) (table 2).

Improvements in the measures employed in the PETI should likely focus on the VPs of respiratory rate and capillary refill because these VPs rely to a large extent on estimates and skill. Regarding triage based on "force majeure", the relative position between stridor and severe recession should be considered when proposing improvements. Additionally, in the table illustrating the reference values for the VPs (appendix 1), detected errors should be revised, including gaps in the reference intervals for heart rate, respiratory rate and temperature. Minor revisions of the lay-out have already been incorporated.

This study has some limitations. First, comparisons with studies on commonly used paediatric triage systems are difficult because the PETI is a four-level system, whereas the others are five-level systems. However, it has been suggested that kappa values in general, and quadratically weighted kappa values in particular, increase as the number of categories in a system is increased.[24, 25] Second, the sample size of this study was relatively small, which resulted in wide CIs and uncertainty in some of the results. For instance, the linear-weighted kappa for the whole group exhibited a 95% CI that stretched below the lower limit of the substantial category. Third, even though the characteristics of the participants in this study resembled those of the patient population, there are some issues regarding the generalizability of the results. The small sample size makes it likely that not all possible presenting complaints of the population were covered in the triage of the participants. In addition, the single-centre design is a cause of concern, as the population, standard practices, and workload can differ at other centres. Furthermore, one could argue that the study design, involving only one research nurse, could affect generalizability to other nurses. However, this should be determined by the total number of nurses performing triage and, the 27 + 1 nurses included in this study should be sufficient. The use of a single research nurse is less likely to be a problem related to generalizability than to an increased risk of underestimating reliability. This situation arises because if the triage level of the sole research nurse is generally higher or lower than those of the ED nurses, it should be manifested as systematic disagreement, resulting in a concomitant underestimation of the reliability.[26, 27] However, there was no evident systematic disagreement in the present study (table 2). In addition, generalizability to other nurses should be strengthened by the design of this study, which resembled a clinical ED setting, in that the ED nurses were not recently trained in the use of the PETI, and six of them lacked formal training in its use, only being trained by their colleagues while working. Fourth, a small proportion of the participants were triaged to the most urgent level (n=3), which seems to be a common problem in studies of "live" triage but does not necessary result in overestimation of kappa values because with most triage systems, triage of the most urgent patients is simple.[5, 7, 9, 25]

In conclusion, our results suggest that the PETI has substantial reliability when used in paediatric patients aged 0-12 years, and almost perfect reliability for patients aged 1-3 years. Moreover, this instrument can be rapidly administered. These findings indicate that triage relying on VPs is advantageous mostly among younger children, in whom the ability to perform triage relying on communication is limited. Because the PETI exhibited promise regarding reliability, the next step should be a validation study.

Funding Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden.

Competing interests None declared.

Acknowledgements The authors thank Antonis Valachis for critically reviewing the manuscript, Lovisa Graflund for data collection, Anna Ekholm for statistical analysis, and the ED nurses for cooperation in the triage procedure.

Contributors Development of the PETI: JK. Design of the study, data interpretation and participation in the statistical analysis: JK and SE. Drafting the manuscript and participation in the acquisition of the data: JK. Critical revision and finaliszation of the manuscript: SE.

Ethics approval The regional board of ethics in Stockholm, reference 2011/11-31/12.

REFERENCES

1 Hostetler MA, Mace S, Brown K, et al. Emergency department overcrowding and children. *Pediatr Emerg Care* 2007;23:507-15.
2 Tang N, Stein J, Hsia RY, et al. Trends and characteristics of US emergency department visits, 1997-2007. *JAMA* 2010;304:664-70.
3 Baumann MR, Strout TD. Evaluation of the Emergency Severity Index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med* 2005;12:219-24.
4 Durani Y, Brecher D, Walmsley D, et al. The Emergency Severity Index Version 4: reliability in pediatric patients. *Pediatr Emerg Care* 2009;25:751-3.
5 Gravel J, Gouin S, Goldman RD, et al. The Canadian Triage and Acuity Scale for children: a prospective multicenter evaluation. *Ann Emerg Med* 2012;60:71-7.
6 Green NA, Durani Y, Brecher D, et al. Emergency Severity Index version 4: a valid and reliable tool in pediatric emergency department triage. *Pediatr Emerg Care* 2012;28:753-7.
7 Travers DA, Waller AE, Katznelson J, et al. Reliability and validity of the emergency severity index for pediatric triage. *Acad Emerg Med* 2009;16:843-9.
8 van Veen M, Moll HA. Reliability and validity of triage systems in paediatric emergency care. *Scand J Trauma Resusc Emerg Med* 2009;17:38.
9 van Veen M, Teunen-van der Walle VF, Steyerberg EW, et al. Repeatability of the Manchester Triage System for children. *Emerg Med J* 2010;27:512-6.
10 Advanced Life Support Group. Advanced paediatric life support: The Practical Approach. 4th ed. Malden, Mass: BMJ Books 2005.
11 Van Gerven R, Delooz H, Sermeus W. Systematic triage in the emergency department using the Australian National Triage Scale: a pilot project. *Eur J Emerg Med* 2001;8:3-7.
12 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.
13 Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epidemiol* 2010;63:384-8.
14 Warren DW, Jarvis A, LeBlanc L, et al. Revisions to the Canadian Triage and Acuity Scale paediatric guidelines (PaedCTAS). *Cjem* 2008;10:224-43.
15 Gilboy N, Tanabe P, Travers D, et al. Emergency Severity Index, Version 4: Implementation Handbook. Rockville, MD: Agency for Healthcare Research and Quality 2005.
16 Duncan H, Hutchison J, Parshuram CS. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. *J Crit Care* 2006;21:271-8.
17 Haines C, Perrott M, Weir P. Promoting care for acutely ill children-development and evaluation of a paediatric early warning tool. *Intensive Crit Care Nurs* 2006;22:73-81.
18 Microsoft Software. Medcalc Software bvba version 12.4. Available from: http://www.medcalc.org/manual/kappa.php. Accessed January 13, 2013.

19 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
20 Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.
21 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.
22 Cooper RJ, Schriger DL, Flaherty HL, et al. Effect of vital signs on triage decisions. *Ann Emerg Med* 2002;39:223-32.
23 Roland D, McCaffery K, Davies F. Scoring systems in paediatric emergency care: Panacea or paper exercise? J *Paediatr Child Health* 2016;52:181-6.
24 van der Wulp I, van Stel HF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol* 2009;62:1196-201.
25 van der Wulp I, van Stel HF. Calculating kappas from adjusted data improved the comparability of the reliability of triage systems: a comparative study. *J Clin Epidemiol* 2010;63:1256-63.
26 Svensson E. Different ranking approaches defining association and agreement measures of paired ordinal data. *Stat Med* 2012;31:3104-3117.
27 Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231-240.

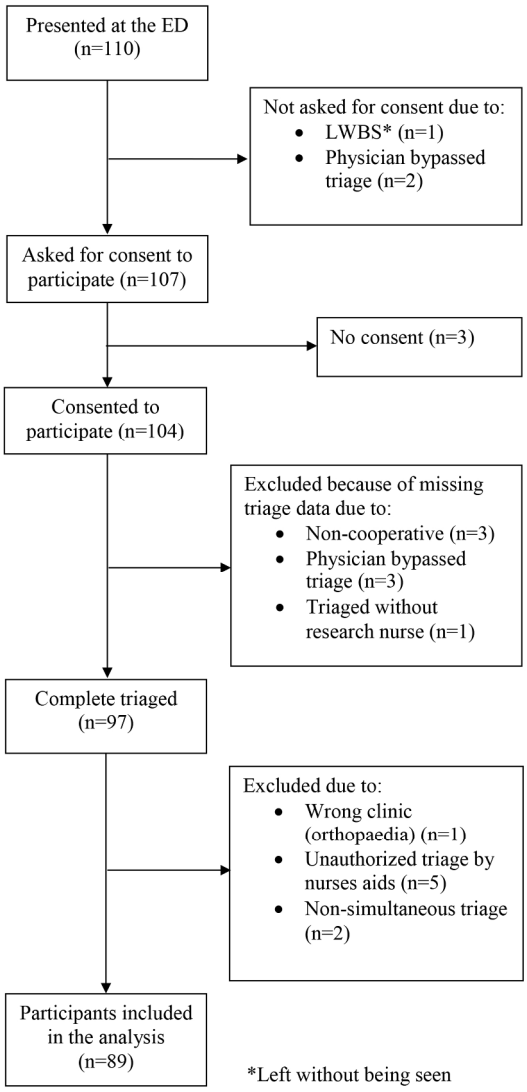

Figure 1. Flow of participant inclusion

148x246mm (300 x 300 DPI)

Appendix 1

Including the PETI-protocol and the reference values of the VIs for use in the PETI

THE PETI - PROTOCOL

Presenting complaint	Date	Time of arrival	Clinic
Time to TRIAGE (by reception desk assessment) RED= 0 min ORANGE=15 min YELLOW= 30 min GREEN=60 min			
Notes			

ID-tag

TRIAGE			4 points EMERGENT	2 points VERY URGENT	1 point URGENT	0 points NON-URGENT	POINTS
A		VS	Airway obstruction Stridor	Airway compromised Severe recession	Mild recession	Normal breathing	
B	SpO_2	%	Spo_2	SaO_2	SaO_2	SaO_2	
	RR	/min	RR	RR	RR	RR	
C	HR	/min	HR	HR	HR	HR	
	CRT	s	CRT	CRT	CRT	CRT	
D	AVPU		U/P Convulsions	V Sloppy infant or irritable infant		A	
E	Temp	°C	Temp	Temp	Temp	Temp	

EMERGENT ≥10p VERY URGENT 6-9p URGENT 2-5 NON-URGENT 0-1p = p

Reason for overruling triage decision	Notes i.e. weight, given medication

THE REFERENCE VALUES OF THE VITAL PARAMETERS (VPs) FOR USE IN THE PETI

	4 points **EMERGENT**	2 points **VERY URGENT**	1 point **URGENT**	0 points **NON-URGENT**
A	**Airway obstruction Stridor**	**Airway compromised Severe recession**	**Mild recession**	**Normal breathing**
B SpO2 %	≤90	≥91	≥93	≥95
RR <1years	>60 or ≤15	50-60	40-50	30-40
RR 1-2years	>45 or ≤12	40-45	35-40	25-35
RR 3-4years	>40 or ≤12	35-40	30-35	25-30
RR 5-12 years	>35 or ≤8	30-35	25-30	20-25
RR >12years	>30 or ≤8	25-30	20-25	15-20
CRT	≥4 seconds	3 seconds	2 seconds	1 second
HR <1years	>180 or ≤70	170-180	160-170	110-160
HR 1-2years	>170 or ≤60	160-170	150-160	100-150
HR 3-4years	>160 or ≤60	150-160	140-150	95-140
HR 5-12years	>140 or ≤50	130-140	120-130	80-120
HR >12years	>130 or ≤40	110-130 or ≤50	100-110	60-100
Disability	**U/P (of AVPU)**	**V (of AVPU) Sloppy infant/Irritable infant**		**A (of AVPU)**
E temperature C° ≤3months	≥39 or ≤35	←	≥38	36,5-37,9
≤3 years	≤35	≥39	≥38	36,5-37,9
>3 years	≤35	≥41	≥38	36,5-37,9

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1 and 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	3
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4 and 5
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4, 5, appendix 1
	10b	Reference standard, in sufficient detail to allow replication	Not applicable
	11	Rationale for choosing the reference standard (if alternatives exist)	Not applicable
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Not applicable
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Not applicable
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	4 and 5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	4 and 5
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5
	15	How indeterminate index test or reference standard results were handled	Not applicable
	16	How missing data on the index test and reference standard were handled	(Page 6), Figure 1
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5
	18	Intended sample size and how it was determined	4
RESULTS			
Participants	**19**	Flow of participants, using a diagram	6 and figure 1
	20	Baseline demographic and clinical characteristics of participants	6
	21a	Distribution of severity of disease in those with the target condition	Not applicable
	21b	Distribution of alternative diagnoses in those without the target condition	Not applicable
	22	Time interval and any clinical interventions between index test and reference standard	4 and 5
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	6, table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	7
	25	Any adverse events from performing the index test or the reference standard	No, and not reported
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	8
	27	Implications for practice, including the intended use and clinical role of the index test	7 and 8
OTHER INFORMATION			
	28	Registration number and name of registry	Not registered
	29	Where the full study protocol can be accessed	Not published/registered
	30	Sources of funding and other support; role of funders	8

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.

BMJ Open

Interrater reliability between nurses for a new paediatric triage system based primarily on vital parameters: The Paediatric Triage Instrument (PETI)

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2016-012748.R1
Article Type:	Research
Date Submitted by the Author:	10-Nov-2016
Complete List of Authors:	Karjala, Jaana; Department of pediatrics, Mälarsjukhuset Hospital; Centre for Clinical research Sörmland, Uppsala university Eriksson, Staffan; Centre for clinical research Sörmland, Uppsala University; Department of Neuroscience, Physiotherapy, Uppsala University
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Emergency medicine
Keywords:	PAEDIATRICS, ACCIDENT & EMERGENCY MEDICINE, HEALTH SERVICES ADMINISTRATION & MANAGEMENT

Interrater reliability between nurses for a new paediatric triage system based primarily on vital parameters: The Paediatric Triage Instrument (PETI)

Jaana Karjala[1,2], Staffan Eriksson[2, 3, 4]

[1]Department of Paediatrics, Mälarsjukhuset Hospital, Eskilstuna, Sweden. E-mail: jaanakarjala@gmail.com

[2]Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden

[3]Department of Neuroscience, Physiotherapy, Uppsala University, Uppsala, Sweden

[4]Department of Community Medicine and Rehabilitation, Physiotherapy, Umeå University, Umeå, Sweden

Corresponding author: Staffan Eriksson, Centre for Clinical Research Sörmland, Uppsala University, 631 88 Eskilstuna, Sweden. Telephone number: +46 16 105251, +46 73 949 99 33. E-mail: staffan.eriksson@germed.umu.se

Key words: paediatrics, triage, reliability, and emergency department.

Word count: 3934 words, 4172 words including strengths and limitations of this study (abstract page), funding, competing interests, acknowledgements, contributors, and ethics approval.

ABSTRACT

Introduction The major paediatric triage systems are primarily based on flow charts involving signs and symptoms for orientation and subjective estimates of the patient's condition. In contrast, the four level Paediatric Triage Instrument (PETI) is primarily based on vital parameters and was developed exclusively for paediatric triage in patients with medical complaints. The aim of this study was to assess the interrater reliability of this triage system in children when used by nurses.

Methods A design was employed in which triage was performed simultaneously and independently by a research nurse and an emergency department (ED) nurse using the PETI. All patients aged ≤12 years who presented at the ED with a medical complaint were considered eligible for participation.

Results The 89 participants exhibited a median age of 2 years and were triaged by 27 different nurses. The interrater reliability between nurses calculated with the quadratic-weighted kappa, was 0.78 (95% CI 0.67–0.89); the linear-weighted kappa was 0.67 (95% CI 0.56–0.80); and the unweighted kappa was 0.59 (95% CI 0.44–0.73). For the patients <1 year old, 1–3 years old, and >3 years old, the quadratic-weighted kappa values were 0.67 (95% CI 0.39–0.94), 0.86 (95% CI 0.75–0.97) and 0.73 (95% CI 0.49–0.97), respectively. The median triage duration was 6 minutes.

Conclusions The PETI exhibited substantial reliability when used in children ≤ 12 years old and almost perfect reliability among 1–3-year-old children. Moreover, rapid application of the PETI was demonstrated. This study has some limitations, including sample size and generalisability, but the PETI exhibited promise regarding reliability, and the next step could be either a larger reliability study or a validation study.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- The design of this 'live' triage study was drawn up to enable blindness and independency at all phases of the triage procedure.
- The design of this study, which resembled a clinical ED setting in that the ED nurses were not recently trained in the use of the PETI and six of them lacked formal training in its use, should strengthen the generalisability to other nurses.
- The sample size of this study was relatively small, which resulted in wide CIs and uncertainty in some of the results.
- Because of the mall sample size and the single-centre design, there are some issues regarding the generalisability of the results.
- A small proportion of the participants were triaged to the most urgent level, but this does not necessary result in an overestimation of the reliability because with most triage systems, triage of the most urgent patients is simple.

INTRODUCTION

Since the early 1990s, there has been a dramatic increase in the number of emergency department (ED) visits.[1, 2] In addition to the increase in emergency visits, several other circumstances have contributed to the overcrowding of EDs, including an inadequate inpatient capacity, the increasing complexity of paediatric patients, the lack of medical staff, and the lack of easy access to primary care.[1] With overcrowding comes greater risks of medical errors and adverse events.[1, 2] The overcrowding of EDs has made triage systems important, and several such systems, such as the Australasian Triage Scale (ATS), the Manchester Triage System (MTS), the Canadian Triage and Acuity Scale (CTAS) and the Emergency Severity Index (ESI) emerged in the 1990s. These four systems are the most established triage systems for adults, and they are also used for paediatric patient populations with some adaptations.[3-9] In addition, the Rapid Emergency Triage and Treatment System (RETTS), including a paediatric version (RETTS-p), is widely used in Scandinavian countries.[10, 11]

The triage of children in an ED setting offers several challenges that differ from adult triage. First, infants and smaller children depend almost entirely on their parents and medical professionals for correct judgements of their status. Second, substantial physiologic variations and immaturity of organ development make small children more susceptible to sudden deterioration, which necessitates the continuous reassessment of children.[12] Some of the currently used paediatric triage systems have reached a substantial level of interrater reliability, although there is still room for improvement. In well-conducted studies of simultaneous 'live' triage, weighted kappa values of 0.57, 0.65, 0.74 and 0.76 have been reported for the ESI version 4, MTS, CTAS and RETTS-p, respectively.[5, 7, 9, 10] In addition, two meta-analyses including studies of both 'live' triage and the triage of paper case scenarios reported correlation coefficients of 0.60 and 0.77 for the CTAS and ESI, respectively, whereas a meta-analysis including only studies applying the triage of paper case scenarios reported a correlation coefficient of 0.40 for the ATS.[13-15]

One factor that may contribute to errors in triage is that triage decisions are based to a large extent on informed but subjective estimates of the patient's presenting condition, such as estimates of pain and future resource utilization in the ATS and ESI, respectively.[16, 17] Another negative factor may be the complexities of triage systems with large numbers of different presenting complaints.[8, 18, 19] To determine acuity levels, these complaints are accompanied by general and complaint-specific discriminating questions in the MTS and sets of general and complaint-specific criteria in the CTAS. The procedure for determining acuity level in the RETTS is similar to that in the CTAS and MTS in the use of presenting complaints and accompanying discriminating criteria, but in addition, it also relies on vital parameters (VPs) [10, 11].

In contrast to the major triage systems, the Paediatric Triage Instrument (PETI) relies primarily on measurements of VPs that are acquired irrespective of the presenting complaints**.** The use of VPs is accepted as important in triage because VPs offer objective measurements on which decisions can be based, and such objective measurements are expected to be especially important in children.[8, 19] Moreover, a triage system based on VPs should be easy and quick to use. An additional possible advantage is increased control of the deterioration of patients because a base-line is established during the first triage, and a rapidly applied triage system makes continuous reassessments more achievable.

The PETI is a four-level triage system that is exclusively applied for paediatric triage and is based primarily on the VPs of patients with medical complaints. In creating this system, the main focus was placed on achieving an initial assessment that is quick and objective.

The aim of this study was to assess the interrater reliability of the PETI in children with medical complaints when used by nurses. The secondary aims were to assess the interrater reliability of the PETI for three different age groups and to assess the duration of the triage procedure associated with the PETI.

METHODS

Study design

This study of interrater reliability applied a design in which each patient was simultaneously and independently triaged by a research nurse and an ED nurse who were blinded to each other's collection of the data and triage assignments. The participants were included prospectively and consecutively.

Study setting and population

The study was conducted at a county hospital in the centre of Sweden. The department of paediatrics provides care for a population of 60,000 individuals aged 0–18 years with a rich ethnic diversity. The ED at the hospital receives 45,000 patients visits annually, and in 2011, 13% of these visits were made by paediatric patients 12 years or younger with medical complaints.

All patients aged 12 years or younger who presented to the ED with a medical complaint were considered eligible for inclusion in the study. Only children between the ages of 0 and 12 years were included because a different triage system has been introduced in the ED for children older than 12 years. The number of participants was decided upon based on the pre-planned data collection time frame, which was limited by resources. Written informed consent was obtained from the parents.

The triage system: PETI

The PETI is a four-level triage system primarily based on measurements of the following five VPs: respiratory rate, heart rate, capillary saturations, capillary refill time, and core temperature (appendix 1). The measurement of each of these VPs is compared with an age-specific reference interval. Depending on the degree of deviation, the VP is assigned 1, 2 or 4 points. The final acuity level is given by the sum of the points assigned to each of the five VPs. Summed scores of 0–1, 2–5, 6–9 and ≥10 corresponds to acuity levels of 'non-urgent' (green), 'urgent' (yellow), 'very urgent' (orange) and 'emergent' (red), respectively. Hence, to limit over-triage, a minimum of 2 points is necessary for triage into the 'urgent' acuity level, and a minimum of 6 points is necessary for triage into the 'very urgent' acuity level. In addition, to emphasize severe cases, extra weight is added for large deviations in the VPs (4 vs. 2 points). The normal reference intervals for the VPs of respiratory rate and heart rate were set according to the Advanced Paediatric Life Support (APLS) system.[12] The normal reference value for the capillary refill time was adjusted based on the APLS value with the intention of increasing reliability. The normal reference intervals for saturation and temperature were set according to established experience. The reference intervals for deviations corresponding to 4 points for the VPs of temperature, capillary saturations, heart rate, and respiratory rate were set according to the cut-off values for danger zone vitals in the ESI, along with clinical experience.[20] The cut-off value for deviations in capillary refill

time corresponding to 4 points was set according to the APLS.[12] The reference values for deviations corresponding to 1–2 points were evenly distributed between normal and 4 points.

Some signs and symptoms included in the PETI are related to the airway and neurology and were selected from the ABCDE- model, including the alert, voice-, pain-, and unresponsive scale (AVPU-scale), which individually creates a 'force majeure' that complements triage based on VPs (appendix 1).[12] Triage based on a "force majeure" is independent from triage based on VPs, and the patient is assigned the highest acuity level between these two methods. These signs and symptoms are assessed prior to or during the collection of VP data. The signs and symptoms of mild recession results in assignment of the patient to the 'urgent' acuity level. Any of the following signs and symptoms result in assignment of the patient to the 'very urgent' acuity level: compromised airway, severe recession, a sloppy or irritable infant, or assessment of the child as voice responsive. Any of the following signs and symptoms result in assignment of the patient to the 'emergent' acuity level: airway obstruction, stridor, convulsions, or assessment of the child as either pain responsive or unresponsive.

The development of the PETI was influenced both by the major triage systems and, more importantly, by paediatric early warning systems, which rely heavily on VPs.[21, 22] During the development of the PETI, feedback was given by groups of paediatricians and other emergency staff.

Data collection

The ED nurses were trained in use of the PETI when the system was introduced at the ED one year prior to the study. This training was implemented via a two-hour lecture and through the opportunity to ask questions for 30 minutes the day the instrument was introduced, or via e-mail, or when the first author was serving at the ED. The research nurse had no previous experience with the PETI and was trained in the use of the system through two one-hour training sessions prior to the study. The research nurse performed 29 shifts of 6 hours each to recruit and triage patients for the study. All but one shift lasted from 4 p.m. to 10 p.m. on normal weekdays. The ED nurses who were working the same 29 shifts during which the research nurse was at the ED participated in the study. According to the established routine in the ED, all triage of children should be performed by an ED nurse.

Triage was performed simultaneously by an ED nurse and the research nurse and included measurements of five VPs and assessments of signs and symptoms related to a "force majeure". Capillary saturation and heart rate was measured using either a Nellcor Puritan Bennett NPB 295 or a Masimo Radical-7 pulse oximeter. Temperature was measured either rectally (in children <1 year old) using a Terumo C402 digital thermometer or aurally using a Braun ThermoScan 6022 tympanic thermometer. The measurements used to calculate the acuity levels of the PETI were performed via the application of two separate sets of instruments. The nurses concealed their data collection from each other by distancing themselves in the room, with the research nurse angling the instrument in use to shield it from the ED nurse. Both the ED nurse and the research nurse calculated acuity levels blindly and separately in different rooms, or separated by distance when in the same room. They were informed not to discuss their data collection or the assignment of acuity levels. The research nurse documented when she believed that the blindness and independency of the triage procedure had not been preserved. Only the ED nurse's triage results were used in patient care. The characteristics of the study participants were documented by the research nurse.

Statistical analysis

Interrater reliability was calculated for the whole group (primary analysis) and for the following post hoc subgroups: <1-, 1–3-, and 4–12-year-olds. The choice of subgroups was based on the purposes of analysing a group of patients <1 year old in whom difficulties in triage have previously been reported and creating groups with a sufficient number of participants for the analyses.[7] The primary test of interrater reliability that was calculated for the primary and subgroup analyses was Cohen's kappa with quadratic weights. The quadratic-weighted kappa was chosen because it accounts for the degree of disagreement and the severity of disagreement at higher acuity levels.[23] Additionally, to enable comparison with other studies, Cohen's kappa with linear weights or no weights was also calculated for the whole group. The kappa values were interpreted according to the following categories: <0.40 poor-fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.0 almost perfect.[24] The duration of triage from the beginning of the collection of the triage data to the assignment of the acuity level was determined for the research nurse. The kappa values and 95% confidence intervals (CIs) were calculated with MedCalc® 12.4.[23]

RESULTS

Data collection was performed from November 3, 2011, to January 11, 2012. Twenty-seven ED nurses participated in the study, six of whom began their employment at the ED after training on the PETI took place and were trained solely by their colleagues while working. The median amount of experience in emergency medicine was 4 years (interquartile range [IQR] 2–15) for the ED nurses, and the research nurse had 1.5 years of experience.

The ED nurses triaged a median of 2 participants each (IQR 1–5). One hundred and four patients agreed to participate in the study, fifteen of whom were excluded; thus, 89 participants were included in the analysis (figure 1). The median age of the included patients was 2 years (IQR 0–11) and 48% were girls (table 1). Overall the characteristics of the study participants corresponded rather well to the characteristics of the patient population (table 1). In nine of the 89 participants, acuity levels were assigned by 'force majeure'. The blindness and independency of the triage procedure between the ED and research nurses was preserved for 75 of 89 participants (84%). The reasons for the failure to preserve blindness included use of the same measurement for temperature due to parental discomfort (n=11) and the need for acute medical procedures in the emergency room (the 3 emergent participants).

Table 1 Characteristics of the participants and the patient population

	Study participants, n=89*	**All patients, n=651†**
Gender female, n (%)	43 (48)	286 (44)
Gender male, n (%)	46 (52)	365 (56)
Age		
<1year, n (%)	29 (33)	169 (26)
1–3 years, n (%)	33 (37)	306 (47)
4–12 years, n (%)	27 (30)	176 (27)
Chief complaints		
Asthma and allergy, n (%)	10 (11)	55 (9)
Fever, n (%)	4 (5)	24 (4)
GI- and urinary tracts, n (%)	20 (23)	96 (15)
Neurological, n (%)	7 (8)	10 (2)
Observation, n (%)	2 (2)	32 (5)
Respiratory tract, n (%)	31 (35)	280 (43)
Other, n (%)	15 (17)	154 (24)

*Participants with a medical complaint, non-surgical, non-orthopaedic
†All patients 0–12 years old; period 3rd November 2010 to 11th January 2011 (i.e. the corresponding period a year prior to the study)
‡Interquartile range

The agreement in the acuity level of the PETI between the research nurse and the ED nurses was 73% (table 2). There was no evident systematic disagreement, as either the research nurse or the ED nurse triaged a participant to a higher acuity level than the other nurse on approximately the same number of occasions: 11 (7 + 3 + 1) and 13 (7 + 6) occasions, respectively (table 2). The agreement by age was 76%, 76% and 67% for participants with ages of <1-, 1–3-, and 4–12 years, respectively (table 3). The median (IQR) duration of the triage procedure was 6 minutes (4.25–7 minutes), n=81.

Table 2 Agreement of acuity levels between the research nurse and the ED nurses, n=89

ED nurse	Research nurse				
	Non-urgent	**Urgent**	**Very urgent**	**Emergent**	**Total**
Non-urgent (Green)	29*	7	1	0	37
Urgent (Yellow)	7	22*	3	0	32
Very urgent (Orange)	0	6	11*	0	17
Emergent (Red)	0	0	0	3*	3
Total	36	35	15	3	89

*Cases showing agreement between the ED nurse and the research nurse

Table 3 Agreement by age, n=89 participants

	<1 yr	**1–3 yr**	**4–12 yr**	**Total**
Agreement	22	25	18	65
Disagreement by one level	6	8	9	23
Disagreement by two levels	1	0	0	1
Total	29	33	27	89

The interrater reliability values for the nurses were 0.78 (95% CI 0.67–0.89) based on the quadratic-weighted kappa, 0.67 (95% CI 0.56–0.80) based on the linear-weighted kappa, and 0.59 (95% CI 0.44–0.73) based on the unweighted kappa. (The corresponding kappa values, including cases with unauthorized triage decisions by nurse aids, were 0.78, 0.68 and 0.58, n=94).

The quadratic-weighted kappa (95% CI) values were 0.67 (0.39–0.94), 0.86 (0.75–0.97), and 0.73 (0.49–0.97) for patients with ages of <1, 1–3, and 4–12 years, respectively.

DISCUSSION

This study demonstrated substantial interrater reliability of the new PETI triage system for paediatric patients ≤12 years old. Additionally, the time required for the completion of the PETI was quite short. The PETI therefore exhibited promise, particularly considering that this study was conducted in a clinical setting in which the ED nurses received no extra training or practice in the triage of case scenarios prior to the study of 'live' triage, which is common in other studies.[3, 7, 9] In addition, an effort was made to perform the triage procedure as independently and blinded as possible, and as reported, this was achieved in 84% of the participants.

The level of reliability observed in our study for the PETI triage instrument is comparable to the best kappa values for simultaneous 'live' triage that have previously been published. Quadratic-weighted kappa values of 0.65, 0.74 and 0.76 have been reported for the MTS, CTAS and RETTS-p, respectively, whereas a linear-weighted kappa for the ESI version 4 was reported as 0.57.[5, 7, 9, 10] While it is difficult to conduct a study of 'live' triage with blindness and independency at all phases of triage, they are important factors.[25] The studies of the RETTS, CTAS, MTS and ESI version 4 were large, well–conducted studies, and the latter three were multi-centre studies with superior generalisability to this study of the PETI. However, by study design, the nurses that performed triage in the studies of the RETTS and CTAS shared VP data to some degree for all of the participants.[5, 10] Similarly, in the studies of the MTS and ESI version 4, the independency regarding the VP-data and other information on which to base the acuity level assignment was not stated.[7, 9] Triage is not only about assigning an acuity level but also about obtaining information on which to base the assignment. Some studies of 'live' triage have obtained higher kappa values in the range of 0.8–0.9 for the ESI version 3 and 4.[3, 6] These studies exhibited drawbacks, including limited sample size and total dependency of the VP data used for the assignment of acuity level, which makes the interpretation of their findings questionable.[25] In addition, some studies that employed the triage of paper case scenarios have also reported high kappa values in the range of 0.8–0.9 for the ESI version 4, MTS and RETTS-p.[4, 7, 8, 10, 11, 26] Similarly, one meta-analysis, which to a large extent, is based on studies of paper case scenarios reported a correlation coefficient of 0.77 for the ESI.[15] One could argue that case scenarios do not reflect the real clinical setting in which the interactions between the nurse, patient, environment and triage system may contribute to mistriage.[18, 25] Indeed, in studies in which both paediatric paper case scenarios and 'live' paediatric patients were triaged within the same study, the kappa values were approximately 0.06–0.2 units higher for case scenarios than for 'live' triage.[3, 7, 9, 10] In contrast, a study that compared 'live' triage with case scenarios based on the use of the CTAS in a mixed population of adults and children found that the kappa value was higher for the 'live' use of the triage algorithm.[27] However, the use of a mixed population in this study makes the comparison of results between studies difficult.

The PETI exhibited a tendency towards showing the best reliability in children aged 1–3 years. A possible explanation for this finding is that the VPs provided a clearly defined framework for the triage of children who lack the ability to efficiently communicate. It has previously been demonstrated that complementing subjective triage decisions with VP data often results in changes in triage decisions in children ≤2 years old and in children whose parents have communication difficulties.[28] The PETI exhibited a tendency towards inferior reliability for children <1 year old, which agrees with previous results for the ESI.[7] In general, triage in infants is particularly difficult because the severity of illness is expressed in multiple and subtle manners and can change rapidly.[12] However, this observation should be

regarded with caution because it stemmed from a discrepancy of two levels in a singly participant.

As assessment using the PETI was shown to be rapid, this tool will facilitate retriage and thereby facilitate the control of patient deterioration. It will also potentially decrease the strain on staff and contribute to resource effectiveness.

It has previously been shown that paediatric triage systems that rely to a large extent on VPs are prone to over-triage (low sensitivity).[29] However, in developing the PETI, the risk of over-triage was compensated through the levels in the scoring system, such that a minimum of 2 points was required for triage into the 'urgent' acuity level, and a minimum of 6 points was required for triage into the 'very urgent' acuity level. As this was not a validation study, there were no available data on the participants' 'true' acuity levels, and it is not possible to answer the question of whether triage with the PETI is prone to over- or under-triage. Nevertheless, it is notable that approximately 40% of the participants were triaged to each of the 2 lowest acuity levels (research nurse: 36/89 'non-urgent' and 35/89 'urgent') (table 2).

Iimprovements in the measures employed in the PETI should likely focus on the VPs of respiratory rate and capillary refill because these VPs rely heavily on estimates and skill. Furthermore, the accuracy of different types and brands of measurement devices have to be taken into account with respect to limits for normal reference values, as consistent measurement deviations from the standard may have an influence on the validity of the PETI. Regarding triage based on 'force majeure', the relative position between stridor and severe recession should be considered in the process of improvements. Additionally, in the table illustrating the reference values for the VPs (appendix 1), detected errors should be revised, including gaps in the reference intervals for heart rate, respiratory rate and temperature. Minor revisions of the lay-out have already been incorporated.

This study has some limitations and some strengths. First, comparisons with studies on commonly used paediatric triage systems are difficult because the PETI is a four-level system, whereas the others are five-level systems. However, it has been suggested that kappa values in general, and quadratically weighted kappa values in particular, increase as the number of categories in a system increase.[30, 31] Second, a small proportion of the participants were triaged to the most urgent level (n=3), which seems to be a common problem in studies of 'live' triage but does not necessary result in the overestimation of kappa values because with most triage systems, triage of the most urgent patients is simple.[5, 7, 9, 31] Third, the sample size of this study was relatively small, which resulted in wide CIs and uncertainty in some of the results. For instance, the linear-weighted kappa for the whole group exhibited a 95% CI that stretched below the lower limit of the substantial category. Fourth, even though the characteristics of the participants in this study resembled those of the patient population, there are some issues regarding the generalisability of the results. The small sample size makes it likely that not all possible presenting complaints of the population were covered in the triage of the participants. In addition, the single-centre design is a cause of concern, as the population, standard practices, and workload can differ at other centres. Furthermore, one could argue that the study design, involving only one research nurse, could affect generalisability to other nurses. However, this should be determined by the total number of nurses performing triage, and the 27 + 1 nurses included in this study should be sufficient. The use of a single research nurse is less likely to be a problem related to generalisability than to an increased risk of underestimating reliability. This situation arises because if the triage level of the sole research nurse is generally higher or lower than those of the ED nurses, it

should be manifested as systematic disagreement, with a concomitant underestimation of the reliability as a result.[32, 33] However, no systematic disagreement was evident in the present study (table 2). In addition, generalisability to other nurses should be strengthened by the design of this study, which resembled a clinical ED setting, in that the ED nurses were not recently trained in the use of the PETI, and six of them lacked formal training in its use, only having been trained by their colleagues while working. Fifth, another strength of this study compared to other studies of 'live' triage was the study design, which was structured to enable blindness and independency at all phases of the triage procedure. Total independency was not achieved for all of the participants but compared with other studies the level of independency seemed high.[3, 5-7, 9, 10]

In conclusion, our results suggest that the PETI has substantial reliability when used in paediatric patients aged 0–12 years and almost perfect reliability for patients aged 1–3 years. Moreover, this instrument can be rapidly administered. These findings indicate that triage relying on VPs is advantageous mostly among younger children, in whom the ability to perform triage relying on communication is limited. This study has some limitations including sample size and generalisability, but the PETI exhibited promise regarding reliability, and the next step could be either a larger reliability study or a validation study.

Funding Centre for Clinical Research Sörmland, Uppsala University, Eskilstuna, Sweden.

Competing interests None declared.

Acknowledgements The authors thank Antonis Valachis for critically reviewing the manuscript, Lovisa Graflund for data collection, Anna Ekholm for statistical analysis, and the ED nurses for cooperation in the triage procedure.

Contributors Development of the PETI: JK. Design of the study, data interpretation and participation in the statistical analysis: JK and SE. Drafting the manuscript and participation in the acquisition of the data: JK. Critical revision and finaliszation of the manuscript: SE.

Ethics approval The regional board of ethics in Stockholm, reference 2011/11-31/12.

REFERENCES

1. Hostetler MA, Mace S, Brown K, et al. Emergency department overcrowding and children. *Pediatr Emerg Care* 2007;23:507-15.
2. Tang N, Stein J, Hsia RY, et al. Trends and characteristics of US emergency department visits, 1997-2007. *JAMA* 2010;304:664-70.
3. Baumann MR, Strout TD. Evaluation of the Emergency Severity Index (version 3) triage algorithm in pediatric patients. *Acad Emerg Med* 2005;12:219-24.
4. Durani Y, Brecher D, Walmsley D, et al. The Emergency Severity Index Version 4: reliability in pediatric patients. *Pediatr Emerg Care* 2009;25:751-3.
5. Gravel J, Gouin S, Goldman RD, et al. The Canadian Triage and Acuity Scale for children: a prospective multicenter evaluation. *Ann Emerg Med* 2012;60:71-7.
6. Green NA, Durani Y, Brecher D, et al. Emergency Severity Index version 4: a valid and reliable tool in pediatric emergency department triage. *Pediatr Emerg Care* 2012;28:753-7.
7. Travers DA, Waller AE, Katznelson J, et al. Reliability and validity of the emergency severity index for pediatric triage. *Acad Emerg Med* 2009;16:843-9.
8. van Veen M, Moll HA. Reliability and validity of triage systems in paediatric emergency care. *Scand J Trauma Resusc Emerg med* 2009;17:38.
9. van Veen M, Teunen-van der Walle VF, Steyerberg EW, et al. Repeatability of the Manchester Triage System for children. *Emerg Med J* 2010;27:512-6.
10. Henning B, Lydersen S, Dollner H. A reliability study of the rapid emergency triage and treatment system for children. *Scand J Trauma Resusc Emerg med* 2016;24:19.
11. Westergren H, Ferm M, Haggstrom P. First evaluation of the paediatric version of the Swedish rapid emergency triage and treatment system shows good reliability. *Acta Paediatr* 2014;103:305-8.
12. Advanced Life Support Group. Advanced paediatric life support: The Practical Approach. 4th ed. Malden, Mass: BMJ Books 2005.
13. Ebrahimi M, Heydari A, Mazlom R, et al. The reliability of the Australasian Triage Scale: a meta-analysis. *World journal of emergency medicine* 2015;6:94-9.
14. Mirhaghi A, Heydari A, Mazlom R, et al. The Reliability of the Canadian Triage and Acuity Scale: Meta-analysis. *N Am J Med Sci* 2015;7:299-305.
15. Mirhaghi A, Heydari A, Mazlom R, et al. Reliability of the Emergency Severity Index: Meta-analysis. *Sultan Qaboos Univ Med J* 2015;15:e71-7
16. Van Gerven R, Delooz H, Sermeus W. Systematic triage in the emergency department using the Australian National Triage Scale: a pilot project. *Eur J Emerg Med* 2001;8:3-7
17. 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
18. Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epidemiol* 2010;63:384-8.
19. Warren DW, Jarvis A, LeBlanc L, et al. Revisions to the Canadian Triage and Acuity Scale paediatric guidelines (PaedCTAS). *Cjem* 2008;10:224-43
20. Gilboy N, Tanabe P, Travers D, et al. *Emergency Severity Index, Version 4: Implementation Handbook*. Rockville, MD: Agency for Healthcare Research and Quality 2005.

21. Duncan H, Hutchison J, Parshuram CS. The Pediatric Early Warning System score: a severity of illness score to predict urgent medical need in hospitalized children. *J Crit Care* 2006;21:271-8.
22. Haines C, Perrott M, Weir P. Promoting care for acutely ill children-development and evaluation of a paediatric early warning tool. *Intensive Crit Care Nurs* 2006;22:73-81.
23. Microsoft Software. Medcalc Software bvba version 12.4. Available from: http://www.medcalc.org/manual/kappa.php. Accessed January 13, 2013.
24. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74
25. Kottner J, Audige L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96-106.
26. Jafari-Rouhi AH, Sardashti S, Taghizadieh A, et al. The Emergency Severity Index, version 4, for pediatric triage: a reliability study in Tabriz Children's Hospital, Tabriz, Iran. *Int J Emerg Med* 2013;6:36.
27. 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.
28. Cooper RJ, Schriger DL, Flaherty HL, et al. Effect of vital signs on triage decisions. *Ann Emerg Med* 2002;39:223-32
29. Roland D, McCaffery K, Davies F. Scoring systems in paediatric emergency care: Panacea or paper exercise? *J Paediatr Child Health* 2016;52:181-6.
30. van der Wulp I, van Stel HF. Adjusting weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol* 2009;62:1196-201.
31. van der Wulp I, van Stel HF. Calculating kappas from adjusted data improved the comparability of the reliability of triage systems: a comparative study. *J Clin Epidemiol* 2010;63:1256-63.
32. Svensson E. Different ranking approaches defining association and agreement measures of paired ordinal data. *Stat Med* 2012;31:3104-17.
33. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 2005;19:231-40.

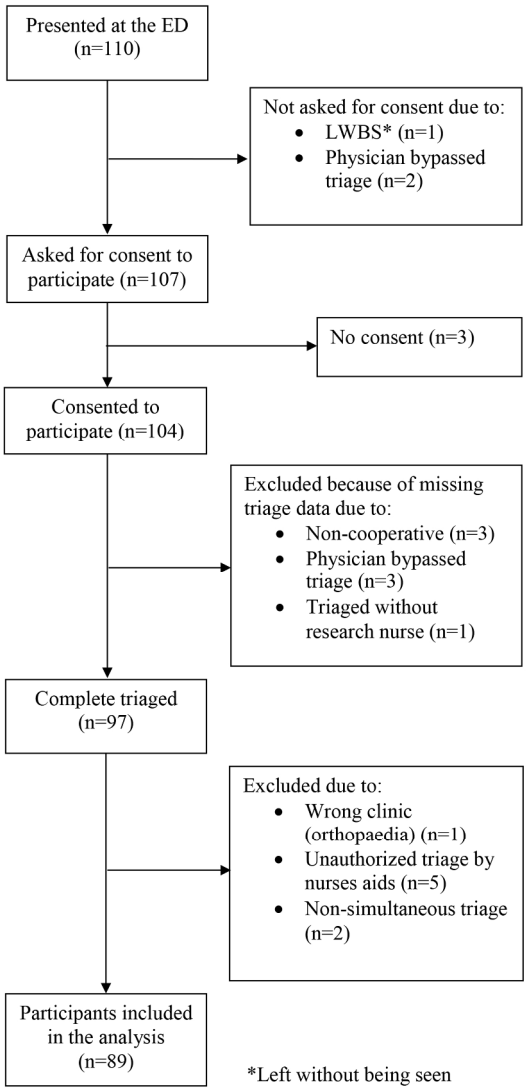

Figure 1. Flow of participant inclusion

148x246mm (300 x 300 DPI)

Appendix 1

Including the PETI-protocol and the reference values of the VIs for use in the PETI

THE PETI - PROTOCOL

Presenting complaint	Date	Time of arrival	Clinic
Time to TRIAGE (by reception desk assessment) RED= 0 min ORANGE=15 min YELLOW= 30 min GREEN=60 min			
Notes			

ID-tag

TRIAGE			4 points EMERGENT	2 points VERY URGENT	1 point URGENT	0 points NON-URGENT	POINTS
A		VS	Airway obstruction Stridor	Airway compromised Severe recession	Mild recession	Normal breathing	
B	SpO_2	%	Spo_2	SaO_2	SaO_2	SaO_2	
	RR	/min	RR	RR	RR	RR	
C	HR	/min	HR	HR	HR	HR	
	CRT	s	CRT	CRT	CRT	CRT	
D	AVPU		U/P Convulsions	V Sloppy infant or irritable infant		A	
E	Temp	°C	Temp	Temp	Temp	Temp	

EMERGENT ≥10p VERY URGENT 6-9p URGENT 2-5 NON-URGENT 0-1p = p

Reason for overruling triage decision	Notes i.e. weight, given medication

THE REFERENCE VALUES OF THE VITAL PARAMETERS (VPs) FOR USE IN THE PETI

	4 points **EMERGENT**	2 points **VERY URGENT**	1 point **URGENT**	0 points **NON-URGENT**
A	**Airway obstruction Stridor**	**Airway compromised Severe recession**	**Mild recession**	**Normal breathing**
B SpO2 %	≤90	≥91	≥93	≥95
RR <1years	>60 or ≤15	50-60	40-50	30-40
RR 1-2years	>45 or ≤12	40-45	35-40	25-35
RR 3-4years	>40 or ≤12	35-40	30-35	25-30
RR 5-12 years	>35 or ≤8	30-35	25-30	20-25
RR >12years	>30 or ≤8	25-30	20-25	15-20
CRT	≥4 seconds	3 seconds	2 seconds	1 second
HR <1years	>180 or ≤70	170-180	160-170	110-160
HR 1-2years	>170 or ≤60	160-170	150-160	100-150
HR 3-4years	>160 or ≤60	150-160	140-150	95-140
HR 5-12years	>140 or ≤50	130-140	120-130	80-120
HR >12years	>130 or ≤40	110-130 or ≤50	100-110	60-100
Disability	**U/P (of AVPU)**	**V (of AVPU) Sloppy infant/Irritable infant**		**A (of AVPU)**
E temperature C° ≤3months	≥39 or ≤35	←	≥38	36,5-37,9
≤3 years	≤35	≥39	≥38	36,5-37,9
>3 years	≤35	≥41	≥38	36,5-37,9

Section & Topic	No	Item	Reported on page #
TITLE OR ABSTRACT			
	1	Identification as a study of diagnostic accuracy using at least one measure of accuracy (such as sensitivity, specificity, predictive values, or AUC)	1 and 2
ABSTRACT			
	2	Structured summary of study design, methods, results, and conclusions (for specific guidance, see STARD for Abstracts)	2
INTRODUCTION			
	3	Scientific and clinical background, including the intended use and clinical role of the index test	3
	4	Study objectives and hypotheses	4
METHODS			
Study design	**5**	Whether data collection was planned before the index test and reference standard were performed (prospective study) or after (retrospective study)	4
Participants	**6**	Eligibility criteria	4
	7	On what basis potentially eligible participants were identified (such as symptoms, results from previous tests, inclusion in registry)	4
	8	Where and when potentially eligible participants were identified (setting, location and dates)	4, 5 and 6
	9	Whether participants formed a consecutive, random or convenience series	4
Test methods	**10a**	Index test, in sufficient detail to allow replication	4, 5, appendix 1
	10b	Reference standard, in sufficient detail to allow replication	Not applicable
	11	Rationale for choosing the reference standard (if alternatives exist)	Not applicable
	12a	Definition of and rationale for test positivity cut-offs or result categories of the index test, distinguishing pre-specified from exploratory	Not applicable
	12b	Definition of and rationale for test positivity cut-offs or result categories of the reference standard, distinguishing pre-specified from exploratory	Not applicable
	13a	Whether clinical information and reference standard results were available to the performers/readers of the index test	4 and 5
	13b	Whether clinical information and index test results were available to the assessors of the reference standard	4 and 5
Analysis	**14**	Methods for estimating or comparing measures of diagnostic accuracy	5-6
	15	How indeterminate index test or reference standard results were handled	Not applicable
	16	How missing data on the index test and reference standard were handled	(Page 6), figure 1
	17	Any analyses of variability in diagnostic accuracy, distinguishing pre-specified from exploratory	5-6
	18	Intended sample size and how it was determined	4
RESULTS			
Participants	**19**	Flow of participants, using a diagram	6 and figure 1
	20	Baseline demographic and clinical characteristics of participants	6, table 1
	21a	Distribution of severity of disease in those with the target condition	Not applicable
	21b	Distribution of alternative diagnoses in those without the target condition	Not applicable
	22	Time interval and any clinical interventions between index test and reference standard	4 and 5
Test results	**23**	Cross tabulation of the index test results (or their distribution) by the results of the reference standard	7, table 2
	24	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	7
	25	Any adverse events from performing the index test or the reference standard	No, and not reported
DISCUSSION			
	26	Study limitations, including sources of potential bias, statistical uncertainty, and generalisability	9-10
	27	Implications for practice, including the intended use and clinical role of the index test	8 and 9
OTHER INFORMATION			
	28	Registration number and name of registry	Not registered
	29	Where the full study protocol can be accessed	Not published/registered
	30	Sources of funding and other support; role of funders	10

STARD 2015

AIM

STARD stands for "Standards for Reporting Diagnostic accuracy studies". This list of items was developed to contribute to the completeness and transparency of reporting of diagnostic accuracy studies. Authors can use the list to write informative study reports. Editors and peer-reviewers can use it to evaluate whether the information has been included in manuscripts submitted for publication.

EXPLANATION

A **diagnostic accuracy study** evaluates the ability of one or more medical tests to correctly classify study participants as having a **target condition.** This can be a disease, a disease stage, response or benefit from therapy, or an event or condition in the future. A medical test can be an imaging procedure, a laboratory test, elements from history and physical examination, a combination of these, or any other method for collecting information about the current health status of a patient.

The test whose accuracy is evaluated is called **index test.** A study can evaluate the accuracy of one or more index tests. Evaluating the ability of a medical test to correctly classify patients is typically done by comparing the distribution of the index test results with those of the **reference standard**. The reference standard is the best available method for establishing the presence or absence of the target condition. An accuracy study can rely on one or more reference standards.

If test results are categorized as either positive or negative, the cross tabulation of the index test results against those of the reference standard can be used to estimate the **sensitivity** of the index test (the proportion of participants *with* the target condition who have a positive index test), and its **specificity** (the proportion *without* the target condition who have a negative index test). From this cross tabulation (sometimes referred to as the contingency or "2x2" table), several other accuracy statistics can be estimated, such as the positive and negative **predictive values** of the test. Confidence intervals around estimates of accuracy can then be calculated to quantify the statistical **precision** of the measurements.

If the index test results can take more than two values, categorization of test results as positive or negative requires a **test positivity cut-off**. When multiple such cut-offs can be defined, authors can report a receiver operating characteristic (ROC) curve which graphically represents the combination of sensitivity and specificity for each possible test positivity cut-off. The **area under the ROC curve** informs in a single numerical value about the overall diagnostic accuracy of the index test.

The **intended use** of a medical test can be diagnosis, screening, staging, monitoring, surveillance, prediction or prognosis. The **clinical role** of a test explains its position relative to existing tests in the clinical pathway. A replacement test, for example, replaces an existing test. A triage test is used before an existing test; an add-on test is used after an existing test.

Besides diagnostic accuracy, several other outcomes and statistics may be relevant in the evaluation of medical tests. Medical tests can also be used to classify patients for purposes other than diagnosis, such as staging or prognosis. The STARD list was not explicitly developed for these other outcomes, statistics, and study types, although most STARD items would still apply.

DEVELOPMENT

This STARD list was released in 2015. The 30 items were identified by an international expert group of methodologists, researchers, and editors. The guiding principle in the development of STARD was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. The list represents an update of the first version, which was published in 2003.

More information can be found on http://www.equator-network.org/reporting-guidelines/stard.