

BMJ Open Value of systematic detection of physical child abuse at emergency rooms: a cross-sectional diagnostic accuracy study

Judith S Sittig,^{1,2} Cuno S P M Uiterwaal,³ Karel G M Moons,³ Ingrid M B Russel,¹ Rutger A J Nievelstein,⁴ Edward E S Nieuwenhuis,¹ Elise M van de Putte¹

To cite: Sittig JS, Uiterwaal CSPM, Moons KGM, *et al.* Value of systematic detection of physical child abuse at emergency rooms: a cross-sectional diagnostic accuracy study. *BMJ Open* 2016;**6**: e010788. doi:10.1136/bmjopen-2015-010788

► Prepublication history and additional material is available. To view please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2015-010788>).

Received 6 December 2015
Revised 9 February 2016
Accepted 29 February 2016



CrossMark

For numbered affiliations see end of article.

Correspondence to
Dr Judith S Sittig;
judith@sittig.nl

ABSTRACT

Objectives: The aim of our diagnostic accuracy study Child Abuse Inventory at Emergency Rooms (CHAIN-ER) was to establish whether a widely used checklist accurately detects or excludes physical abuse among children presenting to ERs with physical injury.

Design: A large multicentre study with a 6-month follow-up.

Setting: 4 ERs in The Netherlands.

Participants: 4290 children aged 0–7 years attending the ER because of physical injury. All children were systematically tested with an easy-to-use child abuse checklist (index test). A national expert panel (reference standard) retrospectively assessed all children with positive screens and a 15% random sample of the children with negative screens for physical abuse, using additional information, namely, an injury history taken by a paediatrician, information provided by the general practitioner, youth doctor and social services by structured questionnaires, and 6-month follow-up information.

Main outcome measure: Physical child abuse.

Secondary outcome measure: Injury due to neglect and need for help.

Results: 4253/4290 (99%) parents agreed to follow-up. At a prevalence of 0.07% (3/4253) for inflicted injury by expert panel decision, the positive predictive value of the checklist was 0.03 (95% CI 0.006 to 0.085), and the negative predictive value 1.0 (0.994 to 1.0). There was 100% (93 to 100) agreement about inflicted injury in children, with positive screens between the expert panel and child abuse experts.

Conclusions: Rare cases of inflicted injury among preschool children presenting at ERs for injury are very likely captured by easy-to-use checklists, but at very high false-positive rates. Subsequent assessment by child abuse experts can be safely restricted to children with positive screens at very low risk of missing cases of inflicted injury. Because of the high false positive rate, we do advise careful prior consideration of cost-effectiveness and clinical and societal implications before de novo implementation.

INTRODUCTION

Physical abuse causes 1% of injuries seen in children attending the emergency room

Strengths and limitations of this study

- This diagnostic study to detect physical child abuse is the first study that meets all Quality Assessment of Diagnostic Accuracy Studies criteria for diagnostic accuracy studies.
- To date, no diagnostic studies of physical child abuse have used the same reference procedure for children that tested positive and negative with checklist and, consequently, accurate predictive values of a negative outcome of a checklist to detect physical child abuse were not available before.
- We used data of over 4000 children.
- The problem of low prevalence, and hence, low predictive value is a core problem.

(ER) according to recent studies.¹ Both direct consequences of injury, and adverse effects on education, mental and physical health, and violent or criminal behaviour^{2 3} justify proper diagnosis at hospital ERs. However, physical child abuse seems under-reported by professionals, mainly due to non-recognition and the lack of confidence that reporting would improve patient outcomes, besides the possible harms of reporting such as avoiding healthcare for the child. Moreover, the doctors are reluctant in reporting because of the fear of being wrong and confrontation with parents.⁴ Indeed, estimated proportions of missed cases range from 11% to 64%.^{1 2 4 5}

Standard diagnostic instruments in children are increasingly being used, or their use is considered in ERs worldwide to detect child abuse.^{6–14} Typically, such instruments register age, type of injury, repeated admission and consistency of medical history and injury.^{1 4} In the Netherlands, from 1996 onwards, the SPUTOVAMO checklist was increasingly used in ERs to detect physical abuse among children attending with

physical injury.⁹ As from 2007, when the Dutch Health Care Inspectorate formulated mandatory ER detection requirements, all hospital ERs use SPUTOVAMO as the standard detection checklist in all children.¹⁵ Local versions were developed¹⁴ with other items, such as evaluation of interactive behaviour of child and caregivers, to suit the checklist for detection of other types of abuse such as neglect, or for detection of need for help in general. Nationwide use of SPUTOVAMO did strongly increase the numbers of potential child abuse.¹⁵

An urgent problem with the use of diagnostic tools for physical child abuse, including SPUTOVAMO, is that comprehensive evidence on their accuracy is lacking.^{1 5} There are general recommendations, such as from the American Academy of Pediatrics, to optimise paediatricians' skills and examinations,¹¹ but the basis for additional use of checklists is currently not supported by evidence.¹⁶ This lack of evidence seems largely due to inherent difficulties in designing such diagnostic studies, such as the need for reference testing in children with negative screens, blinded evaluation of tests, and medical ethical and legal issues. Reference testing in checklist-negatives allows researchers to determine the negative predictive value of the checklist. The outcome must be established without knowledge of the index-test results. This blinded evaluation avoids so-called incorporation bias. Differential (non-) response is another issue with potential medical ethical and legal consequences.¹⁷

The Child Abuse Inventory at Emergency Rooms (CHAIN-ER) study aimed to assess among children presenting with physical injury at Dutch ERs, the diagnostic accuracy of this nationally implemented SPUTOVAMO checklist for physical abuse as primary outcome, and for neglect and need for help from social services as secondary outcome. Secondly, CHAIN-ER assessed if consulting of a 'child abuse paediatrician' (CAP) is a safe and efficient strategy if this consultation is limited to children with positive screens.

CHAIN-ER aimed to comply with published and widely accepted diagnostic study quality criteria (Quality Assessment of Diagnostic Accuracy Studies, QUADAS),¹⁸ and results are here reported by also widely accepted criteria Standards for the Reporting of Diagnostic Accuracy Studies (STARD).¹⁹

METHODS

Participants

All children aged 0–7 years admitted to an ER between June 2009 and December 2010 for any physical injury were included. When the initial symptom was not an injury, but when the trauma became clear during the ER visit, the child was also included. One academic and three non-academic hospitals in the region of Utrecht, a city in the centre of the Netherlands, participated. Evident victims of physical child abuse (admitted by perpetrator at presentation), victims of (witnessed) traffic

accidents and children who had died before arrival were excluded. We restricted our study to young children, for whom self-disclosure is unlikely,¹¹ and for whom ERs are among the few places where abuse can be detected since school attendance is not yet obligatory.

The following characteristics of participants were registered: age; sex; an indication of ethnic origins categorised as North European and non-North European, by the surname of the child; socioeconomic status (low/other), as defined by residing in risk areas (ie, with a low socioeconomic status) through classifications of national zip codes from the Dutch governmental 'Netherlands Institute for Social Research' (<http://www.scp.nl/english/>), based on average income, employment and educational level; previous ER visits in same hospital; time of present visit; type of present injury; mechanism of injury; injury severity.

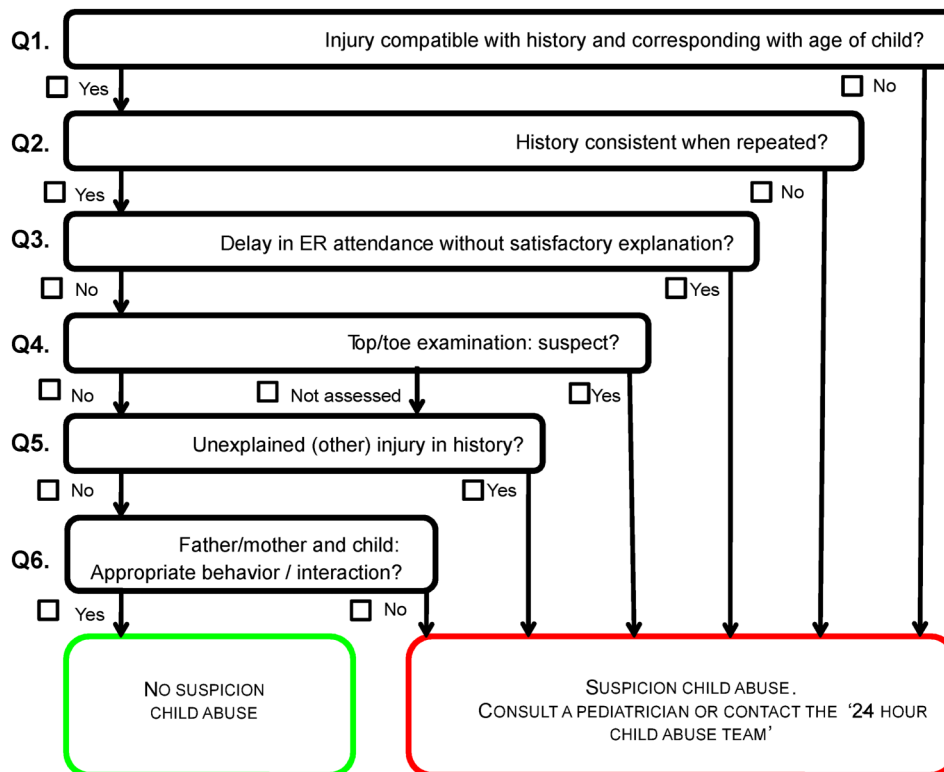
Checklist

All included patients were screened by the SPUTOVAMO-R checklist (index test), the revised version of the original,⁹ containing six questions with yes/no answer options (figure 1), further referred to as the checklist. The checklist classifies positive for suspected child abuse if at least one question scores abnormal. The checklist was a compulsory field in the electronic files of the medical records of all attendees, regardless of the reason of ER attendance. In the clinical process, for every child an ER nurse or physician fills out the checklist directly after clinical examination. A positive screen is followed by a systematic workup starting with an instant paediatric consultation in the ER. All positive cases are evaluated and eventually rejected or substantiated by the so-called Child Abuse Assessment Team, which is a team of several paediatricians and other professionals specialised in child abuse.²⁰ If necessary, the child and parents will be referred to social services.²⁰ The detailed clinical process differs across the four participating hospitals. The outcome of the clinical process is not used in this diagnostic study.

Outcome definition

The primary outcome was injury due to physical abuse by a parent or other caregiver, defined as 'use of physical force or implements against the child that has resulted in physical injury'.²¹ Because physical abuse contributes relatively few cases of child maltreatment that receive child protection intervention and may distract clinical attention from detecting neglect and emotional abuse, which are far more common and, in the long term, also damaging,²² we decided to include injury caused by neglect, and need for help from social services, as secondary outcomes. We evaluated neglect as 'failure to meet a child's basic physical needs or failure to ensure a child's safety'.²¹ We defined need for help from social services as any concern about the situation of the child that requires consultation of social services. Need for help thus included the cases of injury due to physical

Figure 1 SPUTOVAMO-R (index test).



abuse and injury caused by neglect. Need for help from social services is a very generic and summarising description of all possible help from social services, including consultation of prevention or intervention resources such as pedagogical support institutions and community programmes.

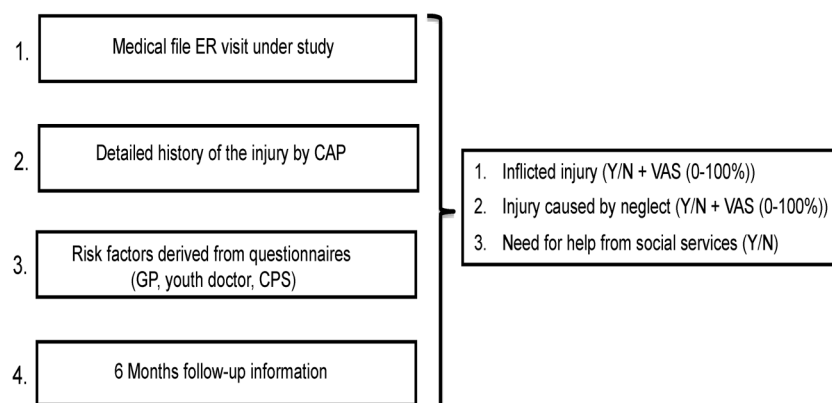
Reference standard procedure

The reference standard procedure was carried out in all children with positive screens and a computer-generated random sample of 15% of children with negative screens. In the absence of objective reference tests as with child abuse, diagnostic research guidelines advise consensus diagnosis by expert panels.^{23–26} The checklist was tested against the majority opinion of a three-member expert panel of cumulative 6-month diagnostic information presented in a structured anonymous paper file format for independent assessment by each panel

member. Diagnostic information was provided by the following consecutive steps (figure 2).

1. All clinical information about the ER visit, including available radiological images, which were evaluated by a child-radiologist specialised in skeletal imaging of suspected intentional physical injuries, blinded to the checklist outcome. Standard radiological imaging for research purposes only was both unfeasible and unethical.
2. Detailed report of additional semistructured history-taking regarding the injury by experienced CAPs based on telephone interviews of parents/caregivers shortly (ie, within 2 weeks) after the ER visit, and on full access to medical records, including the checklist outcome. CAPs are experienced paediatricians with certified child abuse recognition skills, that work according to standardised approaches.²⁰ These CAPs scored physical child abuse and child neglect by the

Figure 2 Different steps of the diagnostic outcome assessment by consensus of the expert panel (reference standard).



same definitions as used by the expert panel (see below).

- Information from healthcare professionals about risk factors for child abuse. Within a month after the ER visit, research nurses requested general practitioners (GPs) and youth doctors of all children with positive screens, and the 15% sample of the children with negative screens, to fill out a structured questionnaire on child, parental and environmental risk factors of child abuse (see appendix). In the same period, the nurses checked for registrations at the Child Protection Services, a national council under the Dutch Ministry of Safety and Justice. Research nurses were aware of the checklist outcome.
- After 6 months, the same research nurses as mentioned at step 3 checked the electronic patient file for additional clinically relevant information, including later ER visits, and summarised this in the expert panel paper file.

Expert panel

Prior to the reference standard procedure, definitions of the outcome measures were clarified in joined sessions to three paediatrician panel members, each nationally acknowledged clinical experts with forensic experience on child abuse. Throughout this reference standard procedure, panel experts were kept blinded to the checklist result²⁴ by deleting that information from steps 2, 3 and 4 by research nurses.

Members independently assessed whether the injury was inflicted (yes/no), and what they thought was the likelihood of the inflicted injury, using a continuous visual analogue scale (VAS) of 0–100%. For this analysis, we defined a positive case of inflicted injury when the likelihood was >50%. We used this 50% likelihood as the cut-off value. The cut-off is arbitrary, but did reflect our caution to keep the risks of both false positive cases and false-negative cases as low as possible, as a priori we considered avoiding each as equally important. The panel members also judged whether the injury was the result of neglect (yes/no) and the likelihood of neglect as cause of the injury, again using a continuous VAS of 0–100% (defined positive at >50%), and whether there was a need for help from social services in this family (yes/no).

For all outcomes, a case was considered positive by majority decision. When a file was incomplete because of missing information, we requested judgment based on all available information.

Child Protection Services child abuse reports

Additional to reference testing, child abuse risk was assessed in all participants with data from the national Child Protection Services registry (all abuse occurring from ER visit up to 4 years after the initial ER visit). This served as a crude overall check on the actual value of a positive screen result. Data were collected in January and February 2014, and merged with the CHAIN-ER

data on an aggregate, non-identifiable basis, by positive or negative checklist outcome.

Informed consent

Specific permission was granted to take oral informed consent only. Written information was given during the ER visit. Further information about CHAIN-ER was given, and informed consent requested, by a specially trained research nurse in a telephone call to parents shortly after the ER visit. A translator was available at request. Informed consent was separately asked for three study steps: extra history-taking by the CAP; contacting the GP, youth doctor and Child Protection Services for information about risk factors; and anonymous processing and evaluation of obtained data. Children with parents refusing anonymous processing and evaluation of data were excluded from the study.

Statistical analysis

Participants' general characteristics were described as means or proportions with corresponding dispersion measures.

For our primary objective to establish the diagnostic accuracy of the checklist, checklist results were evaluated against majority expert panel diagnoses for each outcome, with calculation of predictive values (PVs) and sensitivity and specificity with 95% exact binomial CIs.

For our secondary objective, to assess the value of having CAPs assessing children with positive screens, we assessed the agreement between the physical abuse diagnoses by CAPs and the expert panel.

Finally, child abuse reports to the Child Protection Services were assessed for all CHAIN-ER participants for up to January 2014, and aggregated by the ER checklist outcome categories. A risk ratio for being reported at the Child Protection Services with a positive checklist result as compared with a negative checklist result was calculated with an approximate 95% CI.

Inter-rater agreement between each of the panel members was assessed by two-way intraclass correlation coefficient (ICC). ICC were classified according to arbitrary cut-off values as poor (<0.20), fair (0.21–0.40), moderate (0.41–0.60), good (0.61–0.80) or very good (0.81–1.00) agreement.

Analyses were performed using PASW statistics V.20.0 and STATA/SE V.11.0.

RESULTS

Totally, 4290 children were eligible for inclusion (figure 3). Of these, 112 (2.6%) had a positive screen. With the 15% random sample (n=645) of children with a negative screen, parents of a total of 757 children were asked to participate in the reference standard procedure.

Figure 3 illustrates that the parents of 37 of the 757 children refused to participate in the reference standard procedure, with slightly more refusers among the parents of children with a positive screen (10.7% vs

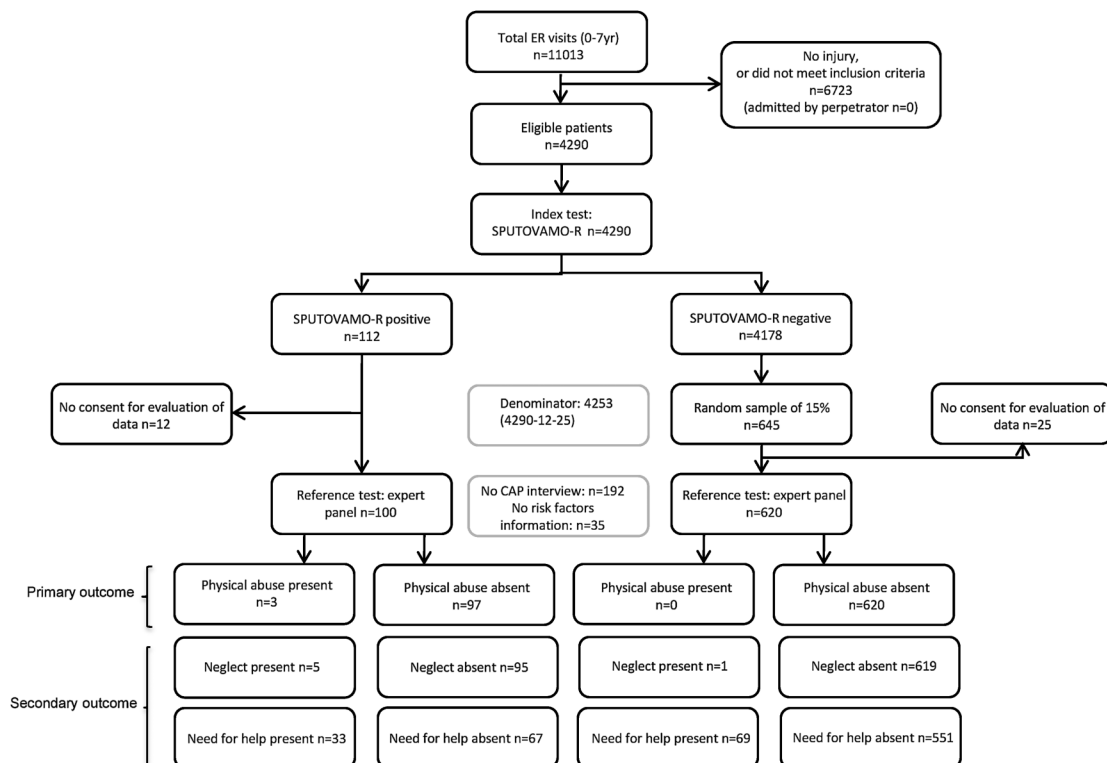


Figure 3 Flow chart inclusion Child Abuse Inventory at Emergency Rooms.

(3.9%), difference not significant). Of these 37 children, we assessed the clinical outcome and found one child (with a positive screen) to have possibly inflicted injury. The remaining study sample thus included 720 children. Of these, the CAP could not interview the parents of 192 children (ie, 63 with a positive screen) about the injury (step 2), because of refusal (n=32) or repeatedly being unreachable (n=160), leaving 528/720 participants being evaluated by the CAP. Of these 720 children, 35 parents (ie, 8 with a positive screen) did not give permission for researchers to contact other healthcare professionals (step 3). The baseline characteristics of the 720 included children are shown in the online supplementary appendix table S1.

Based on the reference standard procedure information, the injury was considered inflicted in 3 of 4253

children, an overall prevalence of 0.07% (95% CI 0.01 to 0.2). Injuries were considered caused by neglect in six children, with an overall prevalence of 0.27% ((5 + (1 × 4153/620))/4253) (95% CI 0.15 to 0.49). Help from social services was considered needed in 102 children, with a prevalence of 11.6% ((33 + (69 × 4153/620))/4253) (95% CI 10.6 to 12.6).

Table 1 shows the diagnostic value of the checklist for inflicted injury, by panel decision on all study outcomes. For every 100 children with a positive screen, three were truly abused (positive PV of 0.03), 97 were truly not abused (false-positive rate of 0.97 (95% CI 0.915 to 0.994), and 0 were missed (false-negative rate of 0.0, 95% CI 0.0 to 0.006). All three children considered by the expert panel to have non-accidental injuries had a positive screen. Details of the three cases of inflicted

Table 1 Value of child abuse checklist for diagnosis of inflicted injury, injury due to neglect, and need for help from social services in children admitted with physical injury at ER

Index test		Expert panel			Injury due to neglect			Need for help		
		Inflicted injury		Total	Yes	No	Total	Yes	No	Total
Checklist	Positive	3	97	100	5	95	100	33	67	100
	Negative	0	620	620	1	619	620	69	551	620
	Total	3	717	720	6	714	720	102	618	720
	PPV	0.030 (0.006 to 0.085)			0.050 (0.116 to 0.113)			0.330 (0.239 to 0.431)		
	NPV	1.000 (0.994 to 1.000)			0.998 (0.991 to 1.000)			0.889 (0.861 to 0.912)		
	Sens	1.000 (0.292 to 1.000)			0.833 (0.359 to 0.996)			0.326 (0.234 to 0.423)		
	Spec	0.865 (0.837 to 0.889)			0.867 (0.840 to 0.891)			0.892 (0.864–0.915)		

ER, emergency room; NPV, negative predictive value; PPV, positive predictive value; Sens, sensitivity; Spec, specificity.

injury are given in the appendix. Similar results were found for the outcome injury caused by neglect. Higher positive PVs were found for need for help from social services but lower negative PVs.

Of all 49 children with a positive screen seen by CAPs, there was 100% (95% CI 93% to 100%) agreement between the CAPs' diagnosis and the expert panel diagnosis.

Table 2 shows the numbers of children with positive and negative screens, and their respective reports to the Child Protection Services up to January 2014. In all 4253 included children, there were a total of 70 reports for physical child abuse to the Child Protection Services, seven in children with a positive screen and 63 in children with a negative screen (risk ratio 4.61, 95% CI 2.14 to 9.95). For general child abuse, there were 203 reports, 15 in 100 children with a positive screen and 188 in 4153 children with a negative screen (risk ratio 3.31, 95% CI 2.03 to 5.39). CPS reported physical child abuse, and CPS reported general child abuse in this sample were 1.6% (70/4253) and 4.8% (203/4253), respectively.

Panel inter-rater agreement for inflicted injury was 0.82 (95% CI 0.80 to 0.84), for injury caused by neglect 0.07 (95% CI 0.02 to 0.11), and for need for help from social services 0.40 (95% CI 0.35 to 0.44).

DISCUSSION

Based on good agreement between expert panel members, only 0.07% of children aged 0–7 years presenting with injury at the ER had been physically abused. The easy-to-use child abuse detection checklist, which is routinely used in Dutch ERs, correctly classified all true-negative cases of physical abuse by expert panel decision. In addition, there was full agreement about inflicted injury between the expert panel and CAPs.

Strengths and limitations

A strong feature of our diagnostic study involving 4253 children aged 0–7 years, is that it is the first that meets all QUADAS criteria for diagnostic accuracy studies, whereas former diagnostic studies met only a minority of all 14 criteria.^{1 14} To date, no diagnostic studies of child abuse have used the same reference standard procedure for children with positive and negative screens.

Consequently, accurate PVs of a negative outcome of a checklist to detect child abuse were not available before.

Some issues need further consideration. Misclassification of the outcome might explain the low prevalence of inflicted injury. However, we used a narrow definition of inflicted injury, and we concur with methodological guidelines indicating that the majority opinion of a panel of experts is the best possible reference standard. Our panel members are regarded as national experts in the field of child abuse.

Moreover, they assessed the information mutually independently, thereby avoiding influence of opinions and group thinking processes.²⁷ Panel members were blinded to the results of the checklist, which avoided so-called incorporation bias.²⁴ Last, the patient files presented to panel members contained all relevant information, including X-ray imaging evaluated by an expert radiologist blinded to checklist results, and information about risk factors obtained from other healthcare professionals. However, because panel members rely on subjective information in the patient file, such as risk factor assessments, we cannot exclude a certain level of implicit bias.

Participation bias is not likely an explanation for the low prevalence of physical abuse. The parents of 37 of the 757 patients (4.9%) refused to participate in the reference standard procedure. Although the expert panel could not make a final diagnosis for the children of these parents, we assessed the clinical outcome for these patients and found only one patient (with a positive checklist outcome) to have possibly inflicted injury, which would not materially influence the low prevalence of physical abuse.

The 0.07% prevalence of true physical abuse in our study is much lower than the 1% prevalence reported by Woodman *et al*,¹ which is estimated based on the data of 66 studies. Differences in measurement, setting and methodology likely explain this difference. Based on our findings, we suspect that most studies have overestimated the prevalence of physical abuse, although we cannot exclude the possibility that there may truly be a wide range of prevalences.

Our finding of not being able to unequivocally diagnose injury due to neglect and need for help, renders prediction of this kind of injury with a checklist, such as SPUTOVAMO-R, inaccurate in young children. However, a checklist is not sufficiently accurate and should not replace skilled assessment by a clinician. Notwithstanding the importance of paying attention to the needs of children in general, we consider it questionable whether systematic screening for injury caused by neglect and need for help should be part of ER professionals' instrumentarium.

In conclusion, this checklist has a very high false-positive rate for physical abuse, which is unhelpful to services and families. For every 100 children that had been screened, only three will have been physically abused, and 33 will require referral to social services. Inaccurate

Table 2 Number of Child Protection Services physical abuse reports up to 4 years after last inclusion

Child Protection Services Reports		
Physical child abuse		
Checklist	Yes	Total
Positive	7	100
Negative	63	4153
Total	70	4253

suspicions may have huge impact on children and their families. Apparently, doing a second screening test on 2.6% of all children presenting with injuries presumably cost-ineffectively increases the workload for a CAP.

Clinical implications and conclusions

For settings where the easy-to-use checklists are widely implemented, we conclude that rare cases of inflicted injury among preschool children presenting at ERs for injury are very likely captured, but at very high false-positive rates. Subsequent assessment by child abuse experts can be safely restricted to children with positive screens at very low risk of missing any cases of inflicted injury.

However, where such checklists are not used yet, we do advise careful prior consideration of cost-effectiveness and clinical and societal implications before de novo implementation.

Author affiliations

¹Department of Paediatrics, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, The Netherlands

²Mental Health Care 'GGZ Centraal', Hilversum, The Netherlands

³Julius Centre for Health Sciences and Primary Care, University Medical Centre Utrecht, Utrecht, The Netherlands

⁴Department of Paediatric Radiology, Wilhelmina Children's Hospital, University Medical Centre Utrecht, Utrecht, The Netherlands

Acknowledgements The authors thank the patients and their parents who took part in this study. They are very grateful for the support and feedback provided by the CHAIN-ER project group: Ad Bosschaart, Albert Vermaas, Arend Groot, Ben Rensen, Chantal Assenbergh, Eline Ockhuijsen, Elske van Mierlo, Eric Hammacher, Erica Post, Frederique van Berkestijn, Ingrid Lukkassen, Iva Bicanic, Jenita Riphagen, Kiki Soetenga, Kirsten Korte, Koen Lansink, Laura Sandbergen, Linde Nijhof, Loudy Priesterbach (research nurse), Luuk Leenen, Marleen Vreeburg, Nan Hepkema, Noor Landsmeer, Piet van Lieshout, Ruth Gilbert, Sandra Rutgers, Sanna van Dam and Stefan Vestjens.

Contributors JSS is primary investigator and responsible for data collection and analysis and for drafting the manuscript. CSPMU, KGMM, EESN and EMP designed and supervised the study. EMP performed structured interviews and obtained funding for the study. IR was expert panel member. RAJN evaluated all available radiological images. JSS and CSPMU performed the data analysis. All authors have read and approved the final manuscript.

Funding This study was funded by the Netherlands Institution for Health Research and Development (ZonMw 80-82405-98-084, project number 15700.1008.) KGMM receives funding from the Netherlands Organization for Scientific Research (project 9120.8004 and 918.10.615).

Competing interests None declared.

Ethics approval Medical Ethics Committee of the University Medical Centre Utrecht.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The additional available data is given in the appendices.

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

REFERENCES

- Woodman J, Pitt M, Wentz R, *et al*. Performance of screening tests for child physical abuse in accident and emergency departments. *Health Technol Assess* 2008;12:iii, xi-xiii 1-95.
- Gilbert R, Widom CS, Browne K, *et al*. Burden and consequences of child maltreatment in high-income countries. *Lancet* 2009;373:68-81.
- Norman RE, Byambaa M, De R, *et al*. The long-term health consequences of child physical abuse, emotional abuse, and neglect: a systematic review and meta-analysis. *PLoS Med* 2012;9:e1001349.
- Gilbert R, Kemp A, Thoburn J, *et al*. Recognising and responding to child maltreatment. *Lancet* 2009;373:167-80.
- Louwers EC, Affourtit MJ, Moll HA, *et al*. Screening for child abuse at emergency departments: a systematic review. *Arch Dis Child* 2010;95:214-18.
- Benger JR, Pearce V. Simple intervention to improve detection of child abuse in emergency departments. *BMJ* 2002;324:780.
- Chang DC, Knight VM, Ziegfeld S, *et al*. The multi-institutional validation of the new screening index for physical child abuse. *J Pediatr Surg* 2005;40:114-19.
- Clark KD, Tepper D, Jenny C. Effect of a screening profile on the diagnosis of nonaccidental burns in children. *Pediatr Emerg Care* 1997;13:259-61.
- Compennolle T. [Eerste hulp bij kindermishandeling en het Strausarts-syndroom]. *Tijdschr Kindergeneeskd* 1996;64:168-79.
- Hornor G. Child maltreatment: screening and anticipatory guidance. *J Pediatr Health Care* 2013;27:242-50.
- Kellogg ND. Evaluation of suspected child physical abuse. *Pediatrics* 2007;119:1232-41.
- Pless IB, Sibald AD, Smith MA, *et al*. A reappraisal of the frequency of child abuse seen in pediatric emergency rooms. *Child Abuse Negl* 1987;11:193-200.
- Sidebotham PD, Pearce AV. Audit of child protection procedures in accident and emergency department to identify children at risk of abuse. *BMJ* 1997;315:855-6.
- Louwers EC, Korfage IJ, Affourtit MJ, *et al*. Accuracy of a screening instrument to identify potential child abuse in emergency departments. *Child Abuse Negl* 2014;38:1275-81.
- Louwers EC, Korfage IJ, Affourtit MJ, *et al*. Effects of systematic screening and detection of child abuse in emergency departments. *Pediatrics* 2012;130:457-64.
- Teeuw AH, Derxk BH, Koster WA, *et al*. Educational paper: detection of child abuse and neglect at the emergency room. *Eur J Pediatr* 2012;171:877-85.
- Sittig JS, Uiterwaal CS, Moons KG, *et al*. Child abuse inventory at emergency rooms: CHAIN-ER rationale and design. *BMC Pediatr* 2011;11:91.
- Whiting P, Rutjes AW, Reitsma JB, *et al*. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 2003;3:25.
- Bossuyt PM, Reitsma JB, Bruns DE, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. The Standards for Reporting of Diagnostic Accuracy Group. *Croat Med J* 2003;44:635-8.
- Sittig JS, Post ED, Russel IM, *et al*. Evaluation of suspected child abuse at the ED: implementation of American Academy of Pediatrics guidelines in the Netherlands. *Am J Emerg Med* 2014;32:64-6.
- Leeb RT, Paulozzi L, Melanson C, *et al*. *Child maltreatment surveillance. Uniform definitions for public health and recommended data elements*. Atlanta: Centers for Disease Control and Prevention, 2008.
- https://http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/469737/SFR41-2015_Text.pdf 2015.
- Bertens LC, Broekhuizen BD, Naaktgeboren CA, *et al*. Use of expert panels to define the reference standard in diagnostic research: a systematic review of published methods and reporting. *PLoS Med* 2013;10:e1001531.
- Moons KG, Grobbee DE. When should we remain blind and when should our eyes remain open in diagnostic studies? *J Clin Epidemiol* 2002;55:633-6.
- Naaktgeboren CA, de Groot JA, van Smeden M, *et al*. Evaluating diagnostic accuracy in the face of multiple reference standards. *Ann Intern Med* 2013;159:195-202.
- Reitsma JB, Rutjes AW, Khan KS, *et al*. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 2009;62:797-806.
- Esser JK. Alive and well after 25 years: a review of groupthink research. *Organ Behav Hum Decis Process* 1998;73:116-41.