Child Psychiatry Hum Dev
DOI 10.1007/s10578-011-0254-1

ORIGINAL ARTICLE

# Children's Depression Screener (ChilD-S): Development and Validation of a Depression Screening Instrument for Children in Pediatric Care

**Barbara Frühe · Antje-Kathrin Allgaier · Kathrin Pietsch · Martina Baethmann · Jochen Peters · Stephan Kellnar · Axel Heep · Stefan Burdach · Dietrich von Schweinitz · Gerd Schulte-Körne**



**Abstract** The aim of the present study was to develop and validate the Children's Depression Screener (ChilD-S) for use in pediatric care. In two pediatric samples, children aged 9–12 ($N_I = 200$; $N_{II} = 246$) completed an explorative item pool (subsample I) and a revised item pool (subsample II). Diagnostic accuracy of each of the 22 items from the revised pool was evaluated in order to select the best items for the brief instrument ChilD-S. Areas under the curve (AUCs) of the revised item pool and the ChilD-S were compared. A diagnostic interview, the Kinder-DIPS, served as gold standard. For the purpose of screening for depressive disorders in children, the eight-item ChilD-S (AUC = 0.97) performed just as well as the revised 22-item pool (AUC = 0.94). For the ChilD-S the optimal cut-off point of $\geq$11 yielded a sensitivity of 0.91 and a specificity of 0.89. The ChilD-S shows high potential for depression screening of children in pediatric care.

**Keywords** Depression · Children · Screening · Validation · Pediatric care

Barbara Frühe and Antje-Kathrin Allgaier contributed equally to this paper.

B. Frühe · A.-K. Allgaier · K. Pietsch · G. Schulte-Körne (✉)
Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital, Ludwig-Maximilians-University Munich, Nussbaumstraße 5a, 80336 Munich, Germany
e-mail: gerd.schulte-koerne@med.uni-muenchen.de

M. Baethmann · J. Peters
Department of Pediatrics, Hospital "Dritter Orden", Munich, Germany

S. Kellnar
Department of Pediatric Surgery, Hospital "Dritter Orden", Munich, Germany

A. Heep
Department of Pediatrics, Hospital "Harlaching", Munich, Germany

S. Burdach
Department of Pediatrics, University Hospital, Technical University Munich, Munich, Germany

D. von Schweinitz
Department of Pediatric Surgery, University Hospital "Dr. von Haunersches Kinderspital", Ludwig-Maximilians-University Munich, Munich, Germany

Published online: 17 September 2011

## Introduction

Depression is one of the most frequent psychiatric disorders across the lifespan: as many as 2.8% of children under the age of 13 meet diagnostic criteria for a depressive disorder as reported by a meta-analysis of epidemiological studies [1]. Child depression affects behavioural, emotional and academic development [2–4]. This is usually accompanied by a lack of social skills, more interpersonal conflicts, higher rates of school absenteeism, and early school dropout [5, 6]. Without treatment, depression in children is associated with a negative outcome and a high risk of symptom progression, recurrence, chronicity and comorbid mental disorders [7]. For example, Harrington et al. [8] reported a 40% recurrence rate for children with first manifestation of depression under the age of 14 over a five year period. Despite the negative consequences of childhood depression and the availability of effective treatment, depressive disorders often remain undiagnosed. A study in pediatric clinics showed, that only 22% of affected patients were recognized [9]. Many pediatricians report a high degree of discomfort with the diagnosis of children's psychiatric disorders [10–12]. This is remarkable especially for depression, since the prevalence rates are higher for children diagnosed with medical conditions [13, 14]. Depression rates have been reported at 14% among children and adolescents with cancer [15], 15% among youths with asthma [16], 23% among children with orthopaedic procedures [17], and 26% among youths with burn injuries [18]. Pediatricians may feel uncomfortable about diagnosing depression, since the disorder is difficult to detect especially in children due to its heterogeneous symptomatology. In addition to depressed mood and irritability, children with depressive disorders also suffer from concentration difficulties, loss of appetite or unspecific physical symptoms such as abdominal pain or headaches [19]. To further compound the difficulty in detecting depression is the lack of time afforded to pediatricians for extensive exploration due to high patient load [20].

Nevertheless, pediatricians play a pivotal role in early identification of depressive disorders [10, 21, 22], since they are usually the first point of contact for affected children and their families. If there is an indication of a depressive disorder, it is the task of the pediatrician to refer the family to the mental health service [22]. That is why pediatricians have a crucial role in facilitating entry into mental health service and why they have a significant impact on the course of the disorder. One way to improve detection rates by physicians is through the use of self-report screening tools [23]. The use of self-report instruments is an accepted method for screening of depressive symptoms in children and adolescents [24–27]. Positive scores on the instrument suggest the need for an extensive exploration of the depressive symptomatology. For screening tools to be accepted in clinical settings, the brevity of the instrument and high criterion validity are important [28]. Therefore, in busy pediatric care a brief and economical instrument is vital. Furthermore, high sensitivity is essential in order to maximize the detection rate, so that as few as possible cases are missed. But at the same time specificity is an important criterion, so that only a few patients are screened as false positive. Self-report instruments for screening depression in children are often used in the area of research. However, criterion validity is reported in comparatively few studies. In one of these few studies, Sorensen et al. [29] analysed the established 27-item Children's Depression Inventory (CDI) [30, 31] in a sample of psychiatric patients aged 8–13 with respect to the gold standard Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) [32] resulting in a sensitivity of 0.63 and a specificity of 0.86. Similarly, Craighead and colleagues [33] validated the CDI in a psychiatric sample of 12- to 18-year-old adolescents by comparing it to the K-SADS-Epidemiological Version [34, 35] and found a sensitivity of

0.81 and a specificity of 0.84. Shemesh et al. [36] also tested the CDI in a pediatric sample of 8- to 19-year-old patients against the K-SADS-PL, resulting in a sensitivity of 0.80 and a specificity of 0.70. The studies of the CDI show diversely fitting values of sensitivity and specificity and thus it is unclear how the measure would perform specifically with pediatric populations. Furthermore, using a 27-item questionnaire as a screening instrument in busy pediatric settings may not be feasible, since time for completing and evaluating the instrument is limited. To address this problem a 10-item CDI Short Form (CDI:S) [31] was developed and is suggested for use for screening in pediatric care. Unfortunately, the CDI:S has not been validated for its use in pediatric settings nor for epidemiological purposes until now. Other instruments such as the 33-item Mood and Feelings Questionnaire-Child Version (MFQ-C) [37] share limitations similar to those noted for the CDI. One investigation of the MFQ-C [38] with both clinical and non-clinical samples of children and adolescents found a sensitivity of 0.68 and a specificity of 0.88 when tested against the K-SADS-PL. A shortened 13-item screening version (SMFQ-C) was validated in an epidemiological sample of 7- to 11-year-old children by Angold and colleagues [39]. When the short version was tested, a sensitivity of 0.60 and a specificity of 0.85 was found compared to the Diagnostic Interview Schedule for Children (DISC) [40]. However, sensitivity of the 33-item as well as for the 13-item MFQ-C are not sufficient.

In summary, to the best of our knowledge, no child-specific depression screening instrument is available today that fulfils the criteria of brevity as well as high sensitivity and high specificity. Therefore, the development of a brief and valid screening instrument for children would be helpful in aiding pediatricians in the identification of depressive symptoms among children. Thus, the aim of our study was to design and validate a depression screening instrument specifically for prepubertal in- and outpatients in pediatric care. The development took part in two steps: In step I, an initial explorative item pool was composed and then validated in a sample of 200 children. In step II, this item pool was revised and validated again in a sample of 246 children. Finally this revised item pool was abbreviated, resulting in the so-called Children's Depression Screener (ChilD-S).

## Method

### Procedure

The multi-centre study was performed between September 2009 and November 2010 in six pediatric and pediatric surgery clinics in Munich, Germany. Data collection took place sequentially in time units of three months for each clinic. The study protocol was approved by the local ethics committees of the Ludwig-Maximilians-University Munich, of the Technical University Munich and of the Bavarian Medical Association.

Every workday, patients newly admitted to the hospital and their parents were invited to participate in the study. Inpatients and their parents were personally informed in the sick rooms while outpatients and their parents were personally informed in the waiting rooms. The research team handed out information sheets that explained the nature of the research and the requirements of the participants. Furthermore, the patients and their parents were assured that refusal to participate would not impact their medical care. When written informed consent was given, a questionnaire consisting of the explorative or revised item pool was completed by the children. Moreover, children as well as parents participated in a diagnostic interview. While the child interview was always conducted face to face, the

parent had the option of completing the interview by telephone within one week after the child's participation. Subjects obtained a gift voucher of 20 Euros.

## Data Collection and Sample Description

In the development process of the ChilD-S an explorative item pool was analysed in step I, using subsample I. In step II, first the revised item pool and then a final brief version (ChilD-S) were evaluated, using subsample II. In Fig. 1, the process of data collection is illustrated.

The following inclusion criteria had to be fulfilled by 9- to 12-year-old patients so that they could complete the screening instrument autonomously: (1) adequate intellectual capacity, (2) satisfactory German language skills and (3) a sufficient health condition. To judge the health condition pediatricians were asked whether the child is able to concentrate for half an hour despite the illness or not. In some cases the children were seriously ill and too weak and so they were not able to participate in the study. The majority of children was fit enough to participate and pleased with an alternative occupation during their long

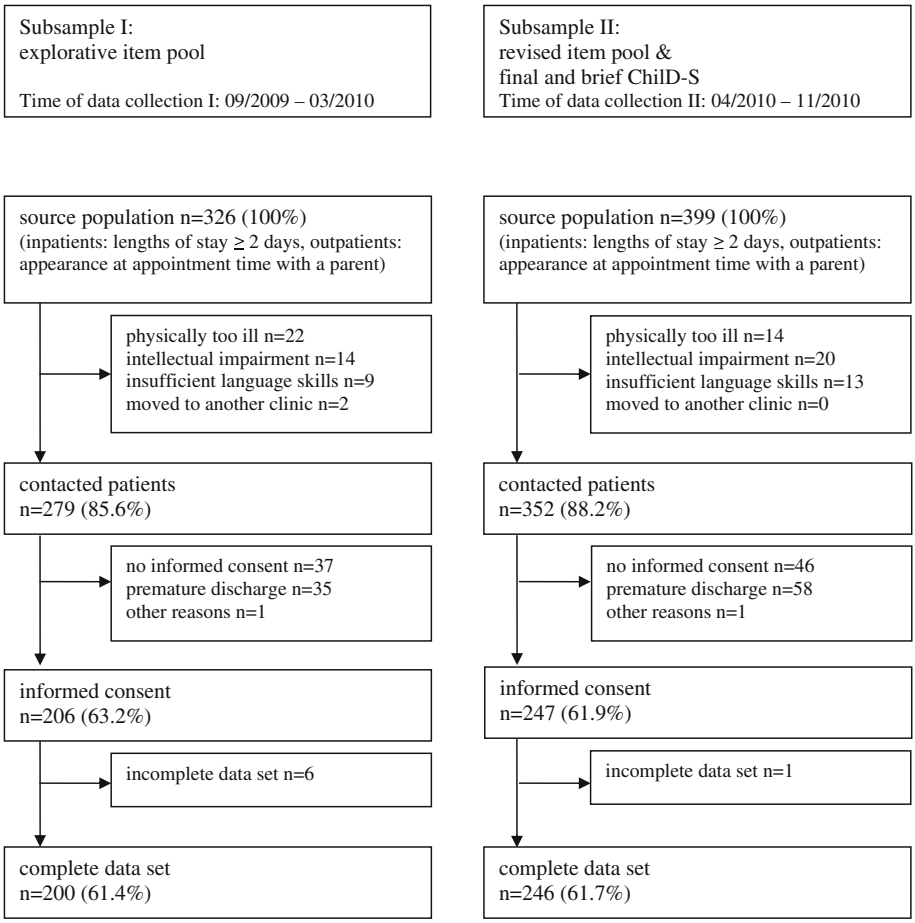


**Fig. 1** Flow chart of data collection

hospital day. For recruitment of inpatients, a minimum stay of two days was required. For outpatients, both patient and parent were required to appear for a medical appointment.

*Subsample I*

In the first recruitment wave, 200 complete data sets were acquired (Fig. 1). The mean age of participating children was 10.48 years (SD 1.15), 42.5% were female and 57.5% were male. Dropouts and included patients did not differ significantly with respect to age ($t = 0.839$; $p = 0.402$) or sex ($\chi^2 = 3.755$; $p = 0.053$).

Within the group of participating children, 68.5% were inpatients and 31.5% were outpatients. A high proportion (72.0%) of the participants came to the hospital with pediatric illnesses (e.g. chronic diseases or infections); the remaining 28% of participants were utilizing pediatric surgery services.

Point-prevalence of 'any depressive disorder' according to DSM-IV-TR [41] was 6.0%: 4.5% fulfilled the criteria for major depression and 2.0% met the criteria for dysthymia (Table 1).

*Subsample II*

In the second recruitment wave, 246 complete data sets were acquired (Fig. 1). The mean age of participating children was 10.54 years (SD 1.07), 43.1% were female and 56.9% were male. Dropouts and included patients did not differ significantly with respect to age ($t = -0.469$; $p = 0.640$) or sex ($\chi^2 = 0.267$; $p = 0.605$).

Of the participating children, 85.0% were inpatients and 15.0% were outpatients. As in the first sample, admission to the hospital was for pediatric illnesses (52.5%) and for pediatric surgery (47.5%).

Point-prevalence of 'any depressive disorder' according to DSM-IV-TR [41] was 4.5%: 3.7% fulfilled the criteria for major depression and 1.2% met the criteria for dysthymia (Table 1).

*Differences in Subsample I and Subsample II*

Subsample I and II did not differ significantly with respect to age ($t = -0.537$; $p = 0.591$), sex ($\chi^2 = 0.016$; $p = 0.900$) or point-prevalence of 'any depressive disorder' ($\chi^2 = 0.527$; $p = 0.468$). Differences between subsample I and II were identified with

**Table 1** Diagnosis in subsample I and II according to DSM-IV-TR

| | Subsample I n = 200 (%) | Subsample II n = 246 (%) |
|---|---|---|
| Any depressive disorder[a] | 12 (6.0)[b] | 11 (4.5)[b] |
| Major depression total | 9 (4.5) | 9 (3.7) |
| Mild | 6 (3.0) | 8 (3.3) |
| Moderate | 1 (0.5) | 1 (0.4) |
| Severe | 2 (1.0) | – |
| Dysthymia | 4 (2.0) | 3 (1.2) |

[a] Any depressive disorder was classified as major depressive disorder or dysthymia

[b] One case with double depression

regard to reason for admission and to status as in- or outpatient: In subsample II significantly more patients were admitted for surgery ($\chi^2 = 31.209$; $p < 0.001$) and significantly more inpatients ($\chi^2 = 18.114$; $p < 0.001$) were included.

## Measures

### *Gold Standard: Kinder-DIPS*

Gold standard for validation were diagnoses based on the module 'affective disorders' of the German structured diagnostic interview for mental disorders in children and adolescents (Kinder-DIPS) [42]. This interview allows diagnostic criteria to be thoroughly assessed, corresponding to ICD-10 [43] as well as DSM-IV-TR [41]. Test–retest reliability for all DSM-IV diagnoses is high (Cohen's $\kappa = 0.85$–$0.94$) [44].

In our study, 'any depressive disorder' was classified as major depressive disorder or dysthymia according to DSM-IV-TR criteria. The interview consisted of a parent and a child interview. Both sources of information were considered for diagnostic decisions. A diagnosis was assigned if either the child or parent endorsed a sufficient number of symptoms to meet diagnostic criteria for the disorder, a method recommended in the Kinder-DIPS manual. In the study, two trained psychologists administered and interpreted the interviews. Interrater reliability was high (Cohen's $\kappa = 0.90$). The interviewers were blind to screening results, as well as to the results of the child or parent interview, respectively.

### *Screening Instrument: ChilD-S*

The new Children's Depression Screener (ChilD-S) was developed to offer a valid and brief self-report screening instrument for prepubertal in- and outpatients in pediatric care. The item pool was especially designed for children between 9 and 12 years of age and measures how participants felt for the past two weeks.

*Step I*: In order to compose an item pool, established English depression questionnaires and their German versions for the concerning age group were reviewed including the Children's Depression Inventory (CDI) [30, 31, 45], the Center for Epidemiological Studies Depression Scale for Children (CES-DC) [46, 47], the Reynolds Adolescents Adjustment Screening Inventory (RAASI$^{TM}$) [48, 49], and the Youth Self Report (YSR) [50, 51]. Items were clustered to symptom groups and selected based on discriminatory power. Out of the items of the particular symptom group, the items with the highest discriminatory power were chosen. These items were rephrased in a simple wording and adapted to our response format. As a result, a pool of 22 newly formulated items was composed, which covers the heterogeneous symptomatology of childhood depression. The pool consisted of items typical for major depression according to DSM-IV-TR [41] as well as age specific items not listed in the diagnostic criteria. Wording was tested with 15 children and afterwards fine-tuned. Positively and negatively formulated items were balanced. A four-point Likert scale was chosen. The sequence of response categories was "disagree" (0) on the first position and "agree" (3) on the last position. High scores indicate a depressive state.

The evaluation resulted in a Cronbach's alpha coefficient of 0.86. Indices of discriminatory power ranged between 0.31 and 0.64, with three items below the critical value of 0.30 ('loss of interest' ($r_{it} = 0.14$), 'sense of guilt' ($r_{it} = 0.17$), and 'irritability' ($r_{it} = 0.28$)). Area under the curve values (AUCs) for single items ranged between 0.61

and 0.76. Five items had insufficient AUCs $\leq$0.60 ['loss of interest' (AUC = 0.53), 'concentration difficulties' (AUC = 0.56), 'stomach ache' (AUC = 0.56), 'feeling offended' (AUC = 0.56), and 'irritability' (AUC = 0.59)].

*Step II*: Based on findings of step I, four items were rephrased for a revised version since wording was probably too complicated (new items: 19 'concentration difficulties', 20 'irritability', 21 'loss of interest', and 22 'sense of guilt'). The other two items with critical values (16 'stomach ache' and 14 'feeling offended') were not modified as they were clearly formulated and had to be evaluated again in subsample II. In order to facilitate the answer choice for children, the sequence of response categories was changed, beginning with affirmation ("agree" 3) and ending with negation ("disagree" 0). This modification was made because several children reported that it was difficult to start answering with a negation of the item.

The revised version with 22 partially rephrased items and with modified scaling was validated. Thereof, items with the highest diagnostic accuracy were chosen for the final brief screening instrument ChilD-S.

### Statistical Analyses

The psychometric properties of the items (i.e. mean scores, standard deviation, and discriminatory power) were computed. The Cronbach's alpha coefficient was calculated so as to check for reliability. A t-test for independent samples was computed in order to compare the average sum scores of depressed and non-depressed subjects. If 20% or less of the items on the ChilD-S were missing, the missing values were replaced using the average score of the completed items of the respective subject. If more than 20% of items were missing, the respective subjects were omitted, which was only necessary in two cases. Therefore, analysis of differences between participants who answered less than 80% than those that completed at least 80% of the instrument was not possible.

To evaluate the validity of the revised item pool as well as for the ChilD-S, receiver operating characteristics (ROC) analyses with corresponding area under the curve values (AUCs) were computed. Therefore, the diagnosis of 'any depressive disorder' served as gold standard. To select items with high diagnostic accuracy for the ChilD-S, AUCs for each item out of the revised item pool were calculated. Ferdinand [52] interpreted the diagnostic accuracy with an AUC of less than 0.70 as poor, an AUC of 0.70–0.80 as fair, an AUC of 0.80–0.90 as good, and an AUC greater than 0.90 as excellent. Overall diagnostic accuracies of the revised item pool and the final brief ChilD-S were compared using a non-parametric measure for correlated samples [53]. For several cut-off points, sensitivity and specificity as well as positive and negative predictive values were investigated and the corresponding Youden-Indices were considered. The comparison of sensitivity and specificity was computed using the McNemar test. Data were processed with the statistical software SPSS (Statistical Package for Social Sciences, version PASW Statistics 18) for Windows and R (Software for Statistical Modelling & Computing, version R 2.11.1).

## Results

### Psychometric Properties and Diagnostic Accuracy of the Revised Item Pool

Psychometric properties (mean scores, standard deviation, and discriminatory power) and diagnostic accuracy (AUCs) of each item are shown in Table 2. Mean scores of the revised

**Table 2** Psychometric properties and diagnostic accuracy of the revised item pool

| No. | Item | Psychometric properties | | Diagnostic accuracy |
|---|---|---|---|---|
| | | M (±SD) | $r_{it}$ | AUC |
| 1. | **I am happy** | **0.54 (±0.64)** | **0.56** | **0.76** |
| 2. | I think nobody really likes me | 0.30 (±0.63) | 0.35 | 0.70 |
| 3. | **I am doing fine** | **0.45 (±0.74)** | **0.46** | **0.79** |
| 4. | I like a lot of things | 0.32 (±0.58) | 0.40 | 0.69 |
| 5. | I feel lonely | 0.39 (±0.75) | 0.60 | 0.77 |
| 6. | **I feel exhausted by everything** | **0.63 (±0.78)** | **0.54** | **0.72** |
| 7. | I like myself | 0.25 (±0.51) | 0.33 | 0.53 |
| 8. | I feel like crying | 0.43 (±0.76) | 0.48 | 0.66 |
| 9. | I feel as fit as always | 0.95 (±0.91) | 0.43 | 0.74 |
| 10. | **I worry a lot** | **0.89 (±0.98)** | **0.58** | **0.80** |
| 11. | I enjoy a lot of things | 0.40 (±0.68) | 0.45 | 0.59 |
| 12. | **I feel sad** | **0.58 (±0.95)** | **0.72** | **0.93** |
| 13. | I sleep just as always | 0.81 (±1.04) | 0.35 | 0.74 |
| 14. | I am offended easily | 0.97 (±1.02) | 0.41 | 0.73 |
| 15. | I play as much as usual with other children | 0.64 (±0.86) | 0.34 | 0.62 |
| 16. | My stomach often hurts | 0.87 (±1.09) | 0.47 | 0.60 |
| 17. | I am more anxious than other children | 0.54 (±0.81) | 0.28 | 0.61 |
| 18. | My appetite is the same as usual | 0.81 (±1.06) | 0.47 | 0.66 |
| 19. | It's often hard for me to concentrate | 0.98 (±0.99) | 0.46 | 0.63 |
| 20. | **I get upset quickly** | **1.14 (±1.08)** | **0.52** | **0.90** |
| 21. | **I am not in the mood for anything** | **0.44 (±0.74)** | **0.56** | **0.80** |
| 22. | **I often think I did something wrong** | **0.84 (±0.95)** | **0.55** | **0.80** |

Items of the screening instrument ChilD-S are printed in bold

*M* mean scores (range 0–3), *SD* standard deviation, $r_{it}$ discriminatory power, *AUC* area under the curve

22-item pool were low, with values between 0.25 and 1.14. Average of sum scores differed significantly between the depressed (M = 31.64; SD = 7.23) and the non-depressed group (M = 13.23; SD = 9.03) (t = −6.654; $p < 0.001$). Discriminatory power had a wide range with indices between 0.28 and 0.72. Reliability of the revised item pool was good with a Cronbach's alpha coefficient of 0.88. Overall diagnostic accuracy was very high with an AUC of 0.94 (95% CI: 0.88–0.99) and differed significantly from the line of no information ($p < 0.001$). AUCs for single items ranged between 0.53 and 0.93.

## The Screening Instrument ChilD-S

### *Design of the ChilD-S*

To compose a valid screening instrument, items with high diagnostic accuracy were selected from the revised 22-item pool. To ensure feasibility, we aimed at a screening instrument not exceeding ten items.

Numerous item combinations were calculated and assessed regarding both content-related and statistical aspects. The best combination consisted of the following eight items:

'I am happy.' (AUC = 0.76), 'I am doing fine.' (AUC = 0.79), 'I feel exhausted by everything.' (AUC = 0.72), 'I worry a lot.' (AUC = 0.80), 'I feel sad.' (AUC = 0.93), 'I get upset quickly.' (AUC = 0.90), 'I am not in the mood for anything.' (AUC = 0.80), and 'I often think I did something wrong.' (AUC = 0.80). This combination resulted in an excellent overall diagnostic accuracy with an AUC of 0.97 (95% CI: 0.93–1.00). Thus, the AUC value of this eight-item ChilD-S did not differ from the AUC value computed for the revised 22-item pool (AUC = 0.94; 95% CI: 0.88–0.99) (z = 1.382; $p = 0.167$) (Fig. 2).

*Psychometric Properties*

The eight-item ChilD-S as reported above showed a satisfactory reliability with a Cronbach's alpha coefficient of 0.81. Mean scores ranged between 0.43 and 1.14. Indices for discriminatory power were between 0.39 and 0.71. Hence, all items of the ChilD-S are above the critical value of 0.30 and indicate a good item-scale-correlation. The average scores differed significantly between the depressed (M = 14.54; SD 2.62) and the non-depressed group (M = 5.10; SD 4.04) (t = −7.669; $p < 0.001$).

*ROC Analyses and Validity Measures*

Figure 2 shows the ROC curve and the AUC value of the ChilD-S tested against the gold standard of 'any depressive disorder'. For direct comparison, the ROC-curve and the AUC of the revised item pool are also included. In this figure, curves moving towards the upper left corner indicate a high AUC value and are associated with higher rates of true positives (high sensitivity) and higher rates of true negatives (high specificity).

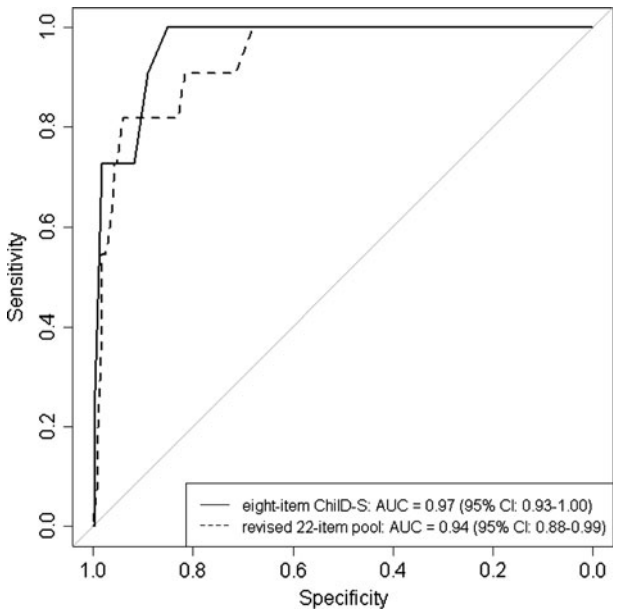


**Fig. 2** ROC-curves of the ChilD-S in comparison with the revised item pool

**Table 3** Cut-off points for the ChilD-S with corresponding validity measures

| Cut-off points | Sensitivity (95% CI) | Specificity (95% CI) | PPV (95% CI) | NPV (95% CI) | Youden (95% CI) |
|---|---|---|---|---|---|
| ≥9 | 1.00 (0.62–1.00) | 0.80 (0.74–0.85) | 0.19 (0.10–0.32) | 1.00 (0.97–1.00) | 0.80 (0.36–0.85) |
| **≥10** | **1.00 (0.62–1.00)** | **0.85 (0.80–0.90)** | **0.24 (0.13–0.40)** | **1.00 (0.97–1.00)** | **0.85 (0.41–0.90)** |
| **≥11** | **0.91 (0.59–1.00)** | **0.89 (0.84–0.93)** | **0.29 (0.15–0.46)** | **1.00 (0.97–1.00)** | **0.80 (0.43–0.93)** |
| ≥12 | 0.73 (0.39–0.94) | 0.92 (0.87–0.95) | 0.30 (0.14–0.50) | 0.99 (0.96–1.00) | 0.64 (0.26–0.89) |

*CI* confidence interval, *PPV* positive predictive value, *NPV* negative predictive value; optimal cut-offs are printed in bold

For different cut-off points of the ChilD-S, the following validity measures are summarized in Table 3: (1) Sensitivity, which is the proportion of children with 'any depressive disorder' correctly identified by the ChilD-S, and (2) specificity, which is the proportion of children without 'any depressive disorder' correctly classified by the ChilD-S. In addition, two predictive values are included in this table: (3) The positive predictive value (PPV) means the proportion of children screened positive by the ChilD-S and actually fulfilling the criteria of 'any depressive disorder'. In turn, (4) the negative predictive value (NPV) quantifies the proportion of children screened negative by the ChilD-S but not actually fulfilling the criteria of 'any depressive disorder'.

Ideal cut-off points can be chosen according to the highest Youden-Index. In the clinical context, predictive values represent key criteria for the choice of cut-off points. Eventually, the selection of the best cut-off point depends on the user's aim [54]. For the ChilD-S two optimum cut-off points could be identified:

First, for the cut-off ≥10, sensitivity was 1.00 and specificity was 0.85, yielding a PPV of 0.24 and a NPV of 1.00. In this sample, the ChilD-S correctly identified all eleven affected children as depressed and screened an additional 34 as false positives. Second, for the cut-off ≥11, sensitivity was 0.91 and specificity was 0.89, yielding a PPV of 0.29 and a NPV of 1.00. There were ten children correctly identified by the ChilD-S. Additionally, 25 children were screened as false positives. To sum up, sensitivity was slightly higher for the cut-off ≥10, but specificity and PPV were slightly higher for the cut-off ≥11.

## Discussion

The aim of the study was to develop and validate the Children's Depression Screener (ChilD-S). The eight-item ChilD-S had good psychometric properties, including a satisfactory discriminatory power of the items as well as a high reliability of the scale. The screening instrument showed excellent diagnostic accuracy according to the gold standard diagnosis 'any depressive disorder' and there was no loss in validity compared to the revised 22-item pool. Hence, the ChilD-S can distinguish between depressed and non-depressed children even when using only a few items. These results are very satisfactory and particularly remarkable in light of the fact our sample consisted of children. Other validation studies achieved AUC-values of 0.72 for the Children's Depression Inventory (CDI) [29] or 0.86 for the Mood and Feelings Questionnaire-Child Version (MFQ-C) [38],

which are not as high as with the ChilD-S (AUC = 0.97). There is one study known to date which obtained an excellent diagnostic accuracy with an AUC value of 0.95 [55]. However, another self-report questionnaire was used for validation [55], the depression and anxiety problem scale of the Youth Self Report (YSR) [56] instead of a high quality gold standard in the form of a diagnostic interview.

For clinical application, cut-off scores with corresponding sensitivity and specificity are of special interest [54]. For the screener ChilD-S, there are two optimal cut-off points with an excellent sensitivity ($\geq$0.91) and at the same time a high specificity ($\geq$0.85). Both cut-off points yielded higher sensitivity values than those found in similar studies [29, 33, 36, 38, 39], in which sensitivity ranged between 0.60 and 0.81. At the same time the ChilD-S achieved an at least comparable specificity to the other questionnaires, since these studies found values of specificity between 0.70 and 0.86.

In order to choose between one of the two optimal cut-off points of the ChilD-S, predictive power should be considered. With the cut-off $\geq$10 no depressive children are missed, but at the same time 34 children are mislabelled as depressed. This increases the strain on healthcare resources as a further diagnostic step is necessary to rule out the incorrect screening suspicion. Additionally, mislabelling children unnecessarily alarms their parents. Therefore, in our opinion the cut-off $\geq$11 is preferable. With this cut-off point, the positive predictive value is better (PPV = 0.29) and fewer children are mislabelled. In general, the PPVs are rather low for most screening instruments (e.g. PPV of MFQ-C = 0.21 [38]; PPV of CDI = 0.38 [29]; PPV of CDI = 0.38 [36]). Therefore, screening can only be considered as a first step in a diagnostic process. A more detailed psychiatric evaluation is always necessary to confirm or to rule out a depression diagnosis [20].

In summary, the present study makes an important contribution to this relevant topic, since little research has been performed on the use of brief and valid screening instruments for identifying depressive disorders in children. Besides the ChilD-S, there is only one other screener, the 10-item CDI:S [31], but its criterion validity has not been investigated to date. One important advantage of the ChilD-S compared to the CDI:S is the simple response format; while for the ChilD-S children have to read and judge eight items on a four-point Likert scale, for the CDI:S the children have to read and judge 30 statements (ten items with three statements each). To establish a self-report screening instrument for the younger age-group of children, a simple response format is beneficial, especially for children with reading disabilities or a migration background.

The strength of the study lies in the thorough development and validation process of the ChilD-S. First of all, the children's feedback concerning the comprehensibility of the items was considered in the development process of the item pool. Furthermore, the item pool was exclusively developed and validated for the target group of prepubertal in- and outpatients in pediatric care settings. The use of a comprehensive item pool enabled the selection of the most promising items for the ChilD-S. Moreover, for validation we used a high quality gold standard in the form of a diagnostic interview, which comprises the diagnostic criteria according to DSM-IV-TR. The interviews were conducted by two specially trained psychologists. However, for diagnostic decisions it was not possible to rule out whether depressive symptoms were caused by somatic conditions (e.g. thyroid hypofunction) or side effects of medication (e.g. corticoid treatment). To make a final psychiatric decision in pediatric care it is important to take into account the medical status. And even if the positive score is not confirmed by the final psychiatric diagnosis, the screening result allows the pediatrician to detect depressive symptoms. This in turn is

important, as it enables the pediatrician to treat the underlying medical condition (e.g. thyroid hormone substitution) or to adjust medication.

Besides the strengths, there are some limitations. The age range of 9–12 years of our sample appears to be narrow, but was intentionally chosen as the target population of the ChilD-S. The age range starts at the age of 9 years, since 9-years-old children have sufficient reading skills to complete the questionnaire autonomously. The age range ends at the age of twelve as depression symptoms are likely to change during puberty. With respect to the representativeness of the sample, we were not able to show whether participants and drop-outs differed in terms of the prevalence of depressive disorders. We also failed to address the full range of pediatric patients. Nevertheless, because a broad range of medically ill children was included (only the diabetological, oncological and nephrological units could not take part) we believe that the results are applicable to the majority of pediatric patients. Since the administration of the screening instrument ChilD-S was embedded in the revised 22-item pool, we cannot exclude the possibility that the results would differ slightly if the ChilD-S would have been used separately. In the present sample for the ChilD-S the described eight-item combination was optimal. Of course, ideal item combinations could differ in other samples. Thus, in further studies not only should our eight-item combination be examined, but a wider range of items from our revised item pool should be investigated. Furthermore, we could not determine, if cut-off points differ for in- and outpatients, since there were too few cases of children with depression for additional sub analyses. Because the depression severity may differ for in- and outpatients, the optimal cut-off points for the subgroups are likely to differ. Consequently, sub analyses for optimal cut-off points for in- and outpatients should be addressed in future research.

## Summary

Since childhood depression can harmfully affect children's development, early identification is necessary. Pediatricians could play a crucial role for recognition, but often report a high degree of discomfort with the responsibility of diagnosing children's psychiatric disorders. Brief and valid screening instruments, which can be a helpful support for pediatricians, are lacking.

Hence, the aim of the present study was to develop and evaluate a new depression screening instrument particularly designed for prepubertal in- and outpatients in pediatric care settings, the Children's Depression Screener (ChilD-S). Therefore, an explorative item pool was composed, evaluated and revised. From the revised pool, items with a high depression specific diagnostic accuracy were selected for the ChilD-S. In two pediatric samples, children aged 9–12 ($N_I = 200$; $N_{II} = 246$) completed the explorative item pool (subsample I) and the revised item pool (subsample II). Diagnoses of major depression or dysthymia based on a diagnostic interview served as gold standard for validation. The prevalence rate for 'any depressive disorder' was 6.0% ($N_I$) and 4.5% ($N_{II}$), respectively. Each of the 22 items from the revised pool was evaluated in order to create the valid and brief instrument ChilD-S. Areas under the curve (AUCs) of the revised item pool and the ChilD-S were calculated and compared. Reliability was high for both the revised item pool ($\alpha = 0.88$) and the ChilD-S ($\alpha = 0.81$). For screening of depressive disorders in children, the eight-item ChilD-S (AUC = 0.97) performed just as well as the revised 22-item pool (AUC = 0.94) ($z = 1.382$; $p = 0.167$). For the ChilD-S the optimal cut-off point of $\geq 11$ yielded a sensitivity of 0.91 and a specificity of 0.89.

The present findings for the screening of depressive disorders in children are promising. If these results can be replicated in a large sample, the ChilD-S can be recommended as a valid and brief screening instrument for pediatric care. Hence, the ChilD-S would be the first child specific depression-screening instrument that fulfils the criteria of brevity as well as high validity.

**Acknowledgments** This project was funded by the Bavarian State Ministry of the Environment and Public Health within the promotional emphasis of Gesund.Leben.Bayern. We thank all participating families, as well as the Department of Pediatrics of the Hospital "Harlaching", the Department of Pediatric Surgery of the University Hospital "Dr. von Haunersches Kinderspital", the Departments of Pediatrics and Pediatric Surgery of the Hospital "Dritter Orden" and the Departments of Pediatric and Pediatric Surgery of the University Hospital "Schwabing". We are very grateful to Anne Hoyler for her support in the organisation of the study and to Anna Maria O'Leary as well as to Calum Anderson for linguistic editing of the manuscript.

## References

1. Costello EJ, Erkanli A, Angold A (2006) Is there an epidemic of child or adolescent depression? J Child Psychol Psyc 47:1263–1271
2. Reynolds CR, Kamphaus RW (2003) Behavior assessment system for children. Pearson Assessment, Bloomington
3. Bettge S, Wille N, Barkmann C, Schulte-Markwort M, Ravens-Sieberer U (2008) Depressive symptoms of children and adolescents in a German representative sample: results of the BELLA study. Eur Child Adolesc Psychiatry 17:71–81
4. Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J, Dahl RE et al (1996) Childhood and adolescent depression: a review of the past 10 years. Part I. J Am Acad Child Adolesc Psychiatry 35:1427–1439
5. Fergusson DM, Woodward LJ (2002) Mental health, educational, and social role outcomes of adolescents with depression. Arch Gen Psychiat 59:225–231
6. Segrin C (2000) Social skills deficits associated with depression. Clin Psychol Rev 20:379–403
7. Brent DA, Birmaher B (2009) Paediatric mood disorders. In: Gelder MG, Andreasen NC, López-Ibor JJ Jr, Geddes JR (eds) New oxford textbook of psychiatry. Oxford University Press, Oxford, pp 1669–1680
8. Harrington R, Fudge H, Rutter M, Pickles A, Hill J (1990) Adult outcomes of childhood and adolescent depression. I. Psychiatric status. Arch Gen Psychiatry 47:465–473
9. Richardson LP, Russo JE, Lozano P, McCauley E, Katon W (2010) Factors associated with detection and receipt of treatment for youth with depression and anxiety disorders. Acad Pediatr 10:36–40
10. Cheung AH, Dewa CS, Levitt AJ, Zuckerbrot RA (2008) Pediatric depressive disorders: management priorities in primary care. Curr Opin Pediatr 20:551–559
11. Coyle JT, Pine DS, Charney DS, Lewis L, Nemeroff CB, Carlson GA et al (2003) Depression and bipolar support alliance consensus statement on the unmet needs in diagnosis and treatment of mood disorders in children and adolescents. J Am Acad Child Adolesc Psychiatry 42:1494–1503
12. Fremont WP, Nastasi R, Newman N, Roizen NJ (2008) Comfort level of pediatricians and family medicine physicians diagnosing and treating child and adolescent psychiatric disorders. Int J Psychiat Med 38:153–168
13. Lack CW, Green AL (2009) Mood disorders in children and adolescents. J Pediatr Nurs 24:13–25
14. Pinquart M, Shen Y (2011) Depressive symptoms in children and adolescents with chronic physical illness: an updated meta-analysis. J Pediatr Psychol 36:375–384
15. Essen L, Enskar K, Kreuger A, Larsson B, Sjoden PO (2000) Self-esteem, depression and anxiety among Swedish children and adolescents on and off cancer treatment. Acta Paediatr 89:341–348
16. Wamboldt MZ, Weintraub P, Krafchick D, Wamboldt FS (1996) Psychiatric family history in adolescents with severe asthma. J Am Acad Child Adolesc Psychiatry 35:1042–1049
17. Kashani JH, Venzke R, Millar EA (1981) Depression in children admitted to hospital for orthopaedic procedures. Brit J Psychiat 138:21–25
18. Thombs BD, Bresnick MG, Magyar-Russell G (2006) Depression in survivors of burn injury: a systematic review. Gen Hosp Psychiat 28:494–502
19. Garber J, Walker LS, Zeman J (1991) Somatization symptoms in a community sample of children and adolescents: further validation of the children's somatization inventory. Psychol Assess 3:588–595

20. Sharp LK, Lipsky MS (2002) Screening for depression across the lifespan: a review of measures for use in primary care settings. Am Fam Physician 66:1001–1008
21. Ani C, Garralda E (2005) Developing primary mental healthcare for children and adolescents. Curr Opin Psychiat 18:440–444
22. Briggs-Gowan MJ, Horwitz SM, Schwab-Stone ME, Leventhal JM, Leaf PJ (2000) Mental health in pediatric settings: distribution of disorders and factors related to service use. J Am Acad Child Adolesc Psychiatry 39:841–849
23. Schubiner H, Tzelepis A, Wright K, Podany E (1994) The clinical utility of the safe times questionnaire. J Adolesc Health 15:374–382
24. Cuijpers P, van Straten A, Smits N, Smit F (2006) Screening and early psychological intervention for depression in schools: systematic review and meta-analysis. Eur Child Adolesc Psychiatry 15:300–307
25. Katon W, Russo J, Richardson L, McCauley E, Lozano P (2008) Anxiety and depression screening for youth in a primary care population. Ambul Pediatr 8:182–188
26. Davanzo P, Kerwin L, Nikore V, Esparza C, Forness S, Murelle L (2004) Spanish translation and reliability testing of the child depression inventory. Child Psychiatry Hum Dev 35:75–92
27. Poli P, Sbrana B, Marcheschi M, Masi G (2003) Self-reported depressive symptoms in a school sample of Italian children and adolescents. Child Psychiatry Hum Dev 33:209–226
28. Henkel V, Mergl R, Coyne JC, Kohnen R, Moller HJ, Hegerl U (2004) Screening for depression in primary care: will one or two items suffice? Eur Arch Psy Clin N 254:215–223
29. Sorensen MJ, Frydenberg M, Thastum M, Thomsen PH (2005) The Children's Depression Inventory and classification of major depressive disorder: validity and reliability of the Danish version. Eur Child Adolesc Psychiatry 14:328–334
30. Kovacs M (1992) The Children's Depression Inventory (CDI). Manual. Multi-Health Systems, North Tonawanda
31. Kovacs M (2003) The Children's Depression Inventory (CDI). Technical manual update. Multi-Health Systems, North Tonawanda
32. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P et al (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. J Am Acad Child Adolesc Psychiatry 36:980–988
33. Craighead WE, Curry JF, Ilardi SS (1995) Relationship of Children's Depression Inventory factors to major depression among adolescents. Psychol Assessment 7:171–176
34. Gammon GD, John K, Rothblum ED, Mullen K, Tischler GL, Weissman M (1983) Use of a structured diagnostic interview to identify bipolar disorder in adolescent inpatients. Am J Psychiatry 140:543–547
35. Puigh-Antich J, Orvaschel H, Tabrizi M, Chambers WJ (1983) Schedule for affective disorders and schizophrenia for school-age children—epidemiological version. State Psychiatric Institute, Department of Child and Adolescent Psychiatry, New York (Unpublished manuscript)
36. Shemesh E, Yehuda R, Rockmore L, Shneider BL, Emre S, Bartell AS et al (2005) Assessment of depression in medically ill children presenting to pediatric specialty clinics. J Am Acad Child Adolesc Psychiatry 44:1249–1257
37. Angold A, Weissman MM, John K, Merikangas KR, Prusoff BA, Wickramaratne P et al (1987) Parent and child reports of depressive symptoms in children at low and high risk of depression. J Child Psychol Psyc 28:901–915
38. Daviss WB, Birmaher B, Melhem NA, Axelson DA, Michaels SM, Brent DA (2006) Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. J Child Psychol Psyc 47:927–934
39. Angold A, Costello EJ, Messer SC, Pickles A (1995) Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. Int J Method Psych 5:237–249
40. Schwab-Stone ME, Fisher PW, Piacentini J, Shaffer D, Briggs M (1993) The diagnostic interview schedule for children—revised version (DISC-R): II. Test retest reliability. J Am Acad Child Adolesc Psychiatry 32:651–657
41. American Psychiatric Association (APA) (2000) Diagnostic and statistical manual of mental disorders, Text revision (DSM-IV-TR), 4th edn. APA, Washington
42. Schneider S, Unnewehr S, Margraf J (2009) Kinder-DIPS—Diagnostisches Interview bei psychischen Störungen im Kindes- und Jugendalter. (Diagnostic interview for mental disorders in children and adolescents). Springer, Heidelberg
43. Dilling H, Mombour W, Schmidt MH (2005) Internationale Klassifikation psychischer Störungen ICD-10, Kapitel V(F). Klinisch-diagnostische Leitlinien. (The ICD-10 classification of mental and behavioural disorders, chapter V(F)). Huber, Göttingen
44. Adornetto C, In-Albon T, Schneider S (2008) Diagnostik im Kindes- und Jugendalter anhand strukturierter Interviews: Anwendung und Durchführung des Kinder-DIPS. (Diagnostics in childhood and

adolescence using structured interviews: application and execution of the Kinder-DIPS.). Klin Diagnostik u Evaluation 1:363–377
45. Stiensmeier-Pelster J, Schürmann M, Duda K (2000) Depressionsinventar für Kinder und Jugendliche (DIKJ). (Depression inventory for children and adolescents.). Hogrefe, Göttingen
46. Weissman MM, Orvaschel H, Padian N (1980) Children's symptom and social functioning self-report scales: comparison of mothers' and children's reports. J Nerv Ment Dis 168:736–740
47. Hautzinger H, Bailer M (1993) Allgemeine Depressionsskala (ADS). (Universal depression scale.). Beltz, Weinheim
48. Reynolds WM (2001) Reynolds adolescent adjustment screening inventory (RAASI). Psychological Assessment Resources, Odessa
49. Hampel P, Petermann F (2005) SPS-J—Screening psychischer Störungen im Jugendalter. (Screening for mental disorders in adolescence.). Huber, Bern
50. Achenbach TM, Rescorla C (2001) Manual for the ASEBA school-age forms and profiles. University of Vermont, Research Center for Children, Youth, and Families, Burlington
51. Döpfner M, Berner W, Lehmkuhl G (1994) Handbuch: Fragebogen für Jugendliche. Forschungsergebnisse zur deutschen Fassung der youth self-report form (YSR) der Child Behavior Checklist. (Manual of the questionnaire for adolescents. Research results of the German version of the youth self-report (YSR) of the child behavior checklist.) Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD), Köln
52. Ferdinand RF (2008) Validity of the CBCL/YSR DSM-IV scales anxiety problems and affective problems. J Anxiety Disord 22:126–134
53. DeLong ER, DeLong DM, Clarke-Pearson DL (1988) Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44:837–845
54. Kresanov K, Tuominen J, Piha J, Almqvist F (1998) Validity of child psychiatric screening methods. Eur Child Adolesc Psychiatry 7:85–95
55. Roelofs J, Braet C, Rood L, Timbremont B, van Vlierberghe L, Goossens L et al (2010) Norms and screening utility of the Dutch version of the Children's Depression Inventory in clinical and nonclinical youths. Psychol Assessment 22:866–877
56. Verhulst FC, van der Ende J, Koot HM (1996) Manual for the youth self report (YSR). Afdeling Kinderen Jeugpsychiatrie. Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam, Rotterdam