Screening for Persistent Psychopathology in 4-year Old Children

Trude Hamre Sveen, PsyD. Department of Psychology, Norwegian University of Science and Technology; NTNU Social Science, Trondheim, Norway

Turid Suzanne Berg-Nielsen, PhD. Regional Centre for Child and Adolescent Mental Health, Eastern and Southern Norway (RBUP), Oslo

Stian Lydersen, PhD. Regional Centre for Child and Youth Mental Health and Child Welfare, the Norwegian University of Science and Technology

Lars Wichstrøm, PhD. Department of Psychology, Norwegian University of Science and Technology; NTNU Social Science, Trondheim, Norway; Child and Adolescent Psychiatric Clinic, St. Olavs Hospital, Trondheim, Norway

Address correspondence to: Trude Hamre Sveen, Department of Psychology, NTNU, 7491 Trondheim, Norway, trude.hamre.sveen@svt.ntnu.no, +4747261822 (phone), +4773591920 (fax)

Short title: Screening for Psychopathology

Financial Disclosure: Cand. Psychol. Hamre Sveen and Drs. Berg-Nielsen, Lydersen and Wichstrøm report no financial relationships with commercial interests.

Funding Source: This research was funded by grants 228685/H10, 185760/V50 from the Research Council of Norway and grant 4396 from the Liaison Committee between the Central Norway Regional Health Authority and NTNU.

Conflict of Interest: No conflict of interest is declared.

Abbreviations:
SDQ: Strengths and Difficulties Questionnaire
PAPA: Preschool Age Psychiatric Assessment interview
AUC: Area under the curve
PPV: Positive Predictive Value
NPV: Negative Predictive Value
ROC: Receiver Operating Characteristic
MAR: Missing at random
MCAR: Missing completely at random

What's Known on This Subject

Preschool-onset psychiatric disorders may continue into school-age if undetected. Whether primary care screening can prospectively identify psychopathology that will persist is unknown.

What This Study Adds

Preschool screening identifies psychopathology that will persist with high sensitivity. However, the false positive rate indicates that brief checklists may not reliably distinguish persistent cases from children presenting symptoms at the time of screening, particularly at low rates of disorder.

Contributors' Statement:

Cand. Psychol. Hamre Sveen participated in the acquisition of data, analysis and interpretation, drafting, and revising of the manuscript, and approved the final manuscript as submitted.
Dr. Lydersen (Medical statistics) served as statistical expert. He participated in analysis and interpretation of data and critically revised the manuscript for important intellectual content, and approved the final manuscript as submitted.
Dr. Berg-Nielsen participated in the conception and design, acquisition of data, drafting, and revising of the manuscript and provided supervision, and approved the final manuscript as submitted.
Dr. Wichstrøm participated in the conception and design, acquisition of data, analysis and interpretation, drafting, and revising of the manuscript, obtained funding and provided supervision, and approved the final manuscript as submitted.

ABSTRACT

Objective: To inform primary care screening and preventive intervention efforts, the authors examined the screening efficiency of the Strengths and Difficulties Questionnaire (SDQP4-16) for persistent disorders relative to transient disorders and its capacity to distinguish between the two.

Methods: Persistence and transience in preschool-onset psychiatric disorders was identified using data from a large population-based cohort study of Norwegian children initially assessed at age 4 and followed up at age 6 (n=1038). DSM-IV diagnoses at both time points were assigned using the Preschool Age Psychiatric Assessment Interview, against which the SDQP4-16 was compared through receiver operating characteristics analysis.

Results: The screening efficiency for persistent disorders exceeded that for transient disorders with a specificity of 86.1%, a sensitivity of 79.3%, and an area under the curve (AUC) value of 0.85. The SDQP4-16 was able to discriminate persistent disorders from transient disorders at an AUC value of 0.71. At the selected cut-point of 10, the negative predictive value was 99.6%, whereas the positive predictive value was 9.5%, partly due to the low prevalence (1.8%) of persistent disorders.

Conclusion: The SDQP4-16 is a sensitive tool for detecting persistent psychiatric disorders in young children. However, a large proportion of screen-positives are non-persistent cases, as indicated by the high false positive rate. Thus, the clinical utility of the SDQ in primary care screening for persistent disorders is uncertain, particularly in samples in which the rate of psychiatric disorders is low.

INTRODUCTION

Psychiatric disorders are prevalent among preschoolers (7-26%),[1-6] yet few are identified and referred.[7-9] Community screening may increase the rate of children being reliably identified and treated.[10] A valid screen could assist clinicians faced with overlapping clinically-concerning and normative behavior[11,12] and inform the decision of further assessment and referral. However, studies indicate that early-onset disorders may follow different developmental pathways relevant to screening. Whereas some preschool diagnoses continue into school-age,[13] approximately half of preschool diagnoses are no longer present at follow-up,[13,14] and half of those with a diagnosis at follow-up did not meet criteria at the initial assessment.[13] Thus, screening may identify preschoolers whose trajectories are diverse; those whose disorder would continue into school-age if undetected; those whose disorder would remit; those whose disorder is emerging at a later time. Previous examinations of preschool screening have not taken into account follow-up assessments.[15-18] Consequently, the screening efficiency of existing screens with respect to persistence as opposed to transience of disorders is unknown. This knowledge is potentially useful to guide prevention efforts.

Early childhood psychiatric disorders are highly comorbid,[13,19] and research indicates that a general psychopathology construct taps much of the variance in preschool disorders.[20,21] Moreover, a positive screen at the primary care level could indicate referral for further evaluation, diagnosing and possible treatment, regardless of the symptom type. Thus, when screening young children at this level it may be just as appropriate to target the presence of any psychopathology as specific diagnoses.

Studies have shown the parent-completed Strengths and Difficulties Questionnaire (SDQ)[22] to adequately screen for concurrent psychiatric problems in preschool community populations.[15,23,24] Moreover, compared with pediatric primary care providers, the SDQ identified substantially more children with possible psychopathology.[25] However, whether brief and user-friendly screens such as the SDQ are able to capture persistent disorders is unknown. This study examined the parent-completed SDQ with respect to 1) the overall screening efficiency for persistent disorders relative to transient disorders and the capacity to distinguish the two; 2) the optimal cut-point for persistent cases. Moreover, we extend the generalizability of our findings to samples with higher rates of persistent disorders by 3) determining the screening efficiency of the SDQ for the most common range of stability rates.

METHOD

Recruitment and participants

The sampling frame was the Trondheim Early Secure Study (TESS), comprising two birth cohorts (2003-2004) of children in the city of Trondheim, Norway, who were invited to the community health check-up for 4-year olds. The TESS has been described in detail elsewhere,[6] including screening with the SDQP4-16. The study was approved by the Regional Committee for Medical and Health Research Ethics. After completely describing the study to the eligible subjects, written informed consent from 2475 (82.1%) parents was obtained. To reduce costs a subsample of 1250, oversampled according to higher SDQ scores to increase statistical power, was invited to participate in a structured diagnostic interview concerning the child's mental health[2] completed at age 4 and re-

administered 2 years later, at age 6. Interview information was obtained for 1038 children, of which 753 (72.5%) completed the age 4 and the age 6 interviews. Descriptive information on participants with completed interview at both time points is provided in Table 1.

/ Table 1 near here/

Measures

Screening scale. The parent version of the SDQ (SDQP4-16) was completed at the age 4 assessment. Of the five 5-item subscales (emotional problems, conduct problems, hyperactivity, peer problems and pro-social behavior), the first four are summed to create a "total difficulties score" (SDQtds), ranging from 0 to 40. The SDQ has documented strong psychometric properties for preschool and school-aged children.[26,27] The Norwegian version has been validated in several large studies.[28,29] In our sample, Cronbach's alpha for the total difficulties score was 0.74.

Diagnostic assessment. Psychiatric diagnoses at both time points were assigned using the Preschool Age Psychiatric Assessment (PAPA), a semi-structured psychiatric interview with parents.[2] Symptoms occurring during the 3 months preceding the interview are rated according to a structured protocol involving both required and optional follow-up questions. Diagnoses were generated by computerized algorithms implementing the Diagnostic and Statistical Manual of Mental Disorders, Fourth edition (DSM-IV)[30]

Interviewers (n=7) had at least a bachelor's degree in relevant fields and extensive prior experience in working with children and families. They received training by the group who developed the measure. Interviewers were blind to the SDQ results at both time points.

To evaluate inter-rater reliability, 9% of the interview audio recordings were re-coded by blinded raters. Pairs of raters obtained the following inter-rater reliabilities;[31] ADHD: *k*=.96; ODD: *k*=.89; CD: *k*=.78; any anxiety disorder: *k*=.89; any depressive disorder: *k*=.86; any sleep disorder: *k*=.87; encopresis: *k*=.92; and any disorder: *k*=.87.

All diagnoses were analyzed as an 'any psychiatric disorder' category for both the age 4 and age 6 assessment. The possible combinations of outcomes at age 4 and age 6 generated the following groups: (1) a 'concurrent disorder' group, consisting of children with diagnosis present at age 4 but absent at age 6; (2) a 'prospective disorder' group, consisting of children with diagnosis absent at age 4 but present at age 6; and (3) a 'persistent disorder' group, consisting of children with diagnosis present at both time points. Group 1 and 2 are also referred to as transient disorders.

Missing data and attrition

Table 2 outlines the distribution of cases (children with PAPA diagnoses), non-cases (children without PAPA diagnoses) and missing interview information across the two time points. There was missing diagnostic data for 4.0% of participants at the age 4 assessment and 23.5% of participants at the age 6 assessment. Parent-completed SDQtds was available for all participants. Participants with complete data differed from participants with missing interview information by a lower SDQ impact score (parent

rated: odds ratio=0.59, 95% CI= 0.38-0.93, teacher rated: odds ratio=0.58, 95% CI= 0.38-0.88) and was less frequently rated by the health nurse as being in "need of help for reported problems" (odds ratio=0.54, 95% CI= 0.41-0.73). Investigating the attrition rate from the age 4 to the age 6 assessment among children with a diagnosis relative to children without a diagnosis gives information about what type of missingness mechanism is at work. The higher, albeit non-significant, attrition rate among diagnosed children (30.2%) relative to non-diagnosed children (23.9%) indicates that data are not missing completely at random (MCAR), but missingness is possibly related to observed data, i.e missing at random (MAR)."

/ Table 2 near here/

Statistical analyses

Complete case (CC) analysis uses data only from cases with complete data, and is unbiased only if data are missing completely at random (MCAR). Thus, participants having completed at least one interview were included in the analyses (n=1038), and missing data were handled by multiple imputation (MI) using chained equations. MI analysis also uses information from cases with partially missing data, yielding higher statistical power than CC analysis. Further, MI analysis is unbiased under the less restrictive MAR assumption, and generally less biased than CC if data are MNAR (missing not at random).[32] We created m=100 imputed data sets,[33] with estimates, confidence intervals and p-values computed using Rubin's rules, and the fraction of

missing information computed according to Buuren.[33] For more information on multiple imputation, see supplement.

Mplus 7.2[34] was used to examine possible differences in rates of diagnoses between the age 4 and 6 assessments by comparing a solution where the means at both time points were fixed to be similar with a model in which they were freely estimated. In this analysis, missing data were addressed via a full information maximum likelihood procedure in which the variables used for imputation were entered as auxiliary variables.

The overall screening efficiency of the SDQP4-16 was evaluated using receiver operating characteristic (ROC) curve analysis, which determines the area under the curve (AUC) for the scale against persistent and transient diagnoses. AUC values were interpreted according to Hosmer and Lemeshow[35] : AUC=0.5 (no discrimination), $0.7 \leq AUC < 0.8$ (acceptable discrimination), $0.8 \leq AUC < 0.9$ (excellent discrimination) and $AUC \geq 0.9$ (outstanding discrimination). Probability-weighted versions of the AUC and ROC, together with 95% confidence intervals (CI), were computed using Roger Newson's programs, -somersd- and -senspec-, which are available for download in Stata.[36,37]

The ROC generated sensitivity/specificity pairs were used to select a threshold for the identification of persistent cases. The sensitivity (proportion of screen positives among diagnosed positives) and specificity (proportion of screen negatives among diagnosed negatives) are more stable across populations than are the positive predictive value (PPV) and negative predictive value (NPV). Thus, the sensitivity and specificity data in the present sample allows for estimating PPV and NPV for various prevalences

(see supplement). Screening efficiency was calculated for the present sample, as well as for prevalences of 10% and 15%.

Due to screen-stratification of the sample, we conducted weighted analyses using weights proportional to the inverse of the drawing probability. Analyses were performed in Stata 13.[38]

RESULTS

Table 3 reports the rates of diagnoses at the age 4 and age 6 assessments; corresponding tendencies are observed in the imputed and complete case data, with the imputed estimates being slightly higher because they account for attrition. The prevalence was similar at both time points (P=.58). Meeting criteria for a diagnosis at age 4 was associated with a fivefold greater risk of meeting criteria for a diagnosis at age 6 in the imputed data (odds ratio=5.31, 95% CI= 2.87-9.84) and in the complete case data (odds ratio=5.17, 95% CI=2.69-9.91).

/ Table 3 near here/

None of the children diagnosed at the age 4 assessment had received treatment during the preceding three months. At age 6, 3 out of 17 (17.6%) persistent cases had received treatment.

Overall screening efficiency

The further analyses of the screening efficiency of the SDQP4-16 are based on imputed data (n=1038). According to Hosmer and Lemeshow's definition,[35] the SDQtds had excellent discrimination for persistent cases (see Figure 1). Acceptable discrimination was obtained for concurrent cases, whereas the AUC was below the acceptable level for prospective cases. The capacity to distinguish between persistent and transient cases was acceptable.

/ Figure 1 near here/

The optimal cut-point for persistent cases

The cut-point maximizing the sum of sensitivity and specificity of the SDQP4-16 was found at a score of 10 or greater (Table 4). Increasing the cut-point by one would lead to a considerable decrease in sensitivity, whereas further reducing the cut-point would imply a decrease in specificity without detecting more persistent cases. The scale ruled in (identified true cases as reflected by the sensitivity) 80% of children with persistent diagnoses. However, it was considerably less sensitive to transient cases; at a cut-point of 10, half of the children with concurrent diagnosis and only one-third of the children with prospective diagnosis were detected.

/ Table 4 near here/

Screening efficiency for varying prevalences

Table 4 shows the screening efficiency of the SDQP4-16 with respect to sensitivity, specificity, the PPV and NPV for persistent diagnoses at a cut-point of 10. In populations with a 10% rate of persistent disorders, the SDQP4-16 would obtain a PPV of 39.1%, further increasing to 50.5% at a15% prevalence. The probability of being a true positive when screening positive (PPV), increases with increasing prevalence. Considerable increases in prevalence only produce minor reductions in NPV.

DISCUSSION

The present study examined screening efficiency for persistent psychiatric disorders. The SDQP4-16's discriminative capacity was twofold; being good (AUC=0.85) at discriminating persistent disorders from children not presenting a persistent pattern but modest (AUC=0.71) at discriminating persistent disorders from transient cases. At the selected cut-point most persistent cases are screen-positives (sensitivity=0.79), whereas most non-persistent cases are screen-negatives (specificity=0.86). However, at the low observed frequency of disorder (<2%) false positives constitute a proportionally larger portion of the screen-positives than true positive cases, yielding a low PPV. At higher rates of persistent disorders, screen-positives would include more true positives and proportionally fewer false positives (increased PPV).

In this study, the probability (AUC=0.85) that a randomly selected child with a persistent disorder would have a higher SDQ-score than a randomly selected child

without a persistent disorder outperforms that obtained for transient cases (AUC=0.74 and 0.68 for concurrent and prospective cases, respectively). Moreover, it exceeds that obtained in studies of preschool and school-aged children not considering diagnostic status at follow-up.[15,39] However, the SDQ's capacity to discriminate persistent cases from children diagnosed at one assessment (concurrent and prospective cases) was lower (AUC=0.71). Screens that could extract persistent cases at an early stage could potentially improve our ability to intervene effectively to prevent continuity of these disorders. However, whereas symptom count as offered by check lists such as the SDQ seems suitable for differentiating persistent pathology from non-pathology, it may not be sufficient to distinguish persistent pathological behavior from transient pathological behavior.

At the selected cut-point of 10, the estimated specificity (86%) and the negative predictive value (99.6%) were high, meaning that the SDQ largely ruled out children that did not show a persistent pattern of disorder. The higher sensitivity for persistent cases (79%) relative to transient cases (50% and 33% for concurrent and prospective cases, respectively) indicates that far fewer persistent cases were missed by the SDQ. However, the accompanying false positive rate was high. This latter finding is consistent with prior findings; when screening in community samples, the proportion of true negatives (NPV) is high, but the proportion of true positives (PPV) is substantially lower, and hence highly related to the prevalence.[40-42] Consequently, increasing the cut-point to 11 in the present sample scarcely affects PPV (rises from 9.5 to 10.7) but yields a substantial decline in sensitivity (from 79% to 64%). Thus, to minimize the false negative rate for persistent cases was our primary guidance when selecting the cut-point. In populations with higher

rates of persistent disorders, a substantially higher rate of true positives (increased PPV) would be detected, but a somewhat larger proportion would be false negatives (decreased NPV) (see Table 4).

A positive screen indicates risk of disorder which requires subsequent assessments to reveal potential presence of psychopathology and possible need of intervention. These subsequent assessments are essential to avoid imposing unnecessary and potentially demanding and risky interventions, e.g. medication, on someone who may not need it. In the present study, the estimated false positive rate includes transient disorders. For these concurrent and prospective cases, a positive screen may be an opportunity for intervention to relieve stress and impairment or for preventive measures before problems become more serious.[10] Moreover, as psychopathology is dimensional in nature, a screen-positive non-case may still experience problems that are possibly impairing even if they do not necessarily warrant a clinical diagnosis. However, screening may cause stress and worry among parents whose children are falsely screened positive and lead to labelling of children who would have been better off unlabeled. Moreover, subsequent assessments of screen-positives would consume considerable resources. When false positive rates are high, a substantial share of the resources would not reach those who need it the most. A more fine-grained screen covering a broader range of childhood psychopathology than the SDQ offers, e.g. the Achenbach System of Empirically Based Assessment[43] (99 items) may offer a better initial differentiation between true and false positives. This would, however, run counter to brevity, a key characteristic of a screen that is suitable in primary care. Moreover, in community populations, the proportion of true positives is lower and milder symptomatology predominates relative to clinical populations.[44-47]

Under these circumstances, it is more challenging to extract children suffering from psychopathology that requires intervention.

Stable prevalences from preschool to school age in the present sample coincide with prior findings from the United States,[48] however the rates (approximately 7% at both time points) were comparatively lower and in line with other Scandinavian findings.[28,29,49] Whereas PPV and NPV are affected by prevalence, sensitivity and specificity--and thus AUC estimates--are reasonably stable across prevalences and populations[50] and may generalize to other populations. Indications of comparable reliability and validity of the SDQ across Western countries[39] support the validity of our results concerning the screening efficiency for other populations. Our results support earlier findings of heterogeneity within early-onset psychopathology; a substantial proportion of children meeting criteria for a diagnosis at age 4 did not meet criteria for a diagnosis at age 6 and vice versa. The stability rate of 1.8% in the present sample reflects the low prevalence at baseline; the fewer children who have a diagnosis at baseline, the fewer children remain diagnosed 2 years later. Statistically, a baseline prevalence of 50% would yield a stability of 25% by chance alone, whereas a baseline prevalence of 25% would indicate a stability of approximately 6% by chance. The fact that the stable cases were seldom referred for treatment, and none when they were preschoolers, underscores the importance of detecting them early on.

Some limitations of this study should be noted. First, some participants were lost to follow-up. However, the use of full information maximum likelihood and multiple imputation, should have minimized the likelihood of inaccurate estimates and increased statistical power relative to using a complete case analysis approach. Second, our subjects

were mostly of Norwegian origin; the findings may thus not generalize to more ethnically diverse populations. Third, parent-reported SDQ scores were compared with the PAPA interview, which was also derived from parental information. Although the PAPA interview is clearly interviewer-based, comparative information (e.g., clinician rating, information from teachers) would minimize potential biases associated with a single informant. Fourth, the 3-month primary period may have limited the identification of cases with onset and remission occurring prior to this period or between assessments. Fifth, the confidence interval in the sensitivity estimation was large. Replication in samples with different stability rates and environments is needed to support our findings.

CONCLUSION

The present study suggests that the SDQ may assist clinicians ruling in persistent cases; 80% of preschoolers with a disorder at age 4 that continues to age 6 were detected. However, a large proportion of screen positives are non-persistent cases, and subsequent and more detailed assessments to distinguish those that require swift intervention from those with different or no interventional needs are essential. Primary care could benefit from screening tools that increase targeted and efficient use of resources. The present findings raise questions regarding the usefulness of the SDQ in primary care screening of persistent psychiatric disorders in young children, particularly at low rates of disorder.

REFERENCES

1. Earls F. Application of DSM-III in an epidemiological study of preschool children. *Am J Psychiatry.* 1982;139(2):242-243.
2. Egger HL, Erkanli A, Keeler G, Potts E, Walter BK, Angold A. Test-retest reliability of the Preschool Age Psychiatric Assessment (PAPA). *J Am Acad Child Adolesc Psychiatry.* 2006;45(5):538-549.
3. Keenan K, Shaw DS, Walsh B, Delliquadri E, Giovannelli J. DSM-III-R disorders in preschool children from low-income families. *J Am Acad Child Adolesc Psychiatry.* 1997;36(5):620-627.
4. Lavigne JV, Gibbons RD, Christoffel KK, et al. Prevalence rates and correlates of psychiatric disorders among preschool children. *J Am Acad Child Adolesc Psychiatry.* 1996;35(2):204-214.
5. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *J Clin Child Adolesc Psychol.* 2009;38(3):315-328.
6. Wichstrom L, Berg-Nielsen TS, Angold A, Egger HL, Solheim E, Sveen TH. Prevalence of psychiatric disorders in preschoolers. *J Child Psychol Psychiatry.* 2012;53(6):695-705.
7. Wichstrom L, Belsky J, Jozefiak T, Sourander A, Berg-Nielsen TS. Predicting service use for mental health problems among young children. *Pediatrics.* 2014;133(6):1054-1060.
8. Lavigne JV, Arend R, Rosenbaum D, et al. Mental health service use among young children receiving pediatric primary care. *J Am Acad Child Adolesc Psychiatry.* 1998;37(11):1175-1183.
9. Lavigne JV, Binns HJ, Christoffel KK, et al. Behavioral and emotional problems among preschool children in pediatric primary care: prevalence and pediatricians' recognition. Pediatric Practice Research Group. *Pediatrics.* 1993;91(3):649-655.
10. Weitzman C, Wegner L, Section on D, et al. Promoting optimal development: screening for behavioral and emotional problems. *Pediatrics.* 2015;135(2):384-395.
11. Wakschlag LS, Briggs-Gowan MJ, Choi SW, et al. Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *J Am Acad Child Adolesc Psychiatry.* 2014;53(1):82-96 e83.
12. Wakschlag LS, Estabrook R, Petitclerc A, et al. Clinical Implications of a Dimensional Approach: The Normal:Abnormal Spectrum of Early Irritability. *J Am Acad Child Adolesc Psychiatry.* 2015;54(8):626-634.

13. Bufferd SJ, Dougherty LR, Carlson GA, Rose S, Klein DN. Psychiatric disorders in preschoolers: continuity from ages 3 to 6. *Am J Psychiatry.* 2012;169(11):1157-1164.

14. Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, Gibbons RD. Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *J Am Acad Child Adolesc Psychiatry.* 1998;37(12):1246-1254.

15. Sveen TH, Berg-Nielsen TS, Lydersen S, Wichstrom L. Detecting psychiatric disorders in preschoolers: screening with the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry.* 2013;52(7):728-736.

16. Luby JL, Heffelfinger A, Koenig-McNaught AL, Brown K, Spitznagel E. ThePreschool Feelings Checklist: a brief and sensitive screening measure for depression in young children. *J Am Acad Child Adolesc Psychiatry.* 2004;43(6):708-717.

17. Dehon C, Scheeringa MS. Screening for preschool posttraumatic stress disorder with the Child Behavior Checklist. *J Pediatr Psychol.* 2006;31(4):431-435.

18. Scheeringa MS, Haslett N. The reliability and criterion validity of the Diagnostic Infant and Preschool Assessment: a new diagnostic instrument for young children. *Child Psychiatry Hum Dev.* 2010;41(3):299-312.

19. Wichstrom L, Berg-Nielsen TS, Angold A, Egger HL, Solheim E, Sveen TH. Prevalence of psychiatric disorders in preschoolers. *Journal of child psychology and psychiatry, and allied disciplines.* 2012;53(6):695-705.

20. Olino TM, Dougherty LR, Bufferd SJ, Carlson GA, Klein DN. Testing models of psychopathology in preschool-aged children using a structured interview-based assessment. *J Abnorm Child Psychol.* 2014;42(7):1201-1211.

21. Wichstrom L, Berg-Nielsen TS. Psychiatric disorders in preschoolers: the structure of DSM-IV symptoms and profiles of comorbidity. *Eur Child Adolesc Psychiatry.* 2013.

22. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psychiatry.* 1997;38(5):581-586.

23. Gomez-Beneyto M, Nolasco A, Moncho J, et al. Psychometric behaviour of the strengths and difficulties questionnaire (SDQ) in the Spanish national health survey 2006. *BMC Psychiatry.* 2013;13:95.

24. Theunissen MH, Vogels AG, de Wolff MS, Reijneveld SA. Characteristics of the strengths and difficulties questionnaire in preschool children. *Pediatrics.* 2013;131(2):e446-454.

25. Brown JD, Wissow LS. Screening to identify mental health problems in pediatric primary care: considerations for practice. *Int J Psychiatry Med.* 2010;40(1):1-19.

26. Mieloo C, Raat H, van Oort F, et al. Validity and reliability of the strengths and difficulties questionnaire in 5-6 year olds: differences by gender or by parental education? *PLoS One.* 2012;7(5):e36805.

27. Niclasen J, Teasdale TW, Andersen AM, Skovgaard AM, Elberling H, Obel C. Psychometric properties of the Danish Strength and Difficulties Questionnaire: the SDQ assessed for more than 70,000 raters in four different cohorts. *PLoS One.* 2012;7(2):e32025.

28. Obel C, Heiervang E, Rodriguez A, et al. The Strengths and Difficulties Questionnaire in the Nordic countries. *Eur Child Adolesc Psychiatry.* 2004;13 Suppl 2:II32-39.

29. Heiervang E, Stormark KM, Lundervold AJ, et al. Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. *J Am Acad Child Adolesc Psychiatry.* 2007;46(4):438-447.

30. Association AP. Diagnostic and Statistical Manual of Mental Disorders (fourth edition) ed. Washington, DC: American Psychiatric Association; 1994.

31. Janson H, Olsson U. A measure of agreement for interval or nominal multivariate observations by different sets of judges. *Educ Psychol Meas.* 2004;64(1):62-70.

32. Sterne JA, White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., & Carpenter, J.R. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *Bmj.* 2009;338.

33. Buuren SV. *Flexible imputation of missing data.* Boca Raton, FL: Chapman and Hall/CRC Press; 2012.

34. Muthén LK, Muthén BO. *Mplus user's guide.* 7th ed. Los Angeles, CA: Muthén & Muthén; 1998-2012.

35. Hosmer DW, Lemeshow S. *Applied logistic regression.* 2nd ed. New York: Wiley; 2000.

36. Newson R. senspec: Stata module to compute sensitivity and specificity results saved in generated variables http://ideas.repec.org/c/boc/bocode/s439801.html. Accessed 15.10.12, 2012.

37. Newson R. Confidence intervals for rank statistics: Somers' D and extensions. *Stata J.* 2006;6(3):309-334.

38. StataCorp. *Stata statistical software: release 11.* College Station, TX: StataCorp; 2009.

39. Stone LL, Otten R, Engels RC, Vermulst AA, Janssens JM. Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: a review. *Clin Child Fam Psychol Rev.* 2010;13(3):254-274.

40. Goodman R, Ford T, Simmons H, Gatward R, Meltzer H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry.* 2000;177:534-539.

41. Sveen TH, Berg-Nielsen TS, Lydersen S, Wichstrom L. Detecting psychiatric disorders in preschoolers: screening with the strengths and difficulties questionnaire. *J Am Acad Child Psy.* 2013;52(7):728-736.

42. Biederman J, Monuteaux MC, Kendrick E, Klein KL, Faraone SV. The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. *Arch Dis Child.* 2005;90(10):1010-1015.

43. Achenbach TM RL. *Manual for ASEBA Preschool Forms & Profiles.* Burlington, VT: Research Center for Children,Youth, and Families, University of Vermont.; 2000.

44. Fombonne E. The use of Questionnaires in Child Psychiatric Research: Measuring their performance and Choosing an Optimal Cut-Off. *Journal of Child Psychology & Psychiatry.* 1991;32(4):677-693.

45. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62(6):617-627.

46. Demyttenaere K, Bruffaerts R, Posada-Villa J, et al. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. *JAMA.* 2004;291(21):2581-2590.

47. Ezpeleta L, de la Osa N, Granero R, Domenech JM, Reich W. The Diagnostic Interview of Children and Adolescents for Parents of Preschool and Young Children: psychometric properties in the general population. *Psychiatry Res.* 2011;190(1):137-144.

48. Bufferd SJ, Dougherty LR, Carlson GA, Rose S, Klein DN. Psychiatric disorders in preschoolers: continuity from ages 3 to 6. *Am J Psychiatry.* 2012;169(11):1157-1164.

49. Rescorla L, Achenbach T, Ivanova MY, et al. Behavioral and emotional problems reported by parents of children ages 6 to 16 in 31 societies. *J Emot Behav Disord.* 2007;15(3):130-142.

50. Zhou XH, Obuchowski NA, McClish DK. *Statistical methods in diagnostic medicine.* 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.

TABLES AND FIGURES

TABLE 1. Demographic Characteristics of Children (N=753) and Parents with Complete Information at Age 4 and Age 6 Available

TABLE 2. Cases, Non-cases and Missing Interview[a] Information across Age 4 and Age 6 Assessments

TABLE 3. Rates of Diagnoses

FIGURE 1. ROC curves representing the discriminative capacity of SDQ for persistent, concurrent and prospective disorders and the ability of SDQ to distinguish persistent and transient disorders [a]

TABLE 4. Sensitivity, Specificity, PPV, NPV for Persistent Disorders