Published in final edited form as:
Eur Child Adolesc Psychiatry. 2009 May ; 18(5): 284–291. doi:10.1007/s00787-008-0729-1.

Five-year predictive validity of *DSM-IV* conduct disorder research diagnosis in 4½–5-year-old children

Julia Kim-Cohen, PhD,
Dept. of Psychology, Yale University, PO Box 208205, New Haven (CT) 06520, USA

Louise Arseneault, PhD,
Institute of Psychiatry, King's College London, London, UK

Rhiannon Newcombe, PhD,
Health Sponsorship Council, Wellington, New Zealand

Felicity Adams, MSc,
Institute of Psychiatry, King's College London, London, UK

Heather Bolton, MSc,
Institute of Psychiatry, King's College London, London, UK

Lisa Cant, MSc,
Institute of Psychiatry, King's College London, London, UK

Kira Delgado, MSc,
Institute of Psychiatry, King's College London, London, UK

Jo Freeman, BA,
Institute of Psychiatry, King's College London, London, UK

National Academy of Parenting, Pactitioners, London, UK

Ania Golaszewski, BSc,
Institute of Psychiatry, King's College London, London, UK

Katerina Kelesidi, MSc,
Institute of Psychiatry, King's College London, London, UK

Charlotte Matthews, BSc,
Institute of Psychiatry, King's College London, London, UK

Nicky Mountain, BSc,
Institute of Psychiatry, King's College London, London, UK

Donna Oxley, MA,
Institute of Psychiatry, King's College London, London, UK

Susanne Watson, BSc,
Institute of Psychiatry, King's College London, London, UK

Helen Werts, BA,

Correspondence to: Julia Kim-Cohen, julia.Kim-cohen@yale.edu.

Institute of Psychiatry, King's College London, London, UK

Avshalom Caspi, PhD, and
Institute of Psychiatry, King's College London, London, UK

Duke University, Durham, USA

Terrie E. Moffitt, PhD
Institute of Psychiatry, King's College London, London, UK

Duke University, Durham, USA

Julia Kim-Cohen: julia.Kim-cohen@yale.edu

Abstract

Objective—This longitudinal study of a non-referred, population-based sample tested the 5-year predictive validity of the *DSM-IV* conduct disorder (CD) research diagnosis in children 4½–5 years of age.

Method—In the E-Risk Study, a representative birth cohort of 2,232 children, mothers were interviewed and teachers completed mailed questionnaires to assess children's past 6-month CD symptoms. A follow-up assessment was conducted when children were 10 years old.

Results—CD-diagnosed 5-year-olds were significantly more likely than controls to have behavioural and educational difficulties at age 10. Increased risk for age-10 educational difficulties persisted after controlling for age-5 IQ and ADHD diagnosis. Although the majority of CD-diagnosed 5-year-olds had no CD symptoms at age 10, findings suggest that these "remitted" children continued to experience behavioural and educational problems 5 years later despite their apparent remission from CD.

Conclusions—*DSM-IV* CD symptoms validly identify preschool-aged children who continue to have behavioural and educational problems in middle-childhood.

Keywords

conduct disorder; preschool; predictive validity; epidemiology; nosology

Introduction

In recent years, evidence has been accumulating in support for the validity of the *DSM-IV* conduct disorder (CD) diagnosis in preschool-aged children [4]. For instance, one study demonstrated that the factor structure of disruptive behaviour disorders in a large sample of children aged 2–5 years was differentiated along the same lines as psychopathology in older children and adolescents [30]. Moreover, studies of both clinic-referred [15-17] and non-referred community samples [18] have found good concurrent and convergent validity for the *DSM-IV* CD diagnosis in preschool-aged children. Thus far, however, only one epidemiological study has tested predictive validity. We previously documented the prospective 2-year predictive validity of the *DSM-IV* CD diagnosis in an epidemiological sample of 4½–5-year-old children [18]. Compared with non-diagnosed children, 5-year-olds diagnosed with CD were significantly at greater risk for a CD diagnosis and behavioural and

educational difficulties when reassessed at age 7. Although many 5-year-olds showed apparent remission from CD by age 7, these children continued to experience clinically significant difficulties [18]. Nevertheless, predictive validity over a 2-year period is a fairly short time span and it is possible that over a longer epoch, preschoolers diagnosed with CD may appear no different from non-diagnosed peers. Here we extend our previous findings over a 5-year period to middle childhood when the same children are 10 years old.

Further testing the predictive validity of *DSM-IV* CD diagnosis in preschoolers is important for several reasons. First, predictive validity is an essential function of a diagnostic system. Good predictive validity of a diagnosis would inform clinicians' ability to predict patients' long-term prognosis and identify children most in need of intervention. Second, when CD symptoms emerge in the preschool period, timely intervention is desirable to prevent chronic CD [14, 26, 32, 37] and a diagnosis is typically necessary for families to qualify for affordable treatment. Third, the predictive validity of diagnosing CD in preschoolers is one of several research priorities for informing the forthcoming *DSM-V* [23, 36]. Evidence is needed regarding whether young children's conduct problems can be validly classified as a disorder based on *DSM-IV* nosology during a developmental period when such problem behaviour is not uncommon [11, 35].

In this study, we tested predictive validity in two ways. First, we compared children who did and did not have an age-5 CD diagnosis on age-10 behavioural and educational outcomes as reported by mothers and teachers. Second, we identified a group of children who had age-5 CD but had no age-10 CD symptoms. Support for predictive validity would be evident if these ostensibly "remitted" children have worse outcomes at age 10 compared to non-CD controls. Although we also report on the continuity of a CD diagnosis between ages 5 and 10 years, the key criteria for assessing predictive validity in this study were children's functional outcomes in middle-childhood, which are important regardless of concurrent diagnostic status.

Method

Participants

Participants are members of the Environmental Risk (E-Risk) Longitudinal Twin Study. The E-Risk sampling frame was two consecutive birth cohorts (1994 and 1995) in a birth register of twins born in England and Wales [33]. Of the 15,906 twin pairs born in these 2 years, 71% joined the register. Bias from non-response was corrected as follows.

The E-Risk Study probability sample was drawn using a high-risk stratification sampling procedure. High-risk families were those in which the mother had her first birth when she was 20 years of age or younger. We used this sampling [1] to replace high risk families who were selectively lost to the register via non-response and [2] to ensure sufficient base rates of children growing up in at-risk environments. Age at first childbearing was used as the risk-stratification variable because it was recorded for virtually all families in the register, it is relatively free of measurement error, and early childbearing is a known risk factor for children's problem behaviours [22, 24]. The sampling strategy resulted in a final sample in which one-third of Study mothers (younger only; $N = 314$) constitute a 160% oversample of

mothers who were at high risk based on their young age at first birth (15–20 years). The other two-thirds of Study mothers (N = 802) accurately represent all mothers in the general population (aged 15–48) in England and Wales in 1994–1995 (estimates derived from the General Household Survey [5]). To provide unbiased statistical estimates that can be generalized to the population of British families with children born in the 1990s, the data reported in this article were corrected with weighting to represent the proportion of young mothers in that population [7].

The E-Risk Study sought a sample size of 1,100 families to allow for attrition in future years of the longitudinal study while retaining statistical power. An initial list of families who had same-sex twins was drawn from the register to target for home visits. Of the families from the initial list, 1,116 (93%) participated in home-visit assessments when the twins were age 5 years, forming the base sample for the study: 4% of families refused, and 3% could not be reached after many attempts. Written informed consent was obtained from mothers. With parent's permission, questionnaires were posted to the children's teachers, and teachers returned questionnaires for 94% of cohort children.

Follow-up home interview data were collected for 96% of the 1,116 families at age 10 and teacher questionnaires were obtained for 90% of the participants taking part in the follow-up. The E-Risk Study has received ethical approval from the Maudsley Hospital Ethics Committee.

CD diagnosis

We derived a research diagnosis of children's CD on the basis of mothers' and teachers' reports on 14 of 15 *DSM-IV* symptoms of CD [18], using the *Child Behaviour Checklist* (1) and *Teacher's Report Form* (2). We supplemented these instruments with items from the *Diagnostic Interview Schedule for Children* [9] to ensure that the interview covered all CD criterion symptoms ("forced sexual activity" was excluded as inappropriate for 5-year-olds). Mothers' reports were obtained in a face-to-face, standardized interview in the family home. Interviewers were blind to child diagnostic status. Teachers' reports were obtained via mailed questionnaires.

A child was considered to have a symptom if either the mother or the teacher reported the symptom as being "very true" or "often true" of the child over the past 6 months at 4½–5 years of age. Some CD items that are very serious behaviours (e.g., "has used a weapon that can cause serious harm") and are considered to be clinically significant if done only once were counted if reported as being "very true." Milder CD items which the *DSM-IV* requires "often" (e.g., "often initiates physical fights") were counted if reported as being "often true." Items for which mothers and teachers responded only "somewhat true" were not counted. We counted a symptom as present if reported by either source, following evidence that this approach enhances diagnostic validity [6, 25]. Symptom counts ranged from 0 to 11. Consistent with *DSM-IV* criteria, children with three or more symptoms were assigned a CD diagnosis. At age 5 years, the prevalence of CD, weighted to represent the population, was 6.6% (N = 189; reported Ns are unweighted) [18]. The same methods were used to diagnose CD at the age-10 assessment, and data were available for 184 of the 189 CD-diagnosed children.

Age-5 control variables

IQ—Each child was individually tested using a short form of the *Wechsler Preschool and Primary Scale of Intelligence-Revised* (*WPPSI-R*) [38], comprising Vocabulary and Block Design subtests. IQs were prorated [27]. The children's IQs ranged from 52 to 145 (M = 98, SD = 14).

ADHD diagnosis—ADHD research diagnoses were based on *DSM-IV* criteria. Children received the diagnosis if they had six or more of the hyperactivity–impulsivity symptoms and/or six or more of the inattentiveness symptoms according to either mother or teacher report. To document pervasiveness, the other rater had to indicate two or more symptoms of either inattentiveness or hyperactivity–impulsivity. Therefore, the diagnostic criteria included the presence of symptoms in more than one setting (home and school), as well as onset before age 7 since all children were first assessed at age 5 years. Symptoms were counted as present only if scored "very or often true". The prevalence of ADHD diagnoses was 5.7% [19].

Age-10 outcome measures

Behavioural scales were created separately for mothers' and teachers' reports of ADHD symptoms, aggression, delinquency, emotional problems, and prosocial behaviour using items from the *Child Behaviour Checklist* (1) and the *Teacher's Report Form* (2), supplemented with items from the *DSM-IV* diagnostic criteria for ADHD and the Rutter Child Scales [28]. Symptoms and behaviours were reported for the preceding 6 months and each item was scored as (0) "not true," (1) "somewhat true," and (2) "very often true". Internal consistencies ranged from 0.66 to 0.95.

Treatment for behavioural/emotional problems by a professional (e.g., physician, psychologist, social worker) in the past year was reported by mothers (N = 351, 15.8% weighted).

Special education service use in the past year was reported by mothers (N = 297, 12.6% weighted).

Children's reading ability was individually tested using the Test of Word Reading Efficiency (TOWRE) [31]. Scores were converted to standardized scores, according to the test manual. We report the percentage of children with a reading score below 90.

English and math school performance—In the Teacher's Report Form [2], teachers were asked to rate the child's current English and math performance, using a 5-point scale (0 = far below average, 1 = somewhat below average, 2 = average, 3 = somewhat above average, and 4 = far above average), compared to pupils of the same age. The sample mean was 2.06 (SD = 0.97) for English and 2.13 (SD = 0.97) for math performance. We report the percentage of children who were rated by teachers as "somewhat below average" or "far below average" in English and math skills.

Teacher's effort—Teachers were asked to report on what it was like to work with this child relative to other children in the class, using a 7-point scale (0 = much less; 3 = average;

6 = much more) to rate six items. Items were: How frequently "…must you act to curb disruptive behaviour by this child?," "…must you give this child extra encouragement to get him/her to take part?," "…must you act to keep this child's attention on task?," "…does this child's behaviour make it rewarding to work with him/her?" (reverse-coded), "…does this child's behaviour make it frustrating to work with him/her?" and "…does this child need one-on-one interaction from you?" The sample mean was 11.66 (SD = 7.99). The internal consistency alpha was 0.88.

Statistical analysis

Group differences were evaluated with t-tests (for continuous variables) and odd ratios (for dichotomous variables), and effect sizes (d) were calculated [13]. Tests were two-tailed and based on the sandwich or Huber/White variance estimator [39], a method available in STATA 9.0 [29], which adjusts estimated standard errors to account for the dependence in the data due to analyzing two children per family.

Results

Predictive validity

At age 10 years, the weighted prevalence of CD was 2.6% ($N = 74$) in the total sample, 3.7% ($N = 49$) for boys, and 1.5% ($N = 25$) for girls. Compared with undiagnosed children, the age-5 CD group was at significantly greater risk for a CD diagnosis at age 10 years (OR: 9.0, 95% CI: 4.9, 16.5). Boys were no more likely than girls to maintain a CD diagnosis at both ages.

Compared with non-CD controls, children meeting criteria for a CD diagnosis at age 5 years had significantly higher levels of age-10 mother- and teacher-reported ADHD symptoms, aggression, and delinquency, lower levels of prosocial behaviour, and were more likely to have received treatment for behavioural/emotional problems (Table 1). Compared with controls, the CD group was approximately 2–3 times more likely to have age-10 educational problems and to require more effort from teachers. After controlling for age-5 IQ and ADHD diagnosis, the CD group continued to be at significantly greater need for special education services [adjusted odds ratio (AOR): 1.8; 95% CI: 1.1, 2.8], to be at increased risk for poor English performance (AOR: 1.6; 95% CI: 1.0, 2.4), and to require more effort from teachers [$t(1,928) = 4.30$, $P < 0.001$]. After controlling for age-5 IQ and ADHD, the CD group was at increased risk for poor reading skills and below average math performance, but these comparisons fell short of significance ($P < 0.09$). Boys with age-5 CD were no more likely than girls with age-5 CD to have any of the age-10 behavioural or educational outcomes, with one exception. Age-5 CD boys were significantly more likely than age-5 CD girls to have higher teacher-rated ADHD symptoms ($P = 0.04$).

Predictive validity in "remitted" children

Among the 184 5-year-olds who met diagnostic criteria for CD and had available data at follow-up, 115 (63% weighted) had no CD symptoms at age 10. We compared age-10 outcome measures for the group of children who had an age-5 CD diagnosis but no CD symptoms at age 10 ("remitted" group) against children who did not have an age-5 CD

diagnosis (comparison group) (Table 2). Findings indicated that "remitted" children continued to experience elevated behavioural problems, had fewer prosocial behaviours, and received treatment for behavioural/emotional problems at significantly higher rates than children with no age-5 CD. Compared with non-conduct-disordered children, "remitted" children were at significantly greater need of special education services, had lower English and math performance, and required more effort from teachers.

Discussion

The predictive validity of a diagnostic test is demonstrated if the diagnostic status is associated with criterion measures of functioning longitudinally over time [3]. This study found that applying *DSM-IV* CD symptoms to 4 ½–5-year-old children in the community is predictive of continuing behavioural and academic concerns 5 years later. One-quarter of 5-year-olds with a CD research diagnosis were in special education services at age 10 and teachers reported children with preschool CD were more burdensome and required more teaching effort in the classroom. Effect sizes comparing diagnosed versus non-diagnosed children on age 10 outcomes were medium to large [8]. Ten-year-olds who appear to be free of CD symptoms nonetheless have elevated behavioural and educational problems if they have a positive history of preschool-aged CD. Thus, it appears that *DSM-IV* CD criteria validly identify a subset of young children who might benefit from early intervention. Evidence for predictive validity is strengthened by the fact that parents and teachers were not informed about children's research diagnosis of CD and treatment referrals were not made. Therefore, the findings herein were not influenced by either "stigma" effects or intervention as a consequence of our research diagnosis.

This study is the first to report on the long-term predictive validity of *DSM-IV* CD in preschool-aged children, but our findings should be interpreted in light of several limitations. First, our sample comprised mostly white British twins and our findings may not generalize to ethnic/racial groups in other countries or to singletons. However, our CD prevalence estimate and sex ratio are comparable to other epidemiological studies of singletons in the U.K. [21] as well as in the U.S. [10]. Second, our "research" diagnosis may differ from typical practice in clinical settings. However, a strength of our study was interviewing mothers face-to-face to assess child symptoms rather than relying on a self-administered questionnaire, and obtaining collateral information from teachers. At the time the E-Risk children were 4 ½–5-years-old, standardized interview [12, 15] and observational methods [34] for making diagnoses in preschool-aged children were not yet available. Third, we used a reporting period of 6 months, which differs from the *DSM-IV* practice of assessing CD symptoms in the past 12 months with at least one criterion present in the past 6 months. At our 4½–5-year assessment, a 6-month period for all symptom criteria would exert a conservative effect on the resulting diagnostic group. However, at our age-10 assessment, the 6-month reporting period might have missed identifying some children who might have met diagnostic criteria over a 12-month period, thus creating a false appearance of recovery. Although this may not be ideal, we needed to keep the reporting period the same across assessments in order to have repeated identical measures for the wider study. Fourth, we could not control for possible confounding diagnoses other than age-5 ADHD

because they were not assessed. Fifth, we diagnosed CD in 4½–5 year old children. Research on predictive validity is needed on children younger than 4 years of age.

Our focus on outcomes at one age window is liable to yield an underestimate of diagnostic stability and an overestimate of "remission" because CD symptoms wax and wane over time.[1] This study examined outcomes at only one age window because this is the situation that resembles what clinicians in practice are likely to face; rarely do they have the benefit of assessing a child on multiple occasions across development. Instead, clinicians will want to know if a preschool child is diagnosed with CD at age 5, what is the probability that he will be diagnosed again if seen by a clinician when in elementary school? What is the probability that, even if he is not diagnosed again with CD, he will continue to have educational difficulties and continue to require mental health and special education services? Should intervention be initiated early in development before problems become entrenched? On the one hand, our definition of "remission" at a single time point at age 10 may be an overestimate. Previous research has shown that whereas 50% of clinic-referred boys meet diagnostic criteria for CD at a single re-assessment, 88% of them meet diagnostic criteria at least once over a 3-year period [20]. On the other hand, our definition of "remission" is relatively conservative, given that remission is defined not as the absence of a CD diagnosis, but as the absence of any CD symptoms (i.e., children with one or two CD symptoms were not classified as "remitted"). Of the 115 children identified as "remitters" at age 10, only 19 (13.1% weighted) of them had met diagnostic criteria for CD at age 7 [18]. Thus, a large majority of our "remitters" appear to have "escaped" a CD diagnosis over more than a single follow-up window and yet continue to have elevated behavioural and educational concerns in middle-childhood.

Without making developmental modifications (aside from dropping the "forced sexual activity" symptom), this study demonstrated that existing *DSM-IV* criteria for CD can statistically predict continuity of behavioural and academic problems in non-referred preschool-aged children, which should help inform decisions regarding revisions for *DSM-V*. Increasingly, the controversy over preschool diagnosis appears to be subsiding and data indicate that psychopathology in preschoolers is similar in prevalence rates and diagnostic features to psychopathology seen in older children and adolescents [4]; however, more research is left do be done. Our findings support further efforts to design and validate developmentally appropriate diagnostic procedures for preschoolers so that greater consensus regarding clinical criteria can be achieved. For instance, studies are needed that specifically test whether frequency and duration criteria and symptom definition should differ for children in different developmental periods [36]. Fortunately, this is an opportune time for investigating potential refinements to the diagnostic criteria for CD prior to the release of the next *DSM*.

[1]To illustrate, 1-week test–retest reliability of the CD diagnosis assessed by the Kiddie Disruptive Behavior Disorders Schedule, a structured clinical interview, in a clinic-referred sample of pre-schoolers was 0.73 [15], which is very good but falls short of perfect even over a few days.

Acknowledgments

We are grateful to the Study families and teachers for their participation. We thank Robert Plomin and Michael Rutter for their contributions, Thomas Achenbach for kind permission to adapt the CBCL, and members of the E-Risk team for their dedication, insights, and hard work. This research was supported by grants from the UK Medical Research Council (G9806489 & G0100527), and the ESRC-SCOPIC Network. Dr. Caspi is a recipient of a Royal Society-Wolfson Research Merit Award.

References

1. Achenbach, TM. Manual for the Child Behavior Checklist/4–18 and 1991 profile. University of Vermont Department of Psychiatry; Burlington, Vermont: 1991.
2. Achenbach, TM. Manual for the Teacher's Report Form and 1991 Profile. University of Vermont Department of Psychiatry; Burlington, Vermont: 1991.
3. Anastasi, A. Psychological testing. Macmillian; New York: 1996.
4. Angold A, Egger HL. Preschool psychopathology: lessons for the life-span. J Child Psychol Psychiatry. 2007; 48:961–966. [PubMed: 17914996]
5. Bennett, N.; Jarvis, L.; Rowlands, O.; Singleton, N.; Haselden, L. Living in Britain: results from the General Household Survey. HMSO; London: 1996.
6. Bird HR, Gould MS, Staghezza B. Aggregating data from multiple informants in child psychiatry epidemiological research. J Am Acad Child Adolesc Psychiatry. 1992; 31:78–85. [PubMed: 1537785]
7. Series FM1 no 23. HMSO; London: 1996. Birth Statistics.
8. Cohen J. A power primer. Psychol Bull. 1992; 112:155–159. [PubMed: 19565683]
9. Costello, A.; Edelbrock, C.; Kalas, R.; Kessler, M.; Klaric, SA. Diagnostic Interview Schedule for Children (DISC). National Institute of Mental Health; Rockville, Maryland: 1982.
10. Costello EJ, Mustillo S, Erkanli A, Keeler G, Angold A. Prevalence and development of psychiatric disorders in childhood and adolescence. Arch Gen Psychiatry. 2003; 60:837–844. [PubMed: 12912767]
11. Cote S, Vaillancourt T, LeBlanc JC, Nagin DS, Tremblay RE. The development of physical aggression from toddlerhood to pre-adolescence: a nation wide longitudinal study of Canadian children. J Abnorm Child Psychol. 2006; 34:71–85. [PubMed: 16565888]
12. Egger, HL.; Angold, A. The pre-school Age Psychiatric Assessment (PAPA): a structured parent interviewfor diagnosing psychiatric disorders in pre-school children. In: DelCarmen-Wiggins, R.; Carter, AS., editors. Handbook of infant and toddler mental health assessment. Oxford; New York: 2004. p. 223-243.
13. Haddock CK, Rinkdskopf D, Shadish WR. Using odds ratios as effect sizes for meta-analysis of dichotomous data: a primer on methods and issues. Psychol Methods. 1998; 3:339–353.
14. Hutchings J, Bywater T, Daley D, Gardner F, Whitaker C, Jones K, Eames C, Edwards RT. Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial. Br Med J. 2007; 334:678. [PubMed: 17350966]
15. Keenan K, Wakschlag L, Danis B, Hill C, Humphries M, Daux J, Donald R. Further evidence of the reliability and validity of DSM-IV ODD and CD in preschool children. J Am Acad Child Adolesc Psychiatry. 2007; 46:457–468. [PubMed: 17420680]
16. Keenan K, Wakschlag LS. Can a valid diagnosis of disruptive behavior disorder be made in preschool children? Am J Psychiatry. 2002; 159:351–358. [PubMed: 11869995]
17. Keenan K, Wakschlag LS. Are oppositional defiant and conduct disorder symptoms normative behaviors in preschoolers? A comparison of referred and nonreferred children. Am J Psychiatry. 2004; 161:356–358. [PubMed: 14754786]
18. Kim-Cohen J, Arseneault L, Caspi A, Polo Tomas M, Taylor A, Moffitt TE. Validity of DSM-IV conduct disorder in 4 1/2–5-year-old children: a longitudinal epidemiological study. Am J Psychiatry. 2005; 162:1108–1117. [PubMed: 15930059]
19. Kuntsi J, Eley TC, Taylor A, Hughes C, Asherson P, Caspi A, Moffitt TE. The co-occurrence of ADHD and low IQ has genetic origins. Neuropsychologia. 2004; 124B:41–47.

20. Lahey BB, Hart EL, Frick PJ, Applegate B, Zhang Q, Green SM, Russo MF. Four-year longitudinal study of conduct disorder in boys: patterns and predictors of persistence. J Abnorm Psychol. 1995; 104:83–93. [PubMed: 7897057]
21. Maughan B, Rowe R, Messer J, Goodman R, Meltzer H. Conduct disorder and oppositional defiant disorder in a national sample: developmental epidemiology. J Child Psychol Psychiatry. 2004; 45:609–621. [PubMed: 15055379]
22. Maynard, RA. Kids having kids: economic costs and social consequences of teen pregnancy. Urban Institute Press; Washington: 1997.
23. Moffitt TE, Arseneault L, Jaffee SR, Kim-Cohen J, Koenen KC, Odgers C, Slutske W, Viding E. Research review: DSM-V conduct disorder: research needs for an evidence base. J Child Psychol Psychiatry. 2008; 49:3–33. [PubMed: 18181878]
24. Moffitt TE. The E-Risk Study Team. Teen-aged mothers in contemporary Britain. J Child Psychol Psychiatry. 2002; 43:727–742. [PubMed: 12236608]
25. Piacentini JC, Cohen P, Cohen J. Combining discrepant diagnostic information from multiple sources: are complex algorithms better than simple ones? J Abnorm Child Psychol. 1992; 20:51–63. [PubMed: 1548394]
26. Reid JB. Prevention of conduct disorder before and after school entry: relating interventions to developmental findings. Dev Psychopathol. 1993; 5:243–262.
27. Sattler, JM. Assessment of Children WISC-III and WPPSI-R Supplement. Jerome M Sattler; San Diego, California: 1992.
28. Sclare, I. The child psychology portfolio. NFER-Nelson Publishing Company; Windsor, Berkshire: 1997.
29. StataCorp. Stata statistical software: Release 8.0. College Station; Texas: 2003.
30. Sterba S, Egger HL, Angold A. Diagnostic specificity and nonspecificity in the dimensions of preschool psychopathology. J Child Psychol Psychiatry. 2007; 48:1005–1013. [PubMed: 17915001]
31. Torgesen, JK.; Wagner, RK.; Rashotte, CA. Test of Word Reading Efficiency. PRO-ED Inc; Austin: 1999.
32. Tremblay RE, Nagin DS, Seguin JR, Zoccolillo M, Zelazo PD, Boivin M, Perusse D, Japel C. Physical aggression during early childhood: trajectories and predictors. Pediatrics. 2004; 114:e43–e50. [PubMed: 15231972]
33. Trouton A, Spinath FM, Plomin R. Twins Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition, and behavior problems in childhood. Twin Res. 2002; 5:444–448. [PubMed: 12537874]
34. Wakschlag, L.; Danis, B. Assessment of disruptive behavior in young children: a clinical-developmental framework. In: DelCarmen-Wiggins, R.; Carter, AS., editors. Handbook of infant and toddler mental health assessment. Oxford; New York: 2004. p. 421-440.
35. Wakschlag LS, Briggs-Gowan MJ, Carter AS, Hill C, Danis B, Keenan K, McCarthy KJ, Leventhal BL. A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. J Child Psychol Psychiatry. 2007; 48:976–987. [PubMed: 17914998]
36. Wakschlag, LS.; Leventhal, BL.; Thomas, JM. Disruptive behavior disorders and ADHD in preschool children: characterizing heterotypic continuities for a developmentally-informed nosology for DSM-V. In: Narrow, WE.; First, MB.; Sirovatka, PJ.; Regier, DA., editors. Age and gender considerations in psychiatric diagnosis: a research agenda for DSM-V. American Psychiatric Publishing, Inc; Washington: 2007. p. 243-257.
37. Webster-Stratton C, Taylor T. Nipping early risk factors in the bud: preventing substance abuse, delinquency, and violence in adolescence through interventions targeted at young children (0–8 years). Prev Sci. 2001; 2:165–192. [PubMed: 11678292]
38. Wechsler, D. The Psychological Corporation. Harcourt Brace & Company; London: 1990. Wechsler Preschool and Primary Scale of Intelligence-Revised.
39. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000; 56:645–646. [PubMed: 10877330]

Table 1

Predictive validity: Comparison of children with and without a CD diagnosis at 5 years of age on outcome measures at age 10

Child functioning at age 10	Age 5 diagnostic grouping		Group difference	Effect size
	Children with CD (*N* = 184)	Children without CD (*N* = 1954)[a]		
Behavioural outcomes				
Mother's report	M (SD)	M (SD)	*t* (*df*)[b]	*d*[c]
Attention deficit hyperactivity (ADHD) symptom scale	14.8 (8.2)	8.0 (6.9)	9.55 (1068)***	0.89
Aggression scale	15.4 (8.7)	7.2 (6.0)	11.23 (1068)***	1.10
Delinquency scale	4.0 (3.3)	1.5 (1.9)	9.50 (1068)***	0.93
Emotional problem scale	9.7 (7.6)	6.2 (5.2)	5.43 (1068)***	0.54
Prosocial behaviour scale	15.0 (3.6)	17.2 (2.7)	7.55 (1068)***	0.70
Teacher's report				
ADHD symptom scale	6.9 (8.1)	3.3 (5.5)	4.98 (990)***	0.52
Aggression scale	9.3 (10.2)	3.6 (6.6)	6.70 (991)***	0.67
Delinquency scale	1.8 (2.5)	0.6 (1.4)	5.86 (989)***	0.62
Emotional problem scale	5.5 (6.2)	4.8 (5.7)	1.36 (991)	0.13
Prosocial behaviour scale	11.9 (4.9)	14.1 (4.6)	4.77 (982)***	0.47
	N (%)	*N* (%)	Odds ratio (95% CI)	
Received treatment for behavioural/emotional problems	65 (36.2)	286 (14.3)	3.4 (2.3, 5.1)***	0.68
Educational outcomes				
Special education service use	53 (27.3)	244 (11.6)	2.9 (1.9, 4.4)***	0.59
Standard reading score below average or poor	62 (30.1)	335 (15.0)	2.4 (1.6, 3.7)***	0.48
English school performance below average	83 (47.3)	512 (25.8)	2.6 (1.7, 3.8)***	0.53
Math school performance below average	73 (43.3)	465 (23.9)	2.4 (1.7, 3.5)***	0.48
	M (SD)	M (SD)	*t* (*df*)	
Teacher's effort scale	16.4 (9.1)	11.3 (7.8)	6.30 (989)***	0.60

[a] Five CD children and 89 comparison children were missing parent data at age 10; an additional 21 CD and 191 comparison children were missing teacher data at age 10. Ns are unweighted; proportions are weighted to represent the population of British families

[b] Continuous variables were analysed with t-tests and their degrees-of-freedom (*df*) and categorical variables with odds ratios (OR) and their 95% confidence intervals (CI). Standard errors, 95% CIs, and test statistics include adjustments for the dependence in the data due to analyzing two children in the same family [39]. Thus, degrees-of-freedom are based on number of families rather than number of children

[c] Differences between groups can be interpreted in terms of standard deviation units (*d*), where $d = 0.2$ is considered a small effect size, $d = 0.5$ is a medium effect size, and $d = 0.8$ is a large effect size [8]

*** $P \leq 0.001$

Table 2

Predictive validity: comparison of children with a CD diagnosis at age 5 and no CD symptoms at age 10 ("remitted") versus children with no CD diagnosis at age 5 (comparison group), on outcome measures at age 10

Child functioning at age 10	Age 5 diagnostic grouping		Group difference	Effect size
	"Remitted" group (CD at age 5 but no age-10 CD symptoms) (*N* = 115)	Comparison group (no age-5 CD) (*N* = 1,954)[a]		
Behavioural outcomes				
Mother's report	M (SD)	M (SD)	*t* (*df*)[b]	*d*[c]
ADHD symptom scale	12.1 (6.9)	8.0 (6.9)	6.04 (1057) ***	0.60
Aggression scale	11.6 (6.1)	7.2 (6.0)	6.31 (1057) ***	0.73
Delinquency scale	2.6 (2.2)	1.5 (1.9)	4.26 (1057) ***	0.55
Emotional problem scale	8.4 (7.0)	6.2 (5.2)	2.87 (1057) **	0.36
Prosocial behaviour scale	15.9 (3.3)	17.2 (2.7)	3.83 (1055) ***	0.45
Teacher's report				
ADHD symptom scale	4.7 (6.0)	3.3 (5.5)	1.95 (979) +	0.23
Aggression scale	6.1 (6.9)	3.6 (6.6)	3.50 (980) ***	0.37
Delinquency scale	1.1 (1.7)	0.6 (1.4)	2.48 (978) *	0.31
Emotional problem scale	4.3 (5.0)	4.8 (5.7)	0.73 (980)	0.08
Prosocial behaviour scale	13.1 (4.4)	14.1 (4.6)	2.01 (971) *	0.22
	N (%)	*N* (%)	Odds ratio (95% CI)	
Received treatment for behavioural/emotional problems	28 (25.3)	286 (14.3)	2.0 (1.2, 3.3) **	0.38
Educational outcomes				
Special education service use	28 (23.2)	244 (11.6)	2.3 (1.4, 3.9) **	0.46
Standard reading score below average or poor	29 (21.2)	335 (15.0)	1.5 (0.9, 2.5)	0.22
English school performance below average	40 (35.7)	512 (25.8)	1.6 (1.0, 2.6) +	0.26
Math school performance below average	33 (31.9)	465 (23.9)	1.5 (1.0, 2.3) +	0.22
	M (SD)	M (SD)	*t* (*df*)	
Teacher's effort scale	13.4 (8.4)	11.3 (7.8)	2.33 (978) *	0.26

[a] A total of 14 "remitted" children and 191 comparison children were missing teacher data at age 10. Ns are unweighted; proportions are weighted to represent the population of British families

[b] Continuous variables were analysed with t-tests and their degrees-of-freedom (*df*) and categorical variables with odds ratios (OR) and their 95% confidence intervals (CI). Standard errors, 95% CIs, and test statistics include adjustments for the dependence in the data due to analyzing two children in the same family [39]. Thus, degrees-of-freedom are based on number of families rather than number of children

[c] Differences between groups can be interpreted in terms of standard deviation units (*d*), where $d = 0.2$ is considered a small effect size, $d = 0.5$ is a medium effect size, and $d = 0.8$ is a large effect size [8]. Statistics for these analyses were not adjusted for age-5 IQ and ADHD diagnosis

* $P \leq 0.05$;

** $P \leq 0.01$;

$^{***}P \leq 0.001$;

$^{+}P \leq 0.10$