Published in final edited form as:

J Abnorm Psychol. 2010 November ; 119(4): 689–698. doi:10.1037/a0019346.

Age of onset, symptom threshold, and expansion of the nosology of conduct disorder for girls

Kate Keenan,

Department of Psychiatry and Behavioral Neuroscience, University of Chicago

Kristen Wroblewski,

Department of Health Studies, University of Chicago

Alison Hipwell,

Department of Psychiatry, University of Pittsburgh

Rolf Loeber, and

Department of Psychiatry, University of Pittsburgh

Magda Stouthamer-Loeber

Department of Psychiatry, University of Pittsburgh

Abstract

The study of conduct disorder (CD) in girls is characterized by several nosologic controversies that center on the most common age of onset, the most valid symptom threshold, and potentially including other manifestations of antisocial behavior and dimensions of personality as part of the definition of CD. Data from a prospective, longitudinal study of a community sample of 2,451 racially diverse girls were used to empirically inform these issues. Results revealed that adolescent-onset CD is rare in girls. There was mixed support for the threshold at which symptoms are associated with impairment: parent-reported impairment provided the clearest evidence of maintaining the current DSM-IV threshold of 3 symptoms. The impact of callousness and relational aggression on impairment varied by informant, with small effects for parent- and youth-reported impairment, and larger effects for teacher-rated impairment relative to the effects for CD. These results support arguments for revising the typical age of onset of CD for girls, but for maintaining the current symptom threshold. The results also suggest the need to consider subtyping according to the presence or absence of callousness. Given its content validity relational aggression requires further study in the context of ODD and CD.

Keywords

conduct disorder; girls; diagnosis; validity; phenotype

Introduction

Conduct Disorder (CD) is a disorder described as occurring more commonly among males than females regardless of age at assessment (American Psychiatric Association, 1994). In a large, community-based sample, the rate of CD in girls was below 1% in childhood and ranged from 1.4–3.3% at ages 13–15, whereas for boys the rate ranged from .5–2.8% in childhood and from 3.2–5.4% at ages 13–15 (Maughan et al., 2004). Similar patterns have been reported in other large epidemiologic samples, in which boys are reported to have higher rates of CD than girls across age, with a narrowing of the sex difference in adolescence due to an increase in new cases of CD among girls during adolescence (Moffitt, 2003). Such findings, however, are largely based on assessing the rate of CD at a single time

point in children at different ages, without determining the age at which the first symptom appeared (e.g., Maughan et al., 2004). The age at which girls meet criteria for CD (i.e., endorse 3 symptoms in a 12-month period) is often interpreted as the age of onset of CD, although that is not consistent with the DSM-IV definition of age of onset of CD. Moreover, some studies of the developmental course of antisocial behavior in girls have incorporated measures of delinquency and or used different informants at different ages (e.g., Moffitt et al., 2001) complicating the assessment of developmental patterns of CD. Thus, the developmental pattern of CD in girls is still not well documented, despite the fact that current etiological theories identify age of onset as a critical dimension in determining causal risk factors and prognosis (e.g., Moffitt, 2003).

Although the empirical basis for many childhood disorders in DSM-IV represented an improvement over previous versions of the DSM, the operationalization of DSM-IV CD was based on a very small number of girls. Only 25% of the 440 participants in the DSM-IV field trials for ODD and CD were female and only 5% (n=24) of these girls met criteria for CD (Lahey et al., 1994; 1998). Similarly, in the NIMH Methods for Epidemiology of Child and Adolescent Mental Disorders Study (MECA), which was also used to test the validity of DSM-IV CD, only 19 girls (1.5% of the total sample) were reported to meet criteria for CD (Lahey et al., 1998). Thus, the validity of DSM-IV CD for females has been questioned (e.g., Hartung & Widiger, 1998; Zoccolillo et al., 1996). One debate focuses on whether the symptom threshold used to determine presence or absence of CD reflects a gender bias in the definition of CD (Hartung & Widiger, 1998; Zoccolillo et al., 1996). Zoccolillo and colleagues (1996) have argued that a threshold of two symptoms of CD is a more sensitive and specific threshold than 3 symptoms for girls. Using a sample of girls who were well characterized in terms of disruptive behavior in early childhood, changing the threshold from three to two symptoms and adding a DSM-III symptom of rule violation resulted in a significant increase in the rate of DSM-III-R CD among girls who were reported to have high and persistent antisocial behavior in childhood (Zoccolillo et al., 1996). The definition of early antisocial behavior, however, was quite broad including symptoms of ADHD and ODD. Additional tests of optimal symptom threshold using alternative criteria are still needed.

Another debate centers on whether additional, female-sensitive symptoms should be included in the definition of CD such as relational forms of aggression (e.g., overtly excluding someone from play; Crick et al., 2006; Crick, & Grotpeter, 1995; Xie et al., 2002) and indirect aggression (e.g., spreading rumors and gossip; Bjorkqvist et al., 1992). Crick and Grotpeter (1995) and Bjorkqvist et al. (1992) have argued that girls are more likely to engage in indirect, than direct, aggression. Whether or not the addition of indirect aggression items to the pool of CD symptoms will increase the validity of the disorder for girls, by accounting for additional variance in impairment for example, remains to be tested. In the only study conducted to date, relational aggression did not appear to explain additional variance in impairment for girls (or for boys) after controlling for DSM-IV CD (Keenan et al., 2008). The rate of CD in that sample, however, was relatively low. In addition, there was no evidence that relational aggression occurred at a higher rate in girls than in boys (Keenan et al., 2008). These results notwithstanding, the research agenda for DSM-V CD includes determining the potential of relational aggression to increase the validity of CD for girls (Moffitt et al., 2008).

A third debate is whether dimensions of psychopathy, such as callousness and lack of emotion need to be incorporated in the defining constructs of childhood CD for both sexes (Frick, et al., 2000; Lynam, 1997; Pardini et al., 2006; Schrum & Salekin, 2006), but perhaps particularly for girls. Ratings of callousness have demonstrated unique moderate associations with disruptive behaviors in both community and clinic-referred samples of

school age children and early adolescents, (Frick et al., 2000). Frick and colleagues (2000) reported that 60% of children scoring high on narcissism but low on the other personality dimensions were girls, and that 25% of that group met criteria for ODD or CD compared with less than 1% of children scoring low on all the dimensions. Thus, the inclusion of personality dimensions that are theoretically linked to antisocial behavior across development may be useful in defining a subtype of CD in girls.

The goal of the present study is to examine the validity of the current DSM-IV nosology of CD in girls by addressing the following questions:

1. Is adolescent onset (i.e., the first CD symptom occurring at or later than age 10, as per DSM-IV) the most common age of onset among girls meeting criteria for DSM-IV CD?
2. Is there evidence to support a change in the symptom threshold for CD in girls, as demonstrated by similar levels of impairment at lower symptom thresholds?
3. Is there significant overlap between CD symptoms and callousness and relational aggression?
4. Do relational aggression and callousness explain unique variance in impairment after controlling for CD?

Providing data to answer the above questions will further shape, and hopefully narrow, the debate on identifying the most valid operational definition of CD for girls. In addition, we explore whether race moderates the findings on onset, symptom threshold, and utility of relational aggression and callousness to the diagnosis of CD.

Methods

Participants

In the Pittsburgh Girls Study (PGS), a stratified, random household sampling, with over-sampling of households in low-income neighborhoods, was used to identify girls who were between the ages of 5 and 8 years. Neighborhoods in which at least 25% of the families were living at or below the poverty level were fully enumerated (i.e., all homes were contacted to determine if the household contained an eligible girl), and a random selection of 50% of the households in non-risk neighborhoods were enumerated during 1998 and 1999. The enumeration identified 3,118 separate households in which an eligible girl resided. From these households, families who moved out of state and families in which the girl would be age ineligible by the start of the study were excluded. When two age-eligible girls were enumerated in a single household, one girl was randomly selected for participation. Of the 2,992 eligible families 2,875 (96%) were successfully re-contacted to determine their willingness to participate in the longitudinal study. Of those families, 85% agreed to participate resulting in a total sample size of 2,451. The 2,451 girls were relatively evenly distributed across the four age groups (5–8 years). Approximately half of the girls were African American (52%), 41% were European American, and the remaining girls were described as multiracial or representing another race. Nearly all the primary caregivers were biological mothers (92%). More than half of the caregivers were cohabiting with a husband or partner, about 47% of parents had completed 12 years or less of education, and 25% of the families had yearly incomes of less than \$15,000.

Retention of the sample has been very high. In Table 1 we present retention data by age, which ranges from a high of 97.5% for age 7 to 87.8% for age 15 data. We note that only two cohorts have been assessed at age 14 and only one at age 15, to date. Some of the variability in retention from year to year is due to difficulty tracking participants; a minority

of families has refused to participate over the years (< 3% at age 15). Comparisons of those assessed and those not assessed at each age were conducted using chi-square tests. Girls who were not assessed at ages 10 through 15 years were more likely to be from families not receiving public assistance; European American girls were less likely to have been assessed at ages 11, 12, and 14 years. In addition, there was no difference in number of CD symptoms or CD diagnosis at the initial assessment between girls who were and were not retained.

The University of Pittsburgh Institutional Review Board approved all study procedures. Written informed consent was obtained from the primary caregiver and verbal assent from the child.

Measures

Information on girls' age, race, and whether the household was in receipt of public assistance (e.g. WIC, food stamps, welfare) was collected by parental report. For these analyses, Asian American girls and girls with unknown race ($n = 26$) were excluded, leaving comparisons between African American girls and European American girls.

The *Child Symptom Inventory-4* (CSI-4, Gadow & Sprafkin, 1994) was used to assess DSM-IV symptoms of CD over the past year via parent and child report. Symptoms were scored on a 4-point scale: *never*, *sometimes*, *often*, or *very often*. CD symptoms that require the frequency to be often (i.e., bullies, fights, lies to con, stays out late, truancy) were scored as present if either informant endorsed the behavior at the level of *often* or *very often*, and all other CD symptoms were scored as present if endorsed at any level of frequency by either informant. As per DSM-IV, age of onset of CD was defined as age at which the first symptom was reported. Internal consistency for the CD scale was high for parent and youth across age (.69–.79).

DSM-IV Oppositional Defiant Disorder (ODD) was assessed using parent report on the CSI. A threshold of four or more symptoms of ODD endorsed at the level of *often* or *very often* was used to define presence or absence of DSM-IV ODD (alpha ranging from .83–.90 across ages). Adequate concurrent validity, and sensitivity and specificity of ODD and CD symptom scores to clinicians' diagnoses are reported for the CSI (Gadow & Sprafkin, 1994).

Parents and teacher were administered the relational aggression subscale of the *Children's Peer Relationship Scale* (CPRS, Crick & Grotpeter, 1995). Concurrent and predictive validity has been established via negative associations with peer acceptance and liking and psychological adjustment (Crick & Grotpeter, 1995; Crick, Ostrov, & Werner, 2006). Internal consistency for this scale was high for parent and teacher reports across age (range = .84–.95). Consistent with the approach to defining a positive endorsement for CD symptoms, the highest level of endorsement by either informant was used to generate a total score for relational aggression (range = 5–25). Girls scoring 1 standard deviation above the mean (at or above approximately the 85th percentile) for the sample (> 15) were defined as relationally aggressive.

The callous/unemotional subscale from the *Psychopathy Screening Device* (PSD, Frick et al., 2000) was administered to the parent and teacher. This subscale has shown good predictive validity by predicting level of severity and stability of antisocial behavior among children with conduct problems (Frick et al., 2005). The highest level of endorsement by parent or teacher was used to generate a callousness score (range = 0–8). Internal consistency was moderate (alphas = .51–.67 across age and informant). Girls scoring 1 standard deviation above the mean for the sample (> 4) were defined as callous.

Parent-rated impairment was assessed via the *Child Global Assessment Scale* (C-GAS; Setterberg et al., 1992), which is a measure of impairment developed for children 4–18 years of age, which has been validated for use by parents (Bird et al., 1996). Scores on the C-GAS range from 1 to 100 with each decile containing a description of the degree of impairment in school, and with family and peer relations. Parent-reported impairment was operationally defined as C-GAS scores of 60 or below (Bird et al., 1996).

Teachers provided global ratings of the child's functioning in school by responding to two questions each scored on a 4-point Likert scale: "During the last two months, how often have you gotten annoyed or upset with this student?" and "During the last two months, how happy has this student seemed?" (scored in the reverse direction). These two scores were combined (range = 2–8), and girls whose scores fell 1 standard deviation above the mean (> 5) were defined as significantly impaired by teacher report. The impairment questions were designed to assess global impairment in psychosocial functioning in the school environment. Data on teacher-rated impairment were available through age 13 years of age.

Children reported on functioning in their peer relations using the total score on the *Loneliness and Social Dissatisfaction Questionnaire* (LSDQ; Asher & Wheeler, 1985) from ages 7 to 10 (alphas = .86–.91 across age), and the self-competence score on the *Perception of Peers and Self* (POPS; Rudolph et al., 1995) at ages 10 through 14 (alphas = .45–.52 across age). Both scales are used to assess the child's perception of her functioning with peers. Changes in the administration of these scales were based on the developmental appropriateness of the items. Administration of the two scales overlapped at age 10, which yielded a moderate level of correlation (*Spearman* $r = 0.46$, $p < 0.001$). Children who score high on the LSDQ have been reported by teachers to be more aggressive and disruptive, but not more shy, than children who score low on the scale (Cassidy & Asher, 1992).

Data reduction and analyses

The PGS uses an accelerated longitudinal design, with relatively equal numbers of girls at ages 5, 6, 7, and 8 years being enrolled in the study at wave 1, followed by annual assessments. For rate and weighted prevalence of DSM-IV CD, we present data from ages 7 through 15, the oldest age for which we have data available on CD symptoms. For all other analyses we included a developmental span from ages 7 to 14 years with the exception of analyses using teacher-rated impairment as the dependent measure, which included ages 7 to 13 years.

All analyses were conducted with weighted data to correct for the over-sampling of the low-income neighborhoods in order to generate prevalence rates that are representative of the population in the City of Pittsburgh. We compared the ratio of girls living in low income versus non-low-income neighborhood in the PGS ($40.92/59.08 = .6926$) to the ratio from the Census data ($27.59/72.41 = .3810$), and divided the two to determine how much more weight should be assigned to a girl from a non-low-income neighborhood $1:1.8178$. Using this weighting factor and the n for the two groups a weight of .6742 was assigned to girls from the low-income neighborhoods and 1.2257 to the girls from the non-low-income neighborhoods.

Given the large sample size, significant differences are reported for p values that are less than .01.

The rate of CD was generated across age and race. Comparisons of the prevalence between groups were conducted using logistic regression analysis, and the difference in the distribution of age of onset by race was tested via Wilcoxon rank-sum test. Generalized estimating equations (GEE, Zeger & Liang, 1986) regression models using STATA

software, Version 11(StataCorp, College Station, TX)were used to test the symptom threshold and the relative contribution of callousness and relational aggression to impairment, after controlling for a diagnosis of CD, and to test the interactive effects of CD and callousness and CD and relational aggression on impairment. GEE is appropriate for nested, longitudinal designs in part because it can specify a working correlation matrix that accounts for within-subject correlations of repeated observations over multiple data waves. Accounting for the correlation structure of the data avoids the assumption that measurements taken at successive points in time are not correlated. This results in a more efficient analysis, unbiased regression parameters, and improved power to detect significant changes over time. GEE models can also be used with unbalanced designs (Diggle et al., 1994) in which some children provide more data points than others, as is the case in the current design.

Results

Prevalence of DSM-IV CD by race and poverty and co-occurrence of CD and ODD

Of the 2,393 girls(i.e. the analytic sample, excluding the 26 girls based on race and 32 girls missing CD data in all waves), 560 (21.2% weighted) met criteria for DSM-IV CD in at least 1 assessment year. The weighted prevalence of CD ranged from 4.9 % at age 11 years to 8.9% at age 15 years (Table 1). One hundred twenty-four European American girls (12.4% weighted) and 436 African-American girls (30.1% weighted) met criteria for CD in at least one assessment year, yielding a significant effect of race on prevalence ($OR = 3.0$, 95% CI (2.5–3.8), $p < 0.001$). Among families never receiving public assistance, 67 of 340 African American girls (19.2% weighted) and 53 of 674 European American girls (7.9% weighted) met criteria for DSM-IV CD ($OR = 2.8$, 95% CI [1.9–4.0], $p < 0.001$). Among girls whose families had received public assistance in at least one year, 369 of 1,062 African American girls (34.1% weighted) and 71 of 317 European American girls (22.1% weighted) met criteria for DSM-IV CD at least once ($OR = 1.8$, 95% CI [1.4–2.4], $p < 0.001$). Thus, poverty is associated with an approximate doubling of the rate of CD in both racial groups, but the effect of race on the rate of CD remains after controlling for poverty. This is further supported by the lack of a statistically significant race by public assistance interaction ($p = 0.08$) and by the OR for race, after adjusting for receipt of public assistance, remaining statistically significant ($OR = 2.1$, 95% CI [1.7–2.7], $p < 0.001$). These results are consistent with those reported by Bird and colleagues (2001), who found that race/ethnicity was associated with the rate of CD in the MECA study in a way that appeared to be independent of poverty.

Of the 560 girls who met criteria for CD in at least 1 assessment year, 255 (47.3% weighted) also met criteria at least once for DSM-V ODD. Of the 1,833 girls who never met criteria for CD, 177 (10.0% weighted) met criteria at least once for DSM-IV ODD.

1. Is adolescent onset (i.e., the first CD symptom occurring at or later than age 10) the most common age of onset among girls meeting criteria for DSM-IV CD?

Age of onset of CD was defined as the age at first reported symptom among the 560 girls who met criteria for DSM-IV CD. The majority of girls had childhood onsets and this did not vary significantly by race ($p = 0.24$ by Wilcoxon rank-sum test). Of the girls who met criteria for DSM-IV CD, 504 (89.8% weighted) had an age of onset between 7 and 9 years of age and 56 (10.2% weighted) had an age of onset between 10 and 15 years of age (see Figure 1). An age of onset after 9 years of age, therefore, was rare in this sample. In fact, 62% of the 560 girls who met criteria for CD had three symptoms within a 12-month period before the age of 10 years.

Among girls meeting criteria for CD, the most common symptoms reported at 7, 8, and 9 years of age were destruction of property (41.3–44.0%), stealing without confrontation (37.9–40.8%), cruelty to others (22.6–24.0%), and lying to con (17.6–24.6%).

2. Is there evidence to support a change in the symptom threshold for DSM-IV CD in girls, as demonstrated by similar levels of impairment at lower symptom thresholds?

—The proportion of girls with impairment at each symptom level of CD (none, 1, 2, and 3 or more) was examined for each of the three measures of impairment (parent-, teacher-, and youth-rated). Results are depicted in Figure 2. Using parent-reported impairment the threshold of 3 symptoms is generally supported. The rate of impaired girls increased by 100% from 0 to 1 symptoms (4%–9%), 50% from 1 to 2 symptoms (9% to 14%), and 100% from 2 to 3 or more symptoms (14–28%). For teacher and youth-reported impairment, the pattern was more linear.

The likelihood of impairment for girls with 3 or more symptoms was compared to the likelihood for girls with none, one, and two symptoms of CD. A lack of a significant difference between symptom levels would suggest that functional impairment is equivalent at each level. Results are presented in Table 2. The likelihood of parent-rated impairment was significantly higher at 3 or more symptoms than at 0, 1, or 2 symptoms. For example, the likelihood of impairment was twice as high among girls manifesting 3 symptoms compared to those with 2 symptoms ($OR = 2.1 [1.6–2.8]$, $p < 0.001$). This effect did not vary by race.

The same was true for teacher-rated impairment, although the magnitude of effect was not quite as strong: the likelihood of impairment was increased by 70% among girls manifesting 3 symptoms compared to those with 2 symptoms ($OR = 1.7 [1.3–2.2]$, $p = 0.002$). Race did not interact with CD symptoms to predict teacher rated impairment. According to youth report of impairment, there was no significant difference in the likelihood of being impaired between those with 2 and those with 3 or more symptoms ($OR = 1.3 [1.0–1.6]$, $p = 0.018$); this was true across racial groups.

3. Is there significant overlap between CD symptoms and callousness and relational aggression?

—A high level of callousness (i.e., greater than or equal to 1 SD above the mean for the sample) was reported at least once for approximately 44% of girls (40.5% weighted). High scores on callousness were more common among girls who met criteria for CD (68.2%; 65.5% weighted) than among girls who did not meet criteria for CD (36.9%; 33.8% weighted). This almost two-fold increase in the rate of high level of callousness was statistically significant ($OR = 3.7$, 95% $CI [3.0–4.6]$, $p < 0.001$) and was not affected by race. Most girls who scored high on callousness (63.9%; 65.7% weighted), however, did not meet criteria for CD.

Nearly half of the sample (46.1%; 43.4% weighted) scored a standard deviation above the mean on relational aggression in at least one year. High scores on relational aggression were more common among girls who met criteria for CD (73.8%; 72.5% weighted) than among girls who did not meet criteria for CD (37.6%; 35.5% weighted). This level of overlap was statistically significant ($OR = 4.8$, 95% $CI (3.8–5.9)$, $p < 0.001$) and was not affected by race. Most girls who scored high on relational aggression (62.5%; 64.6% weighted) did not meet criteria for CD.

4. Do relational aggression and/or callousness explain unique variance in impairment after controlling for DSM-IV CD and impairment at the previous assessment?

—Because there were no a priori hypotheses about the nature of the association between callousness and relational aggression, and because the research to date

has been conducted separately, their respective contribution to concurrent and later impairment controlling for CD was tested in separate GEE models. In each model, prior impairment (at time T-1), diagnosis of CD, and either callousness or relational aggression, were entered as main effects. The interaction between CD diagnosis and callousness/relational aggression also was tested. Interaction terms were dropped from the model if not statistically significant. For the predictive models, diagnosis of CD at time T-1 and either callousness or relational aggression at time T-1 were included to examine their contribution to impairment at time T. For the concurrent models, diagnosis of CD at time T and either callousness or relational aggression at time T were included with impairment at time T as the dependent variable. Results reported in Tables 3 and 4 are from weighted analyses.

Due to the accelerated longitudinal design of the PGS, sample sizes varied by age. Although unbalanced designs are permitted with GEE regression modeling, to verify that girls who were included at the older ages were not unrepresentative of the sample as a whole, the interaction terms involving the cohort the participant belonged to (cohorts 5, 6, 7, or 8) and prior impairment at time T-1 were tested in models predicting impairment at time T. None were found to be statistically significant (data not shown).

Across the three indices of impairment, callousness was significantly associated with a greater likelihood of impairment both concurrently and prospectively, with one exception (Table 3). Relational aggression accounted for increased risk of parent-and teacher rated impairment both concurrently and prospectively, but not youth-reported impairment (Table 4).

The pattern of association between relational aggression and callousness and impairment, however, differed by informant. For parents, there was a three-to four-fold increase in the odds of concurrent and later impairment as a function of meeting criteria for CD, with relational aggression and callousness generating odds ratios of about 2.0 or less. The reverse was true according to teacher-rated impairment; callousness and relational aggression were associated with a four-to five-fold increase in the odds of concurrent impairment, whereas the association between impairment and CD diagnosis was weaker. Prospectively, however, the effects of relational aggression and callousness on teacher-rated impairment were diminished ($OR = 1.9$). For youth report, CD increased the risk of impairment modestly, and only callousness accounted for unique variance in the risk for impairment, and only concurrently.

Only one significant interaction effect was detected: the interaction of relational aggression and CD on concurrent teacher-rated impairment ($OR = 0.4$, 95% $CI [0.3-0.7]$ $p = .001$) (Table 4). The effect, however, was in the opposite direction that one would have expected. Among those not relationally aggressive, the odds of teacher reported impairment given the presence of CD was 3.1 times greater than the odds given the absence of CD ($p < 0.001$). In contrast, among girls who were high on relational aggression the odds of teacher reported impairment given the presence of CD was only 1.3 times greater than the odds given the absence of CD ($p = 0.076$).

Discussion

Results from the present study support and extend the existing research on DSM CD in girls. The use of a prospective, longitudinal dataset, beginning in early childhood, provided an opportunity to generate information about the developmental course, symptom threshold, overlap with and incremental validity in risk for impairment resulting from relational aggression and callousness that extends the results generated via cross-sectional studies and studies incorporating measures of delinquency.

In the present sample, in which DSM-IV CD symptoms were assessed in girls from ages 7–15 years by parents and youth, the lifetime, weighted prevalence of CD was 21.2%, the prevalence across age ranged from 4.9–8.9%, and the most common age of onset was prior to the age of 10 years. Approximately half of the more than 500 girls who met criteria for CD in at least 1 year during the period of assessment were reported to have manifest the first symptom at 7 years of age, and close to 90% who met DSM-IV criteria for CD had an onset before age 10.

At first glance, our results on prevalence and age of onset appear to be in contrast to the findings from large epidemiologic studies such as the B-CAMHS99 described by Maughan and colleagues (2004) and the Dunedin Longitudinal Study (Moffitt et al., 2001). There are a few explanations for this. The first is that the period of assessment in the present study did not extend far enough into adolescence and therefore a significant increase in prevalence during adolescence was not observed. Maughan et al. (2004) reported that the rate of CD in girls in the B-CAMHS99 increased from 0.3% at age 12 to 1.3% at age 13, 2.0% at age 14 and 3.3% at age 15. Thus, because we only assessed girls through age 15 and had the fewest observations at that age, our estimates of onset are biased toward the younger ages.

However, B-CAMHS99 is a cross-sectional study, not longitudinal. Thus, the rate of CD at each age is generated by separate groups of girls. Moreover, because the base rate of CD in the B-CAMHS99 was low ($n=42$), age comparisons are based on small N s, potentially yielding somewhat unreliable comparisons. Two possibilities may explain the very low rate of CD in the B-CAMHS99. Symptoms of CD were assessed via a screening measure, the response to which determined whether the remaining CD symptoms should be assessed. Following this, clinician based diagnoses were generated. It is possible that the screening measure was not very sensitive for girls. In addition, the demographics of the CAMHS99 sample are much different than that in the PGS, even after weighting back to the population of the City of Pittsburgh. In the B-CAMHS99, 91% of the participants were white, and about 20% were living in single parent households.

Another large epidemiologic and prospective study to which our data should be compared is the Dunedin Longitudinal Study, in which the rate of CD and development of new cases in girls was reported from ages 11–18 (Moffitt et al., 2001). In that study, the estimated prevalence of DSM-IV CD in girls was 18%, 16%, 22%, and 26% at ages 11, 13, 15, and 18, yielding a lifetime prevalence of 46%. These rates are higher than what were reported in the PGS and demonstrate only modest variability across ages 11–15 years. Because the rate of DSM-IV CD was so high, the authors used a cut-off of 5 or more CD symptoms for all further analyses on sex differences in CD. This new cut-off resulted in prevalence rates of 3%, 5%, 8%, and 3%, respectively. Although the “peak” for girls in both cumulative and new cases (i.e. 5 or more CD symptoms in the past 12 months) was age 15 (Moffitt et al., 2001), age of onset of the first symptom was not ascertained. One could reasonably posit, however, that accumulating 5 or more CD symptoms in a single year, would be relatively rare, and thus the most common age of onset, was probably less than 15 years of age.

We note that impairment was not used to define CD in the present study. In the DSM-IV, an impairment criterion was added for many mental disorders, including CD. There was little empirical support for adding the criterion. Recent evidence for depression using data from the National Comorbidity Survey Replication suggests that an impairment criterion is redundant with the distress and impairment inherent in the symptoms (Wakefield, Schmitz & Baer, in press). For CD, in which the symptoms are meant to reflect violating the basic rights of others and violations of social norms, the same case could be made vis-à-vis “impaired social functioning.” We suggest that future studies of CD provide prevalence data

with and without the DSM-IV impairment criteria, so that comparisons can be made based on diagnoses derived from symptom criteria.

In only a few studies has age of onset of CD, as defined by the age at which the first symptom is manifest, in girls been assessed, but the results are consistent with our findings. Lahey et al (1998) reported age of onset of CD using retrospective report in two separate samples: a clinical sample of 4–17 year olds and a household sample of 9–17 years olds, both of which were used for the DSM-IV field trials. Among the 24 girls meeting criteria for CD in the clinic sample, 15 (62.5%) reported a childhood onset. Of the 19 girls with CD in the household sample, 14 (73.7%) had a childhood onset (Lahey et al., 1998). In a study designed to test the theory that girls with CD followed a delayed onset pathway, McCabe et al. (2004) found that among girls with a history of social service use, close to half had an onset of CD before the age of 10. Prospective studies are the best method for determining age of onset of CD, but none have been conducted in which CD symptoms are assessed from early in life. Cote et al. (2001) reported that of 28 girls diagnosed with DSM-IV CD at age 15, 64.3% were characterized as following medium-to high disruptive behavior problem trajectories from ages 6–12.

Together, data on age of onset assessed via retrospective recall or estimated from prospective studies of broad measure of disruptive behavior problems are generally consistent with those of the present study, in which a prospective assessment of age of onset was conducted, and suggest that the most common pathway to CD for girls is via an *exacerbation or intensification* of symptoms from childhood to adolescence rather than initiation or acute onset of CD during adolescence. Clearly, such findings are highly significant for prevention and intervention research, with the present data suggesting that indicated preventions for the majority of girls should target the period of early to middle childhood. Moreover, these results call into question approaches to conceptualizing CD in girls that have at their foundation the assumption that most girls begin to manifest CD in adolescence, such as the “delayed-onset pathway” (Silverthorn & Frick, 1999).

Regarding the test of symptom threshold, the results vary by informant. For parent reported impairment using the C-GAS, which has substantial data supporting its validity as a measure of global impairment, there is a significant difference in the likelihood of having a C-GAS score that falls at or below 60 at the level of 3 or more symptoms compared with 2 symptoms. The same result was found for teachers, but the comparison was not significant when the informants were the girls themselves. Because adults are the most commonly used informants regarding a child’s level of functional impairment, one could make an argument that the results using parent-and teacher-report should be weighed more heavily than youth-report in determining symptom threshold for a disorder. This could be refuted, however, by the possibility that when youth report specifically on the domain within which they are expert, their reports of impairment demonstrate greater validity, and thus generating more reliable results.

Overall, it is difficult to fully defend the threshold of 3 symptoms based on the results of the present study, but there is even less support for changing the symptom threshold. The association between impairment and number of symptoms was essentially linear findings for all three informants, although less so for parent informants. The linearity of the results, however, does not necessarily support lowering the threshold from 3 symptoms to 2 symptoms. Rather, the results renew the debate regarding the artificiality of categorical diagnoses and the loss of information on severity when ignoring the dose response nature of the association between symptoms of CD and outcomes (e.g., Fergusson & Horwood, 1995).

Finally, we addressed the extent to which callousness and relational aggression add to the concurrent and predictive association with impairment among girls with CD. These two constructs evolve from different theoretical traditions. Assessing individual differences in callousness in children has emerged as a means by which to link personality and disorder in children and identify a possible subtype of CD that is likely to persist into adulthood (Frick, et al., 2000). Relational aggression has been proposed as the primary means by which girls engage in aggressive acts, and thus serves as a female-specific form of aggressive behavior that could be included in the diagnostic nosology.

With regard to expanding the nosology to include either relational aggression or callousness, it is important to note that the current approach to identifying atypical rates of relational aggression and callousness (i.e., using a standard deviation above the mean to define high scores; Crick & Grotpeter, 1995; Frick et al., 2000) generates a lifetime rate that it approximately twice as high as the rate of CD. Although the overlap is significant, the majority of high scorers on relational aggression and callousness do not meet criteria for CD. This means that simply adding symptoms of relational aggression and callousness would likely result in a significant increase in the rate of CD.

One approach to determining whether these two constructs would be useful expansions of the current CD nosology is whether they explain unique variance in impairment, and/or interact with CD to predict impairment. In general, the two constructs do provide additional information regarding current and later impairment after controlling for CD and previous level of impairment. In fact, teachers appear to experience callousness and relational aggression as more impairing than CD, although the magnitude of effect is not maintained from one year to the next. The criterion of explaining unique variance, however, was essentially met.

Given the fact that the rate of callousness and relational aggression is relatively high, an alternative approach to simply adding symptoms to the current nosology is to consider including subtypes. For callousness, this appears to be a viable option. The items are not better measured by another childhood disorder, and conceptually it would provide a link to personality dimensions relevant to the study of antisocial behavior. The lack of significant interaction effects between callousness and CD, however, calls for caution in deciding how to move forward in incorporating psychopathy into the DSM nosology for CD in girls.

The data from the present study on relational aggression provide little guidance on how to proceed. Relational aggression explained unique variance in current and later parent- and teacher-rated impairment, but not youth-rated impairment. The relationship between CD and teacher-reported impairment appeared to be moderated by relational aggression status, but in a way that was unexpected: the odds ratio for CD (1.3) among girls with high relational aggression was lower compared to that for girls with low relational aggression (OR = 3.1). Although this may be a spurious finding, the results suggest less stability in the nature of the association between relational aggression, CD, and teacher-reported impairment. From a measurement perspective, it is still not clear whether relational aggressive behaviors overlap too highly with ODD symptoms. Most items measuring relational aggression (e.g., spreading rumors about someone in order to make other kids not like that person) appear to be more consistent with the ODD symptom 'spiteful and vindictive.' Further tests of the utility of relational aggression in the context of both CD and ODD are needed. It may be that once a diagnosis of ODD is included in the model, relational aggression no longer provides unique information on impairment.

Despite the effect of race on the rate of CD, even after controlling for poverty, there were no significant effects of race on age of onset, symptom threshold, overlap with relational

aggression and callousness, or utility of relational aggression and callousness to the diagnosis of CD in this sample of girls. Explaining the disproportionately higher rate of CD among African American girls will require further investigation that incorporates measures of perceived discrimination (Coker et al., 2009) and neighborhood context (Zalot et al., 2007), which have demonstrated empirical links to girls conduct problems, and which are associated but do not completely overlap with receipt of public assistance. The evidence for the *validity* of the DSM-IV CD for African American girls, however, is equally strong as the evidence for the validity of DSM-IV CD for European American girls.

Although, the lack of males in the present sample precludes the testing of sex differences in the present study, the results do have implications for the descriptions of purported sex differences in CD in DSM-V, and the specificity of the language used to describe those differences. For example, in the text accompanying the description of DSM-IV CD the statement that, “The ratio of males to females with Conduct Disorder is lowest for the Adolescent-Onset Type than for the Childhood-Onset Type (APA, 1994; p. 87),” is unlikely to be true given the rarity of adolescent-onset CD in the present sample, and is not consistent with data from the MECA study, in which the proportion of girls among youth with the adolescent-onset and childhood-onset types were generally equivalent (Lahey et al., 1998). The statement that, “Whereas confrontational aggression is more often displayed by males, females tend to use more non-confrontational behaviors,” may be somewhat misleading. “Non-confrontational aggression” behaviors are not listed in DSM-IV CD, so it is unclear how this construct would be operationally defined. If relational aggression was being considered as “non-confrontational,” then it will be important to first confirm that such behaviors provide evidence of construct validity before listing them as alternative manifestations of aggression within the diagnosis of CD.

In summary, the most definitive finding of the present study was that the onset of CD for most girls is in childhood. The lack of a definitive finding on symptom threshold speaks to the continued struggle of forcing a set of behaviors with a continuous distribution into categories, but this is not specific to girls, or to CD for that matter. Thus, there is no evidence supporting the need to lower the threshold of CD from three to two symptoms, and minimal support for maintaining the current symptom threshold of three symptoms. There does appear to be some support for continued exploration of callousness as a possible subtype of an additional set of symptoms within the diagnosis of CD, but the recommendation for relational aggression is to conduct further tests of its utility in the context of both CD and ODD.

Acknowledgments

This work was supported by grants R01 MH56630, R01 MH66167, and K01 MH71790 from the National Institute of Mental Health.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4. Washington, DC: 1994.
- Asher SR, Wheeler VA. Children’s loneliness: A comparison of rejected and neglected peer status. *Journal of Consulting and Clinical Psychology* 1985;53:500–505. [PubMed: 4031205]
- Bird HR, Andrews H, Schwab-Stone M, Goodman S, Dulcan M, Richters J, Rubio-Stipec M, Moore RE, Chiang P, Hoven C, Canino G, Fisher P, Gould MS. Global measures of impairment for epidemiologic and clinical use with children and adolescents. *International Journal of Methods in Psychiatric Research* 1996;6:295–307.
- Bird HR, Canino GJ, Davies M, Zhang H, Ramirez R, Lahey BB. Prevalence and correlates of antisocial behaviors among three ethnic groups. *Journal of Abnormal Child Psychology: An official*

- publication of the International Society for Research in Child and Adolescent Psychopathology 2001;29:465–478.
- Björkqvist K, Lagerspetz KM, Kaukiainen A. Do girls manipulate and boys fight? Developmental trends in regard to direct and indirect aggression. *Aggressive Behavior* 1992;18:117–127.
- Cassidy J, Asher SA. Loneliness and peer relations in young children. *Child Development* 1992;63:350–365. [PubMed: 1611939]
- Coker TR, Elliott MN, Kanouse DE, Grunbaum JA, Schwebel DC, Gilliland MJ, Tortolero SR, Peskin MF, Schuster MA. Perceived racial/ethnic discrimination among fifth-grade students and its association with mental health. *American Journal of Public Health* 2009;99:878–84. [PubMed: 19299673]
- Côté S, Zoccolillo M, Tremblay RE, Nagin D, Vitaro F. Predicting girls' conduct disorder in adolescence from childhood trajectories of disruptive behaviors. *Journal of the American Academy of Child & Adolescent Psychiatry* 2001;40:678–684. [PubMed: 11392346]
- Crick NR, Grotpeter JK. Relational aggression, gender, and social-psychological adjustment. *Child Development* 1995;66:710–722. [PubMed: 7789197]
- Crick NR, Ostrov JM, Werner NE. A longitudinal study of relational aggression, physical aggression, and children's social-psychological adjustment. *Journal of Abnormal Child Psychology* 2006;34:131–142. [PubMed: 16741683]
- Diggle, PJ.; Liang, KY.; Zeger, SL. *Analysis of longitudinal data*. Oxford: Clarendon Press; 1994.
- Fergusson DM, Horwood LJ. Predictive validity of categorically and dimensionally scored measures of disruptive childhood behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry* 1995;34:477–485. [PubMed: 7751262]
- Frick PJ, Bodin SD, Barry CT. Psychopathic traits and conduct problems in community and clinic-referred samples of children: Further development of the Psychopathy Screening Device. *Psychological Assessment* 2000;12:382–393. [PubMed: 11147105]
- Frick PJ, Stickle T, Dandreaux D, Farrell J, Kimonis E. Callous-unemotional traits in predicting the severity and stability of conduct problems and delinquency. *Journal of Abnormal Child Psychology* 2005;33:471–487. [PubMed: 16118993]
- Gadow, K.; Sprafkin, J. *Child Symptom Inventories Manual*. Stonybrook, NY: Checkmate Plus; 1994.
- Hartung C, Widiger T. Gender differences in the diagnosis of mental disorders: Conclusions and controversies of the DSM-IV. *Psychological Bulletin* 1998;123:260–278. [PubMed: 9602559]
- Keenan K, Coyne C, Lahey BB. Should relational aggression be included in the DSM-IV nosology for disruptive behavior disorders? *Journal of the American Academy of Child and Adolescent Psychiatry* 2008;47:86–93. [PubMed: 18174829]
- Lahey BB, Applegate B, Barkley RA, Garfinkel B, McBurnett K, Kerdyk L, Greenhill L, Hynd GW, Frick PJ, Newcorn J, Biederman J, Ollendick T, Hart E, Perez D, Waldman I, Shaffer D. DSM-IV field trials for oppositional defiant disorder and conduct disorder in children and adolescents. *American Journal of Psychiatry* 1994;151:1163–1171. [PubMed: 8037251]
- Lahey BB, Loeber R, Quay HC, Applegate B, Shaffer D, Waldman I, Hart EL, McBurnett K, Frick PJ, Jensen PS, Dulcan MK, Canino G, Bird HR. Validity of DSM-IV subtypes of conduct disorder based on age of onset. *Journal of the American Academy of Child & Adolescent Psychiatry* 1998;37:435–442. [PubMed: 9549965]
- Lynam DR. Pursuing the psychopath: Capturing the fledgling psychopath in a nomological net. *Journal of Abnormal Psychology* 1997;106:425–438. [PubMed: 9241944]
- Maughan B, Rowe R, Messer J, Goodman R, Meltzer H. Conduct Disorder and Oppositional Defiant Disorder in a national sample: Developmental epidemiology. *Journal of Child Psychology and Psychiatry* 2004;45:609–621. [PubMed: 15055379]
- McCabe KM, Rodgers C, Yeh M, Hough R. Gender differences in childhood onset conduct disorder. *Development and Psychopathology* 2004;16:179–192. [PubMed: 15115070]
- Moffitt, TE.; Caspi, A.; Rutter, M.; Silva, PA. *Sex differences in antisocial behavior: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge, UK: Cambridge University Press; 2001.

- Moffitt, TE. Life-course persistent and adolescence-limited antisocial behavior: A 10-year research review and a research agenda. In: Lahey, B.; Moffitt, TE.; Caspi, A., editors. *The causes of conduct disorder and serious juvenile delinquency*. New York: Guilford; 2003. p. 49-75.
- Moffitt TE, Arseneault L, Jaffee SR, Kim-Cohen J, Koenen KC, Odgers CL, Slutske WS, Viding E. Research review: DSM-V conduct disorder: Research needs for an evidence base. *Journal of Child Psychology and Psychiatry* 2008;49:3–33. [PubMed: 18181878]
- Pardini D, Obradovic J, Loeber R. Interpersonal callousness, hyperactivity/impulsivity, inattention, and conduct problems as precursors to delinquency persistence in boys: A comparison of three grade-based cohorts. *Journal of Clinical Child and Adolescent Psychology* 2006;35:46–59. [PubMed: 16390302]
- Rudolph KD, Hammen C, Burge D. Cognitive representations of self, family, and peers in school-age children: Links with social competence and sociometric status. *Child Development* 1995;66:1385–1402. [PubMed: 7555222]
- Setterberg, S.; Bird, H.; Gould, M. Parent and interviewer version of the Children's Global Assessment Scale. New York: Columbia University; 1992.
- Schrum CL, Salekin RT. Psychopathy in adolescent female offenders: An Item response theory analysis of the Psychopathy Checklist: Youth Version. *Behavioral Sciences and the Law* 2006;24:39–63. [PubMed: 16491478]
- Silverthorn P, Frick PJ. Developmental pathways to antisocial behavior: The delayed-onset pathway in girls. *Development and Psychopathology* 1999;11:101–126. [PubMed: 10208358]
- Wakefield JC, Schmitz MF, Baer JC. Does the DSM-IV Clinical Significance Criterion for Major Depression Reduce False Positives? Evidence From the National Comorbidity Survey Replication. *American Journal of Psychiatry*. (in press).
- Xie H, Cairns RB, Cairns BD. The development of social aggression and physical aggression: A narrative analysis of interpersonal conflicts. *Aggressive Behavior* 2002;28:341–355.
- Zalot AA, Jones DJ, Forehand R, Brody G. Self-regulation and conduct problems among low-income African American youth from single-mother homes: The roles of perceived neighborhood context and child gender. *Journal of Black Psychology* 2007;33:239–259.
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–130. [PubMed: 3719049]
- Zoccolillo M. Gender and the development of conduct disorder. *Development and Psychopathology* 1993;5:65–78.
- Zoccolillo M, Tremblay R, Vitaro F. DSM-III-R and DSM-III criteria for conduct disorder in preadolescent girls: Specific but insensitive. *Journal of the American Academy of Child and Adolescent Psychiatry* 1996;35:461–470. [PubMed: 8919708]

Figure 1.
Cumulative age of onset of DSM-IV CD among girls (n = 560)

Figure 2.
Impairment by level of CD symptoms

Table 1

Number of participants contributing data to each age of assessment, and unweighted and weighted prevalence of CD

Age	7	8	9	10	11	12	13	14	15
Total interviewed	1783	2370	2326	2303	2272	2250	1696	1105	546
Targeted N	1829	2451	2451	2451	2451	2451	1863	1233	622
Retention rate	97.5	96.7	94.9	94.0	92.7	91.8	91.0	89.6	87.8
DSM-IV CD	7	8	9	10	11	12	13	14	15
Unweighted prevalence (%)	7.8	8.0	7.5	5.7	5.4	6.2	7.6	8.8	9.4
Weighted prevalence (%)	7.1	7.3	6.9	5.0	4.9	5.5	6.9	7.9	8.9

Table 2

Likelihood of impairment at different symptom thresholds

Parent Rated Impairment	OR	95% CI	p value
3 or more symptoms versus 0	6.2	5.0–7.7	<0.001
3 or more symptoms versus 1	3.1	2.4–4.0	<0.001
3 or more symptoms versus 2	2.1	1.6–2.8	<0.001
Teacher Rated Impairment	OR	95% CI	p value
3 or more symptoms versus 0	3.4	2.7–4.2	<0.001
3 or more symptoms versus 1	2.0	1.5–2.5	<0.001
3 or more symptoms versus 2	1.7	1.3–2.2	<0.001
Youth Rated Impairment	OR	95% CI	p value
3 or more symptoms versus 0	2.1	1.8–2.5	<0.001
3 or more symptoms versus 1	1.5	1.2–1.8	<0.001
3 or more symptoms versus 2	1.3	1.0–1.6	0.018

Table 3

Incremental utility of callousness to impairment ratings by parent, teacher, and youth

	Concurrent			Predictive		
	OR	95% CI	p value	OR	95% CI	p value
Parent-rated Impairment						
Impairment T-1	5.8	4.6–7.2	<0.001	5.6	4.5–6.9	<0.001
DSM-IV CD	4.8	3.9–5.8	<0.001	3.2	2.5–3.9	<0.001
Callousness	1.5	1.2–1.8	<0.001	1.5	1.3–1.8	<0.001
Teacher-rated Impairment						
Impairment T-1	3.2	2.7–3.9	<0.001	3.2	2.7–3.9	<0.001
DSM-IV CD	2.2	1.7–2.8	<0.001	2.3	1.9–3.0	<0.001
Callousness	4.2	3.6–5.0	<0.001	1.9	1.6–2.3	<0.001
Youth-rated Impairment						
Impairment T-1	4.2	3.7–4.8	<0.001	4.2	3.7–4.7	<0.001
DSM-IV CD	1.8	1.5–2.2	<0.001	1.7	1.4–2.0	<0.001
Callousness	1.3	1.2–1.6	<0.001	1.2	1.0–1.3	0.030

Table 4

Incremental utility of relational aggression to impairment ratings by parent, teacher, and youth

	Concurrent			Predictive		
	OR	95% CI	p value	OR	95% CI	p value
Parent Rated Impairment						
Impairment T-1	5.5	4.4–6.9	<0.001	5.3	4.2–6.6	<0.001
DSM-IV CD	4.2	3.5–5.2	<0.001	2.9	2.4–3.6	<0.001
Relational Aggression	2.1	1.7–2.5	<0.001	1.7	1.2–2.1	<0.001
Teacher Rated Impairment						
Impairment T-1	2.8	2.3–3.4	<0.001	3.2	2.6–3.8	<0.001
DSM-IVCD	3.1	2.1–4.4	<0.001	2.3	1.8–2.9	<0.001
Relational Aggression	5.4	4.6–6.4	<0.001	1.9	1.6–2.3	<0.001
DSM-IV CD x Relational Aggression	0.4	0.3–0.7	0.001			
Youth Rated Impairment						
Impairment T-1	4.2	3.7–4.8	<0.001	4.2	3.7–4.7	<0.001
DSM-IV CD	1.8	1.5–2.2	<0.001	1.7	1.4–2.0	<0.001
Relational Aggression	1.2	1.0–1.3	0.048	1.1	0.9–1.2	0.300