

Beyond Symptom Counts for Diagnosing Oppositional Defiant Disorder and Conduct Disorder?

¹Oliver Lindhiem, ²Charles B. Bennett, ¹Alison E. Hipwell, and ¹Dustin A. Pardini

¹University of Pittsburgh

²University of Pittsburgh Medical Center

Corresponding Author:

Oliver Lindhiem, Ph.D.

Assistant Professor

University of Pittsburgh, School of Medicine

Department of Psychiatry

3811 O'Hara St.

Pittsburgh, PA 15213

Office: 537 Bellefield Towers

Phone: 412-246-5909

Email: lindhiemoj@upmc.edu

Acknowledgements: This study was supported by a grant from the National Institute of Mental Health (NIMH) to the first author (MH 093508). The NCS dataset was funded by the National Institute of Mental Health (Grants R01 MH/DA46376 and R01 MH49098), the National Institute of Drug Abuse (through a supplement to R01 MH/DA46376), and the W. T. Grant Foundation (Grant 90135190).

Abstract

Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD) are among the most commonly diagnosed childhood behavioral health disorders. Although there is substantial evidence of heterogeneity of symptom presentations, *DSM* diagnoses of CD and ODD are formally diagnosed on the basis of symptom counts without regard to individual symptom patterns. We used unidimensional item response theory (IRT) two-parameter logistic (2PL) models to examine item parameters for the individual symptoms of CD and ODD using data on 6,491 adolescents (ages 13-17) from the National Comorbidity Study: Adolescent Supplement (NCS-A). For each disorder, the symptoms differed in terms of severity and discrimination parameters. As a result, some adolescents who were above *DSM* diagnostic thresholds for disruptive behavior disorders exhibited lower levels of the underlying construct than others below the thresholds, based on their unique symptom profile. In terms of incremental benefit, our results suggested an advantage of latent trait scores for CD but not ODD.

Keywords: Conduct Disorder, Oppositional Defiant Disorder, item response theory, assessment

Beyond Symptom Counts for Diagnosing Oppositional Defiant Disorder and Conduct Disorder?

Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD) are among the most common childhood behavioral health concerns (Costello, Mustillo, Erkanli, Keeler, & Angold, 2003; Merikangas et al., 2010; Shivram et al, 2009). Both disorders increase in prevalence across childhood (Ford, Goodman, & Meltzer, 2003; Merikangas et al, 2010), have high levels of functional impairment (Breslau, Miller, Chung, & Schweitzer, 2011; Loeber, Burke, & Pardini, 2009), and elevated rates of concurrent and consecutive comorbidity with other psychiatric disorders (Barker, Oliver, & Maughan, 2010; Beauchaine, Hinshaw, & Pang, 2010; Fergusson, Horwood, Ridder, 2007; Loeber, Burke, Lahey, Winters, & Zera, 2000). Not surprisingly, individuals with CD and ODD also have especially high rates of involvement with mental health services (Farris, Nicholson, Borkowski, & Whitman, 2011; Merikangas et al., 2010).

Construct Validity and Heterogeneity of Disruptive Behavior Disorders

Although symptoms associated with ODD and CD frequently co-occur, considerable theoretical and empirical evidence has supported the uniqueness of the two disorders. The diagnosis of ODD refers to a persistent pattern of negativistic, hostile, defiant, and disobedient behaviors toward others, while CD is characterized by a persistent pattern of behavior that involves significant violations of the rights of others and/or major societal norms. Confirmatory factor analytic studies using multiple informants (e.g., parents, teachers, youth, interviewer ratings) and assessment methods (e.g., self-administered rating scales, structured interviews) have generally found that the symptoms associated with ODD and CD seem to be tapping distinct, yet highly related constructs in children and adolescents, with some studies reporting minor symptom overlap between the two disorders (Bezdjian, et al, 2011; Fergusson, Horwood, & Lynskey, 1994; Frick et al., 1993; Lahey et al., 2008). Temporally, evidence suggests that

symptoms of ODD tend to predict changes in CD symptoms over time (Lahey, McBurnett, & Loeber, 2000; Loeber et al., 2000), but the reverse does not tend to be true (Burke, Loeber, Lahey, & Rathouz, 2005; Kolko & Pardini, 2010; Pardini & Fite, 2010). Further evidence for the distinction between the disorders comes from longitudinal studies indicating that ODD symptoms tend to predict the emergence of internalizing problems (Burke, Hipwell, & Loeber, 2010; Copeland, Shanahan, Costello, & Angold, 2009; Fergusson, Boden, & Horwood, 2010; Pardini & Fite, 2010; Rowe, Costello, Angold, Copeland, & Maughan, 2010), whereas CD symptoms are more robustly associated with the development of substance use disorders, persistent criminal behavior, and features of antisocial and psychopathic personality (Burke, Loeber, & Lahey, 2007; Byrd, Loeber, & Pardini, 2011; Copeland et al., 2009; Fergusson et al., 2010; McMahon, Witkiewitz, Kotler, & Conduct Problems Prevention Research Group, 2010; Pardini, White, & Stouthamer-Loeber, 2007; Pardini & Fite, 2010).

Evidence suggests that even within the diagnostic categories of ODD and CD there remains significant heterogeneity in symptomatology. Several studies have now demonstrated that ODD symptoms associated with negative affectivity (e.g., angry and resentful) can be distinguished from more headstrong (e.g., argues with adults) and vindictive (e.g., spiteful) behaviors (Krieger et al., 2013; Rowe et al., 2010; Stringaris, & Goodman, 2009b; Whelan, Stringaris, Maughan, & Barker, 2013). In addition, the negative affectivity component of ODD seems to account for the association with internalizing problems in youth (Burke et al., 2010; Whelan, et al., 2013), whereas the headstrong and vindictive symptoms seem to be more robustly associated with the development of CD (Kolko & Pardini, 2010; Krieger, et al., 2013; Stringaris, & Goodman, 2009a). Studies have similarly noted that CD symptoms can be further subdivided into overt (e.g., aggression, destruction of property) and covert/rule breaking (e.g., stealing,

runaway) behaviors (Bezdjian et al., 2011; Frick et al., 1993; Tackett, Krueger, Sawyer, & Graetz, 2003). Longitudinal studies examining the relative predictive utility of these two dimensions have provided somewhat mixed results, with some evidence indicating that covert CD symptoms are more strongly related to later antisocial personality disorder (APD) (Lahey, Loeber, Burke, & Applegate, 2005) and others reporting that overt CD symptoms are more robustly related to later APD (Le Corff & Toupin, 2013).

In light of the heterogeneity in ODD and CD symptoms, some studies have attempted to delineate subtypes of youth exhibiting each disorder based on individual symptom profiles. Nock and colleagues (2006) identified five subtypes of youth exhibiting CD based on symptom endorsement: rule violations, deceit/theft, aggressive, severe covert, and pervasive. The rule violations, deceit/theft, and aggressive subtypes show a unique symptom set per diagnosis. The severe covert and pervasive subtypes have similarities in symptoms, but differed in symptom severity and count. Similarly, Lacourse and colleagues (2010) identified subtypes of CD: non-aggressive (involving acts such as property offenses), physically aggressive (involving acts of physical harm and violence), and severe/mixed (involving patients experiencing a greater number of symptoms). Similar studies have attempted to identify subgroups of children based on ODD symptom profiles, with one recent investigation finding evidence for three subgroups in clinical referred youth: high behavioral and negative affective symptoms, high behavioral symptoms only, and low symptom levels (Burke, 2012). Findings from this study indicated that youth in the combined behavioral/affective group were more likely to meet criteria for depression and exhibit high levels of neuroticism in adulthood relative to youth in the other two groups.

Symptom Counts Versus Symptom Profiles

Although delineating symptom profiles within disorders may prove useful, current diagnostic criteria in the Diagnostic and Statistical Manual of Mental Disorders (*DSM-IV*, APA, 2000 and *DSM-5*, APA, 2013) for CD and ODD are based on symptom counts. In order to meet diagnostic criteria for CD, an individual must present with three or more of 15 symptoms. Each symptom contributes equally to the diagnostic threshold, regardless of the particular combination of symptoms. Out of a possible 32,768 (2^{15}) symptom patterns, 32,647 can result in a diagnosis of CD. It is even possible for several individuals to meet criteria for CD without having a single symptom in common. In order to meet diagnostic criteria for ODD, an individual must present with four or more of eight symptoms. Out of a possible 256 (2^8) symptom patterns, 163 will result in a diagnosis of ODD. The possibility of non-overlapping symptom profiles within CD and ODD (defined as ‘polythetic’, Needham, 1975) and the consequent heterogeneity among individuals diagnosed with CD and ODD, raises questions about the utility of symptom counts as the basis of firm diagnoses given that different symptoms may also differ in regard to etiology, severity, progression and response to treatment (Krueger & Bezdjian, 2009).

One of the benefits of item response theory (IRT) is to efficiently reduce response patterns to a latent trait score (theta). It should be noted that some patterns of symptoms are very unlikely to occur while others may result in similar theta scores. In our own work applying latent trait IRT models to DSM diagnoses, however, we have found that two children with the same number of symptoms might have different levels of the underlying construct in terms of a latent trait score (Lindhiem, Kolko, & Yu, 2013). For example, we found that a child with the symptoms, “often argues with adults,” “often deliberately annoys people,” and “often blames others for his or her mistakes,” has a lower theta score than a child with the symptoms, “often loses temper,” “often actively defies or refuses to comply with adults’ requests or rules,” and “is

often angry or resentful.” We have even found that children with more symptoms ODD can sometimes have lower levels of the underlying construct than those with fewer symptoms, depending on their individual symptom pattern. As a result, it would seem that diagnostic criteria should consider the symptom profile (*which* symptoms) and not just the symptom count (how many symptoms). A greater focus on symptom profiles may also have utility for the identification of youth most at risk for exhibiting a persistent pattern of problem behavior.

The examination of symptom profiles may also have utility for making finer grained distinctions between levels of the underlying construct than simply counting the number of symptoms present. Along these lines, a longitudinal study by Lahey and colleagues (1995) found that a significant number of clinic-referred boys who were diagnosed with CD in one year did not meet criteria the next year. More than half of the boys who did not meet diagnostic criteria at the second assessment still had two symptoms of CD, which is just below the diagnostic threshold. Furthermore, few of the boys were symptom free, and many were diagnosed again in a later year.

Current Study

We seek to replicate our findings that individuals with more symptoms sometimes have lower theta scores than those with fewer symptoms, depending on the symptom pattern, when latent trait severity scores (theta scores) are estimated using IRT (Lindhiem et al., 2013). In our prior study with a clinical sample of children (age 5 to 12), we found that some children with more ODD symptoms had lower theta scores than those with fewer symptoms, depending on which symptoms were endorsed. In this study we seek to extend this prior finding in a large nationally representative sample of adolescents (age 13 to 17). We also seek to examine whether similar findings hold for CD. We expect that a latent trait approach might be even more relevant

to CD than ODD given the greater range of severity of CD symptoms (from lying to fire-setting). Specifically, we hypothesize that some adolescents above the diagnostic threshold will have lower theta scores than other adolescents below the diagnostic threshold, when latent trait severity scores (theta scores) are estimated using IRT. We expect this will be true for both ODD and CD. These analyses will build on our ongoing effort to improve the accuracy of diagnoses for childhood disorders.

Method

Sample

We conducted secondary analyses using data from the National Comorbidity Study: Adolescent Supplement (NCS-A; see Kessler et al., 2009; Kessler et al., 2012 for full details) which we obtained from the Inter-university Consortium for Political and Social Research (ICPSR). The goal of the NCS-A was to produce nationally representative data on the prevalences and correlates of mental disorders among adolescents ages 13-17 between February 5, 2001 and January 31, 2004. The data collected closely matched the 2000 US census data on key sociodemographic variables including age, sex, race, and ethnicity (Kessler et al., 2004). The overall sample was 51.1% female and 48.9% male (Kessler et al., 2012). The participants were 55.7% non-hispanic white, 19.3% non-hispanic black, 18.9% hispanic, and 6.1% other. The total dataset included 10,148 adolescents with data from 8,485 parents. In the current study, we analyzed the data from a subset of 6,491 parents who completed the full set of questionnaires. The NCS-A sample included 586 children who met diagnostic criteria for CD (lifetime) and 1,047 children who met diagnostic criteria for ODD (lifetime).

Procedure

Adolescents were recruited through both household and school sub-samples. The

household sub-sample included adolescents from families who were contacted during the National Comorbidity Survey – Replication (NCS-R; see Kessler et al., 2004 for full details). To control for the socio-demographic and geographic distributions (according to the 2000 Census) and probability of selection, the household sub-sample was weighted (further details have been described by Kessler et al., 2009). The school sub-sample was identified by initially contacting 289 schools. Of these schools, 81 agreed to participate. The most common reason for refusal was concerns about releasing private student information. Every school that refused was replaced with similar schools in order to avoid potential biases. However, because multiple replacement schools were contacted for each refusal, the survey included students from 320 schools. Within each school, 40-50 students were randomly selected as participants. The NCS-A data was collected by 197 staff members from the Survey Research Center at the University of Michigan and were supervised by 18 experienced supervisors. The staff conducted face-to-face interviews with adolescent participants to evaluate the prevalence of mental health disorders using the *DSM-IV*. During the face-to-face interviews, parents of participating adolescents were asked to complete self-administered questionnaires.

Measures

Thirty day, 12 month, and lifetime prevalence rates for adolescent mood, anxiety, substance, and behavior disorders were measured with both parent and adolescent report during the NCS-A. Adolescents were interviewed using the Composite International Diagnostic Interview Version 3.0 (CIDI). During the interviews, parents completed a paper questionnaire that assessed disorders with established parent report reliability (i.e., attention-deficit/hyperactivity disorder, conduct disorder, oppositional defiant disorder, major depressive episode, and dysthymic disorder; see Merikangas, Avenevoli, Costello, Koretz, & Kessler, 2009

for full measure details). In the current study we analyzed the parent-reported lifetime prevalence of CD and ODD symptoms.

Data Analyses

Assessment of unidimensionality. We conducted exploratory factor analyses (EFAs) separately for the ODD items and CD items to assess unidimensionality of each, a prerequisite for IRT analyses. Our criterion for unidimensionality was a ratio of first to second eigenvalues of ≥ 3.0 for an unrotated factor solution (e.g., Hawes et al., 2014; Morizert, Ainsworth, & Reise, 2007).

Two-parameter logistic (2PL) IRT models. Primary IRT analyses were conducted using IRTPRO (Cai, du Toit, & Thissen, 2011). In separate analyses, the symptoms of CD and ODD were fit to two-parameter logistic (2PL) IRT models for dichotomous items. 2PL models, first described by Birnbaum (1968), are consistent with DSM assumptions (unidimensional constructs with dichotomous symptoms) and therefore these models are most widely applied to DSM disorders (e.g. Cole et al., 2011, Gelhorn et al., 2009, Lindhiem et al., 2013). They have severity parameters (β s) and discrimination parameters (α s). The severity parameter β is defined as the latent trait level at which a respondent has a 50% probability of endorsing the item (in this case a symptom). The discrimination parameter α (the slope of the item characteristic curve [ICC] at β) measures how well the item (symptom) discriminates between those with theta levels above and below β . Higher α s indicated better discrimination. For each model, we estimated threshold parameters (β s) and discrimination parameters (α s) for each of the DSM symptoms. We also estimated latent trait levels (θ) for each of the symptom patterns that were represented. Scoring was based on the expected a posteriori (EAP) estimation method (Bock & Mislevy, 1982).

Incremental validity of latent trait scores. To test the incremental validity of latent trait scores compared to symptom counts, we examined partial correlations with impairment data. Our impairment variable was a composite of three items measuring impairment across home, school, and peer settings. Specifically, we tested whether theta scores were significantly associated with clinical impairment, controlling for symptom counts.

Results

Descriptive and Preliminary Analyses

Conduct Disorder. Internal consistency for the fifteen CD items was adequate, Cronbach's alpha = .78. Total symptom counts ranged from 0 to 15 ($M = 0.44$; $SD = 1.23$). Item frequencies (proportion endorsed) for the sample were as follow: "lies" = .15, "steals" = .05, "breaks in" = .01, "set fire" = .01, "damages" = .02, "stays out" = .02, "truant" = .02, "run away" = .02, "bullies" = .04, "fights" = .03, "weapon" = .01, "harm animal" = .01, "harm person" = .03, "mugs" = .01, and "forced sex" = .00. The exploratory factor analysis (EFA) of the fifteen CD items resulted in an unrotated factor solution with a first to second order eigenvalue ratio of 3.02, indicating adequate unidimensionality to conduct IRT analyses.

Oppositional Defiant Disorder. Internal consistency for the eight items was good, Cronbach's alpha = .93. Total symptom counts ranged from 0 to 8 ($M = 1.58$; $SD = 2.57$). Item frequencies (proportion endorsed) for the sample were as follow: "argues" = .26, "temper" = .19, "defies" = .21, "annoys" = .17, "blame" = .19, "touchy" = .24, "angry" = .24, and "spiteful" = .09. The exploratory factor analysis (EFA) of the eight ODD items resulted in an unrotated factor solution with a first to second order eigenvalue ratio of 8.04, indicating adequate unidimensionality to conduct IRT analyses.

IRT Results

Conduct Disorder. Table 1 summarizes the discrimination parameters (α s) and severity parameters (β s) for each of the 15 symptoms of CD. All items had reasonable discrimination parameters that ranged from 2.11 (“steals”) to 3.96 (“bullies”). The item “lies” had the lowest severity parameter ($\beta = 1.35$). At this latent trait level (1.35 *SDs* above the sample mean) there is a 50% chance that this symptom would be present per parent report. The item “forced sex” had the highest severity parameter ($\beta = 3.24$). At this latent trait level (3.24 *SDs* above the sample mean) there is a 50% chance that this symptom would be present per parent report. There were 285 unique symptom patterns represented in the sample. These are plotted in Figure 1 with theta scores on the x-axis and symptom counts on the y-axis. From the figure, it can be seen that there is overlap in severity (latent trait score) between some adolescents with just two symptoms (below the DSM diagnostic threshold) and those with three symptoms (above the DSM diagnostic threshold). Similarly, many adolescents had the same symptom count but different levels of θ due to their different symptom patterns and the different discrimination and severity parameters associated with their symptoms. Consistent with our hypothesis, adolescents with three symptoms may have lower theta scores than adolescents with two symptoms, depending on which symptoms are present.

Oppositional Defiant Disorder. Table 2 summarizes the discrimination parameters (α s) and severity parameters (β s) for each of the eight symptoms of ODD. All items had good discrimination parameters that ranged from 3.96 (“spiteful”) to 6.25 (“argues”). The item “argues” had the lowest severity parameter ($\beta = 0.71$). At this latent trait level (0.71 *SDs* above the sample mean) there is a 50% chance that this symptom would be present per parent report. The item “spiteful” had the highest severity parameter ($\beta = 1.48$). At this latent trait level (1.48 *SDs* above the sample mean) there is a 50% chance that this symptom would be present per

parent report. There were 189 unique symptom patterns represented in the sample. These are plotted in Figure 2 with theta scores on the x-axis and symptom counts on the y-axis. As with CD, there was overlap in latent trait scores between some adolescents below the DSM diagnostic threshold (three symptoms) and those above the DSM diagnostic threshold (four symptoms). Again, many adolescents had the same symptom count but different levels of theta due to their different symptom patterns and the different discrimination and severity parameters associated with their symptoms. Consistent with our hypothesis, adolescents with four symptoms may have lower theta scores than adolescents with three symptoms, depending on which symptoms are present.

Incremental Validity of Latent Trait Scores

Theta scores for the CD items predicted clinical impairment above beyond symptom counts, partial $r = .184$, $p < .001$. Bivariate correlations were .415 (clinical impairment and theta scores) versus .378 (clinical impairment and symptom counts). Theta scores, however, did not perform any better for ODD items than symptom counts. Theta scores for the ODD items did not predict clinical impairment above beyond symptom counts, partial $r = -.033$, $p = .272$.

Discussion

Consistent with our hypotheses, latent trait models of ODD symptoms and CD symptoms resulted in several cases in which adolescents above the DSM diagnostic thresholds had lower theta scores than those below the DSM diagnostic thresholds. In terms of incremental validity, our results evidenced incremental benefit of latent trait scores above and beyond symptom counts for CD but not ODD. These results support the view that diagnostic criteria for CD should take into consideration not just the symptom count, but also the symptom profile associated with a given disorder. For example, destruction of property and fire-setting are particularly severe

symptoms of CD and should perhaps be weighted more heavily than lying, which is a common symptom even among adolescents without clinically significant conduct problems. The advantage of symptom profiles was less clear for ODD. One implication of these results may be that diagnoses of disruptive behavior disorders should be made cautiously, perhaps only assigning a provisional diagnosis for mild cases (minimum number of symptoms for a diagnosis) at initial assessments, particularly for CD.

Symptom Profiles and the Posterior Probability of Diagnosis (PPOD) Index

Given the results of this study and others with similar results, we suggest the possibility that diagnostic criteria could take into consideration not just the symptom count, but also the symptom profile of an individual patient, particularly for CD. The Posterior Probability of Diagnosis (PPOD) Index has recently been proposed as a way to quantify the likelihood that a patient meets or exceeds a latent trait diagnostic threshold, based on the patient's symptom pattern rather than symptom count (Lindhiem et al., 2013; Lindhiem, Yu, Grasso, Kolko, & Youngstrom, in press). Patients with the same symptom count may have different PPOD Index values depending on their symptom profile. Although the PPOD Index certainly needs more clinical studies before being integrated into routine clinical care, it is a promising model of a dimensional approach to diagnostics based on symptom profiles. At the very least, it provides a way for clinicians to quantify the degree of confidence in a diagnosis and is consistent with the tenets of evidence-based assessment (e.g., Hunsley & Mash, 2005; Jensen-Doss & Weisz, 2008). Although the PPOD Index does not eliminate diagnostic uncertainty, it quantifies the uncertainty and can therefore be clinically useful. For example, a clinician could ask all patients with a 30% or higher likelihood of the disorder to follow-up in 3 months.

Provisional Diagnoses

Another practical solution that does not involve the complex statistics required to estimate the PPOD Index is to assign provisional diagnoses for borderline cases or even mild cases (minimum number of symptoms for a diagnosis). A clinician may choose to wait and see “which way the needle moves” before assigning a diagnosis that will become a permanent part of patient’s permanent medical history. This may be especially prudent for initial assessments in light of Lahey and colleagues’ (1995) study showing boys with CD tend to fluctuate above and below DSM diagnostic criteria from year to year. Guidance on the use of provisional diagnoses has not changed from the DSM-IV (APA, 2000) to DSM-5 (APA, 2013). “The clinician can indicate the diagnostic uncertainty by recording “(provisional)” following the diagnosis” (APA, 2013, p.23).

Limitations

In the current study, we relied on parent-reported symptoms of ODD and CD. Although individual item-parameters would almost certainly differ for adolescent-reported symptoms or combined data, the meta-result that individual symptoms differ in terms of severity and discrimination parameters would likely be unchanged. In the current study, parent-report data was simply chosen to provide a clean illustration of the application of IRT to DSM diagnoses. The approach, however, could easily be expanded to handle both parent and adolescent reports. A typical approach would be for a symptom to be considered endorsed if *either* a parent *or* the adolescent endorsed the symptom. We also chose not to analyze the data separately for males and females. Even though males have considerably higher rates of ODD and CD than females, we analyzed the data together for two reasons. First, current diagnostic criteria for ODD and CD are the same for males and females. Second, studies to date show little evidence of differential item functioning (DIF) for males and females for most symptoms (e.g. Gelhorn et al., 2009).

Finally, it should be noted that diagnoses are made on the basis of both symptoms and impairment. Specifically, a diagnosis is not made without clinically significant impairment regardless of symptoms. In practice, however, this could be applied to symptom profiles no differently than to symptom counts.

Future Studies

It will be important for the results of this study to be replicated with a different dataset to examine the stability of the results. Future studies might also extend the current study by applying multidimensional IRT models to explore the hypothesized subtypes of CD and ODD further. Finally, this study could be extended by the inclusion of additional constructs such as age-of-onset and the stability of symptoms, and by examining the positive predictive power (PPP) and negative predictive power (NPP) of theta scores relative to symptom counts.

Summary and Conclusions

Based on latent trait (IRT) models, some adolescents above DSM diagnostic thresholds for disruptive behavior disorders may actually be exhibiting less severe (in terms of a latent trait) manifestations of ODD and CD than others below the thresholds. The clinical utility of symptom profiles, however, is perhaps more important for CD than ODD. Based on our results and review of the literature, we recommend that clinicians use caution in assigning diagnoses for borderline and mild cases of CD and ODD. Specifically, we recommend efforts to quantify the level of uncertainty associated with diagnoses or the use of provisional diagnoses for mild cases.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders*, Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*, Fifth Edition. Washington, DC: American Psychiatric Association.
- Arbuckle, J. L. (2012). IBM® SPSS® Amos™ 21. Chicago: Amos Development Corporation.
- Barker, E. D., Oliver, B. R., & Maughan, B. (2010). Co-occurring problems of early onset persistent, childhood limited, and adolescent onset conduct problem youth. *Journal of Child Psychology and Psychiatry*, 51(11), 1217-1226.
- Beauchaine, T. P., Hinshaw, S. P., & Pang, K. L. (2010). Comorbidity of attention-deficit/hyperactivity disorder and early-onset conduct disorder: Biological, environmental, and developmental mechanisms. *Clinical Psychology: Science and Practice*, 17(4), 327-336.
- Bezdjian, S., Krueger, R. F., Derringer, J., Malone, S., McGue, M., & Iacono, W. G. (2011). The structure of DSM-IV ADHD, ODD, and CD criteria in adolescent boys: A hierarchical approach. *Psychiatry Research*, 188(3), 411-421.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord, & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-479). Reading, MA: Addison-Wesley.
- Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.
- doi:10.1177/014662168200600405

- Breslau, J., Miller, E., Chung, W., & Schweitzer, J. B. (2011). Childhood and adolescent onset psychiatric disorders, substance use, and failure to graduate high school on time. *Journal of Psychiatric Research, 45*(3), 295-301.
- Burke, J. D. (2012). An affective dimension within oppositional defiant disorder symptoms among boys: Personality and psychopathology outcomes into early adulthood. *Journal of Child Psychology and Psychiatry, 53*(11), 1176-1183. doi:10.1111/j.1469-7610.2012.02598.x
- Burke, J. D., Hipwell, A. E., & Loeber, R. (2010). Dimensions of oppositional defiant disorder as predictors of depression and conduct disorder in preadolescent girls. *Journal of the American Academy of Child and Adolescent Psychiatry, 49*(5), 484-492.
- Burke, J. D., Loeber, R., & Lahey, B. B. (2007). Adolescent conduct disorder and interpersonal callousness as predictors of psychopathy in young adults. *Journal of Clinical Child and Adolescent Psychology, 36*(3), 334-346.
- Burke, J. D., Loeber, R., Lahey, B. B., & Rathouz, P. J. (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. *Journal of Child Psychology and Psychiatry, 46*(11), 1200-1210.
- Byrd, A. L., Loeber, R., & Pardini, D. A. (2012). Understanding desisting and persisting forms of delinquency: The unique contributions of disruptive behavior disorders and interpersonal callousness. *Journal of Child Psychology and Psychiatry, 53*(4), 371-380. doi:10.1111/j.1469-7610.2011.02504.x
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling*. Chicago, IL: Scientific Software International.

- Cole, D. A., Cai, L., Martin, N. C., Findling, R. L., Youngstrom, E. A., Garber, J., . . . Forehand, R. (2011). Structure and measurement of depression in youths: Applying item response theory to clinical data. *Psychological Assessment, 23*(4), 819-833. doi:10.1037/a0023518
- Copeland, W. E., Shanahan, L., Costello, E. J., & Angold, A. (2009). Childhood and adolescent psychiatric disorders as predictors of young adult disorders. *Archives of General Psychiatry, 66*(7), 764.
- Costello, E., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and Development of Psychiatric Disorders in Childhood and Adolescence. *Archives of General Psychiatry, 60*(8), 837-844.
- Farris, J. R., Nicholson, J. S., Borkowski, J. G., & Whitman, T. L. (2011). Onset and progression of disruptive behavior problems among community boys and girls: A prospective longitudinal analysis. *Journal of Emotional and Behavioral Disorders, 19*(4), 233-246.
- Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2010). Classification of behavior disorders in adolescence: Scaling methods, predictive validity, and gender differences. *Journal of Abnormal Psychology, 119*(4), 699-712.
- Fergusson, D. M., Horwood, L., & Ridder, E. M. (2007). Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: Results of a 25-year longitudinal study. *Drug and Alcohol Dependence, 88*, s14-s26.
- Fergusson, D. M., Horwood, L., & Lynskey, M. T. (1994). Structure of DSM-III-R criteria for disruptive childhood behaviors: Confirmatory factor models. *Journal of the American Academy of Child & Adolescent Psychiatry, 33*(8), 1145-1157.

- Ford, T., Goodman, R., & Meltzer, H. (2003). The British Child and Adolescent Mental Health Survey 1999: The prevalence of DSM-IV disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*(10), 1203-1211.
- Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S., & Dane, H. E. (2003). Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology, 31*(4), 457-470.
- Frick, P. J., Lahey, B. B., Loeber, R., Tannenbaum, L., Van Horn, Y., Christ, M. A. G., . . . Hanson, K. (1993). Oppositional defiant disorder and conduct disorder: A meta-analytic review of factor analyses and cross-validation in a clinic sample. *Clinical Psychology Review, 13*(4), 319-340.
- Gelhorn, H., Hartman, C., Sakai, J., Mikulich-Gilbertson, S., Stallings, M., Young, S., . . . Crowley, T. (2009). An item response theory analysis of DSM-IV conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(1), 42-50. doi:10.1097/CHI.0b013e31818b1c4e
- Hawes, S. W., Byrd, A. L., Henderson, C. E., Gazda, R. L., Burke, J. D., Loeber, R., & Pardini, D. A. (2014). Refining the parent-reported Inventory of Callous–Unemotional Traits in boys with conduct problems. *Psychological Assessment, 26*(1), 256-266. doi:10.1037/a0034718
- Hunsley, J., & Mash, E. J. (2005). Introduction to the special section on developing guidelines for the evidence-based assessment (EBA) of adult disorders. *Psychological Assessment, 17*, 251-255.

Jensen-Doss, A., & Weisz, J. R. (2008). Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *Journal of Consulting and Clinical Psychology, 76*, 711-722.

Kessler, R. C., Avenevoli, S., Costello, E. J., Georgiades, K., Green, J. G., Gruber, M. J., . . . Merikangas, K. R. (2012). Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the national comorbidity survey replication adolescent supplement. *Archives of General Psychiatry, 69*(4), 372-380. doi:10.1001/archgenpsychiatry.2011.160

Kessler, R. C., Avenevoli, S., Costello, E. J., Green, J. G., Gruber, M. J., Heeringa, S., . . . Zaslavsky, A. M. (2009). National comorbidity survey replication adolescent supplement (NCS-A): II. overview and design. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 380-385. doi:10.1097/CHI.0b013e3181999705

Kessler, R. C., Berglund, P., Chiu, W. T., Demler, O., Heeringa, S., Hiripi, E., . . . Zheng, H. (2004). The US national comorbidity survey replication (NCS-R): Design and field procedures. *International Journal of Methods in Psychiatric Research, 13*(2), 69-92. doi:10.1002/mpr.167

Kolko, D. J., & Pardini, D. A. (2010). ODD dimensions, ADHD, and callous-unemotional traits as predictors of treatment response in children with disruptive behavior disorders. *Journal of Abnormal Psychology, 119*(4), 713-725.

Krieger, F. V., Polanczyk, G. V., Goodman, R., Rohde, L. A., Graeff-Martins, A., Salum, G., . . . Stringaris, A. (2013). Dimensions of oppositionality in a Brazilian community sample: Testing the DSM-5 proposal and etiological links. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*(4), 389-400. doi:10.1016/j.jaac.2013.01.004

Krueger, R. F., & Bezdjian, S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: Toward DSM-V and ICD-11. *World Psychiatry, 8*(1), 3-6.

Lacourse, E., Baillargeon, R., Dupéré, V., Vitaro, F., Romano, E., & Tremblay, R. (2010). Two-year predictive validity of conduct disorder subtypes in early adolescence: A latent class analysis of a Canadian longitudinal sample. *Journal of Child Psychology and Psychiatry, 51*(12), 1386-1394. doi:10.1111/j.1469-7610.2010.02291.x

Lahey, B. B., Loeber, R., Burke, J. D., & Applegate, B. (2005). Predicting future antisocial personality disorder in males from a clinical assessment in childhood. *Journal of Consulting and Clinical Psychology, 73*(3), 389-399.

Lahey, B. B., Loeber, R., Hart, E. L., Frick, P. J., Applegate, B., Zhang, Q., . . . Russo, M. F. (1995). Four-year longitudinal study of conduct disorder in boys: Patterns and predictors of persistence. *Journal of Abnormal Psychology, 104*(1), 83-93. doi:10.1037/0021-843X.104.1.83

Lahey, B. B., McBurnett, K., & Loeber, R. (2000). Are attention-deficit/hyperactivity disorder and oppositional defiant disorder developmental precursors to conduct disorder? Sameroff, Arnold J [Ed], 431-446.

Lahey, B. B., Rathouz, P. J., Van Hulle, C., Urbano, R. C., Krueger, R. F., Applegate, B., . . . Waldman, I. D. (2008). Testing structural models of DSM-IV symptoms of common forms of child and adolescent psychopathology. *Journal of Abnormal Child Psychology, 36*(2), 187-206. doi:10.1007/s10802-007-9169-5

Le Corff, Y., & Toupin, J. (2013). Overt versus covert conduct disorder symptoms and the prospective prediction of antisocial personality disorder. *Journal of Personality Disorders, 1*-9.

- Lindhiem, O., Kolko, D. J., & Yu, L. (2013). Quantifying diagnostic uncertainty using item response theory: The posterior probability of diagnosis index. *Psychological Assessment*, 25, 456-466. doi:10.1037/a0031392
- Lindhiem, O., Yu, L., Grasso, D. J., Kolko, D. J., & Youngstrom, E. A. (in press). Adapting the Posterior Probability of Diagnosis (PPOD) Index to enhance evidence-based screening: An application to ADHD in primary care. *Assessment*. doi: 10.1177/1073191114540748
- Loeber, R., Burke, J., & Pardini, D. A. (2009). Perspectives on oppositional defiant disorder, conduct disorder, and psychopathic features. *Journal of Child Psychology and Psychiatry*, 50(1-2), 133-142.
- Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: A review of the past 10 years, Part I. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(12), 1468-1484.
- Loeber, R., Pardini, D. A., Hipwell, A., Stouthamer-Loeber, M., Keenan, K., & Sembover, M. A. (2009). Are there stable factors in preadolescent girls' externalizing behaviors? *Journal of Abnormal Child Psychology*, 37(6), 777-791. doi:10.1007/s10802-009-9320-6
- Merikangas, K. R., He, J.-P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010). Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. *Pediatrics*, 125(1), 75-81.
- Merikangas, K. R., Avenevoli, S., Costello, E. J., Koretz, D., & Kessler, R. C. (2009). National comorbidity survey replication adolescent supplement (NCS-A): I. background and measures. *Journal of the American Academy of Child & Adolescent Psychiatry*, 48(4), 367-379. doi:10.1097/CHI.0b013e31819996f1

- Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., . . . Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(10), 980-989.
- McMahon, R. J., Witkiewitz, K., Kotler, J. S., & Conduct Problems Prevention Research Group. (2010). Predictive validity of callous-unemotional traits measured in early adolescence with respect to multiple antisocial outcomes. *Journal of Abnormal Psychology, 119*(4), 752-763.
- Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. Robins, R. C., Fraley, & R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 407-421). New York, NY: Guildford Press.
- Needham, R. (1975). Polythetic classification: Convergence and consequences. *Man, 10*(3), 349-369. doi:10.2307/2799807
- Nock, M. K., Kazdin, A. E., Hiripi, E., & Kessler, R. C. (2006). Prevalence, subtypes, and correlates of DSM-IV conduct disorder in the national comorbidity survey replication. *Psychological Medicine, 36*(5), 699-710. doi:10.1017/S0033291706007082
- Pardini, D., White, H. R., & Stouthamer-Loeber, M. (2007). Early adolescent psychopathology as a predictor of alcohol use disorders by young adulthood. *Drug and Alcohol Dependence, 88*, S38-S49.
- Pardini, D. A., & Fite, P. J. (2010). Symptoms of conduct disorder, oppositional defiant disorder, attention-deficit/hyperactivity disorder, and callous-unemotional traits as unique predictors of psychosocial maladjustment in boys: Advancing an evidence base for DSM-

- V. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49(11), 1134-1144.
- Rowe, R., Costello, E. J., Angold, A., Copeland, W. E., & Maughan, B. (2010). Developmental pathways in oppositional defiant disorder and conduct disorder. *Journal of Abnormal Psychology*, 119(4), 726-738.
- Shivram, R., Bankart, J., Meltzer, H., Ford, T., Vostanis, P., & Goodman, R. (2009). Service utilization by children with conduct disorders: Findings from the 2004 Great Britain child mental health survey. *European Child & Adolescent Psychiatry*, 18(9), 555-563.
- Stringaris, A., & Goodman, R. (2009a). Longitudinal outcome of youth oppositionality: Irritable, headstrong, and hurtful behaviors have distinctive predictions. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48(4), 404-412.
- Stringaris, A., & Goodman, R. (2009b). Three dimensions of oppositionality in youth. *Journal of Child Psychology and Psychiatry*, 50(3), 216-223.
- Tackett, J. L., Krueger, R. F., Sawyer, M. G., & Graetz, B. W. (2003). Subfactors of DSM-IV conduct disorder: Evidence and connections with syndromes from the Child Behavior Checklist. *Journal of Abnormal Child Psychology*, 31(6), 647-654.
- Whelan, Y. M., Stringaris, A., Maughan, B., & Barker, E. D. (2013). Developmental continuity of oppositional defiant disorder subdimensions at ages 8, 10, and 13 Years and their distinct psychiatric outcomes at age 16 years. *Journal of the American Academy of Child & Adolescent Psychiatry*, 52(9), 961-969.

Table 1

Item Parameters for the 15 Symptoms of CD (2PL Model) in Order of Severity (β)

Item	Item Parameters		Standard Errors	
	α	β	σ_{α}	σ_{β}
Lies	2.20	1.35	0.11	0.04
Bullies	3.96	1.95	0.29	0.04
Fights	3.94	1.98	0.30	0.05
Harm person	3.09	2.10	0.22	0.05
Steals	2.11	2.19	0.13	0.07
Damages	3.13	2.25	0.24	0.06
Run away	2.64	2.40	0.20	0.08
Stays out	2.62	2.47	0.20	0.08
Breaks in	3.16	2.47	0.28	0.08
Truant	2.59	2.53	0.21	0.09
Mugs	3.78	2.57	0.41	0.09
Weapon	3.49	2.62	0.39	0.09
Harm animal	2.56	2.70	0.23	0.11
Set fire	3.05	2.84	0.35	0.13
Forced sex	2.92	3.24	0.45	0.20

Table 2

Item Parameters for the Eight Symptoms of ODD (2PL Model) in Order of Severity (β)

Item	Item Parameters		Standard Errors	
	α	β	σ_{α}	σ_{β}
Argues	6.25	0.71	0.29	0.02
Touchy	6.03	0.76	0.28	0.02
Angry	6.10	0.77	0.28	0.02
Defies	5.03	0.88	0.22	0.02
Temper	5.26	0.95	0.24	0.02
Blames	4.77	0.98	0.21	0.02
Annoys	4.44	1.04	0.19	0.02
Spiteful	3.96	1.48	0.22	0.03

Figure 1. Theta scores (x-axis) plotted against symptoms counts (y-axis) for each CD symptom pattern represented in the dataset.

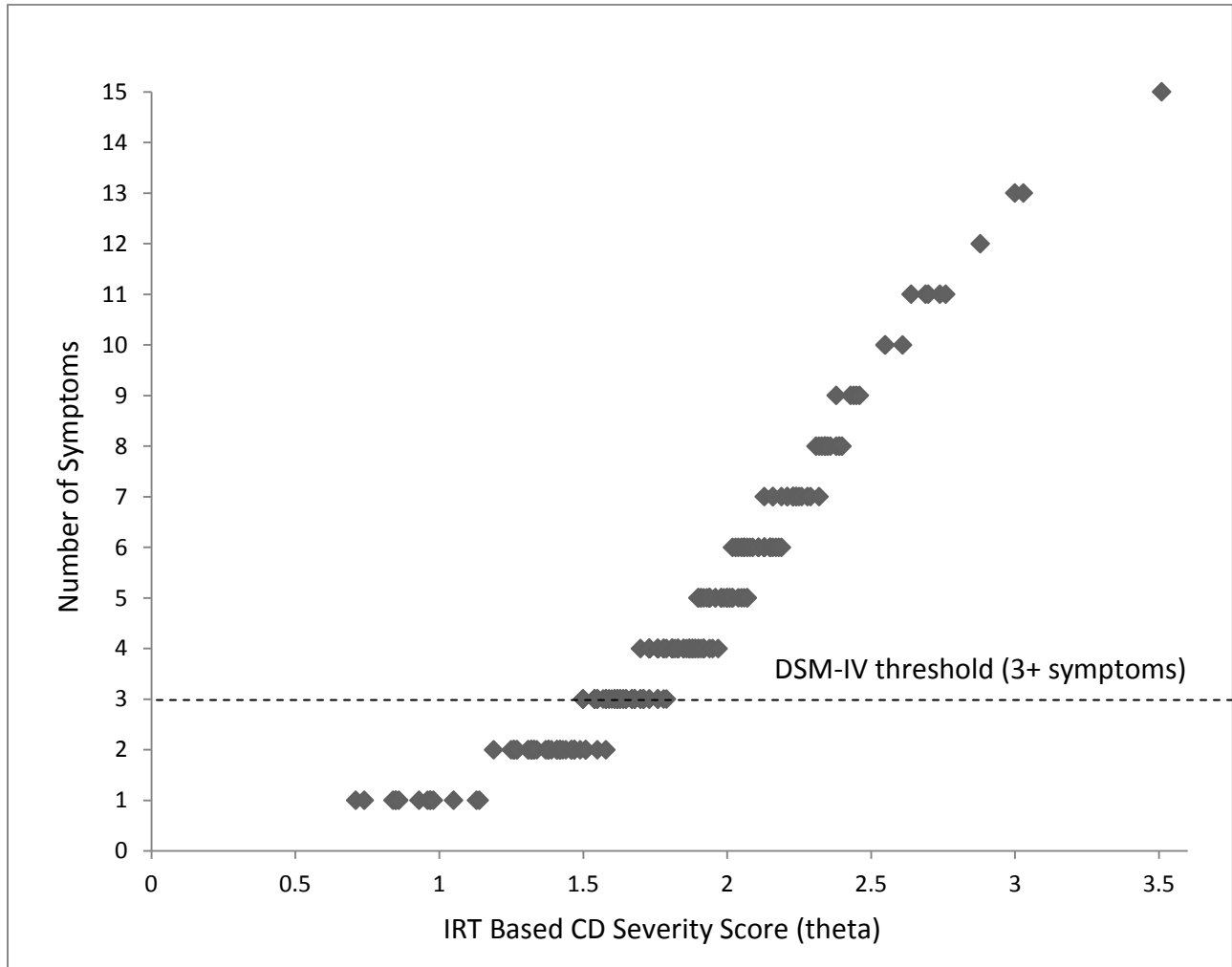


Figure 2. Theta scores (x-axis) plotted against symptoms counts (y-axis) for each ODD symptom pattern represented in the dataset.

