HHS Public Access

Author manuscript

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2020 January 01.

Published in final edited form as:

J Clin Child Adolesc Psychol. 2019 ; 48(1): 93–107. doi:10.1080/15374416.2017.1280806.

Age 4 predictors of Oppositional Defiant Disorder in early grammar school

John V. Lavigne,
Department of Child and Adolescent Psychiatry, Ann & Robert H. Lurie Children's Hospital of Chicago, Feinberg School of Medicine, Northwestern University, and Mary Ann and J. Milburn Smith Child Health Research Program, Children's Hospital of Chicago Research Center, (jlavigne@luriechildrens.org)

Fred B. Bryant,
Department of Psychology, Loyola University Chicago, (fbryant@luc.edu)

Joyce Hopkins, and
College of Psychology, Illinois Institute of Technology, (hopkins@iit.edu)

Karen R. Gouze
Department of Child and Adolescent Psychiatry, Ann & Robert H. Lurie Children's Hospital of Chicago, and Feinberg School of Medicine, Northwestern University, (kgouze@luriechildrens.org)

Abstract

Objective.—Our ability to predict which children will exhibit Oppositional Defiant Disorder (ODD) at the time of entry into grammar school at age 6lags behind our understanding of the risk factors for ODD.This study examined how well a set of multi-domain risk factors for ODD assessed in 4-year-old childrenpredictedage 6 ODD diagnostic status.

Method.—Participantswere a diverse sample of 7964-year-old children (391 males).The sample was 54% White, non-Hispanic; 16.8%, African American, 20.4%, Hispanic; 2.4%, Asian; 4.4%, Other or mixed race.), The classification accuracy of two models of multi-domain risk factors,using either a measure of overall ODD symptoms or dimensions of ODD obtainedat age 4,were compared to one another, to chance, and to a parsimonious model based solely on parent-reported ODD using Automated Classification Tree Analysis (CTA).

Results.—Effect Strength for Sensitivity(ESS),a measure of classification accuracy, indicated a multi-domain model including a general measure of ODD symptoms at age 4 yielded a large effect (56.29%), a 13.7% increase over the ESS for the parsimonious model (ESS = 42.9%). The ESS (51.23%)for a model includingtwo ODD dimensions (behavior and negative affect) was smaller than that for the model including a measure of overall ODD symptoms.

Conclusions.—The CTA approach showeda small but distinct advantage that would be useful in screening for which childrenwould most likely meet criteria for age 6 ODD.

Correspondence should be addressed to John V. Lavigne, Department of Child and Adolescent Psychiatry (#10), Ann & Robert H. Lurie Children's Hospital of Chicago, 225 East Chicago Avenue, Chicago, IL 60611. jlavigne@luriechildrens.org.

Keywords

Oppositional Defiant Disorder; classification accuracy; Optimal Data Analysis; Classification Tree Analysis

Symptoms of difficult-to-manage, oppositional behavior are common in preschoolers(Egger & Angold, 2006). Studies suggest that these symptoms are reasonably stable from preschool to the early school years (Campbell, 1990; Lavigne et al., 1998). Nonetheless, a significant number of oppositional preschoolers may not exhibit these symptoms by early grammar school (Lavigne et al., 1998). Campbell(1995) estimatesthat approximately half of difficult-to-manage childrenwill continue to have high levels of symptoms of disruptive behavior 3–7 years later. Lavigne et al. (1998) report that children with early indications of a disruptive disorder are 47% - 50% more likely to have such a disorder in early grammar school. Overall, this pattern of results suggests that, while about 50% of children with the disorder in preschool continue to exhibit high levels of symptoms ofOppositional Defiant Disorder (ODD) in early grammar school, for others,ODD symptoms decline over time.

Oppositional behavior in early grammar school can be highly problematic. Studies identify: (a) a life-coursepattern of homotypic continuity in which a subset of youngoppositional children later develop conduct disorder later (Moffitt, 1993); (b) a pattern of heterotypic continuity in which oppositional preschoolers later develop anxiety or depression(Burke & Loeber, 2010; Drabick, Ollendick, & Bubier, 2010; Lavigne et al., 2001); and (c) preschool children with ODD exhibit problems with emotional regulation (Lavigne, Gouze, Hopkins, Bryant, & LeBailly, 2012) linked to problems with academic achievement (McClelland et al., 2007; Ponitz, McClelland, Matthews, & Morrison, 2009).

Because ODD in early childhood may have deleterious long-term effects, early intervention could be critical, particularly because empirically-supported interventions for ODD are available(Barlow, Bergman, Konner, Wei, & Bennett, 2016).Because ODD prevalenceis high in young children—approximately 8–10% when impairment is included in the diagnosis(Lavigne, LeBailly, Hopkins, Gouze, & Binns, 2009), mustering the treatment resources to address the need is potentially problematic.As a result, from a population health perspective, finding ways to identify young children most in need of treatment may be important. Such concerns have prompted pediatricians to recommend screening for behavioral problems in primary care (Foy, Kelleher, & Laraque, 2010). Since symptoms of ODD remit in some children and not others before grammar schoolentry, determining which children are most likely to have persistent problems with ODD, or those whose subthreshold ODD symptoms might become more problematic, is animportantissue.

Our understanding of the factors associated with the onset, stability, and course of ODD is increasing, and the knowledge gained might help identifychildren needing a targeted intervention to reduce the likelihood of an ODD diagnosis in early grammar school.Thelarge number of risk factors associated with ODD symptomshave been described elsewhere(Lavigne et al., 2012). In general, the association of these risk factors with ODD has been examined in small-scale studies of individual or small sets of risk factors at a time.Few studies have integrated those risk factors into an overall modelincluding risk

factorsfrommultipledomainswhile also specifying the pathways through which such risk factors are associated with ODD symptom levels.One such model has been reported, linking multi-domain factors associated with the stability of ODD syptoms in both cross-sectional (Lavigne et al., 2012)and longitudinal models(Lavigne, Gouze, Hopkins, & Bryant, 2015). Risk factorsin the multi-domain modelinclude: (a) at the contextual level,lower SES(Evans, 2004), parental stress(McMahon, Grant, Compas, Thurm, & Ey, 2003), and family conflict(Grant et al., 2006); (b) at the parent level,parental depression(Goodman et al., 2011); (c) at the parenting level,poor parenting(Hipwell et al., 2008); and (d) at the child level,less secure attachment (DeVito & Hopkins, 2001) andtemperament variables of negative affect, poor effortful control(Eisenberg et al., 2009), and poor sensory regulation(Gouze, Hopkins, LeBailly, & Lavigne, 2009).Finally, the best predictor of meeting criteria for a diagnosis at a future date may be indications that symptoms of that disorder were present previously(Lavigne, Gouze, Hopkins, &Bryant, 2015). In addition, because symptoms of psychopathology often are correlated, and comorbidity among disorders is often high (Angold, Costello, & Erkanli, 1999),symptoms of ODD, or ODD symptoms combined with other types of symptoms, may be strong predictors of who meets criteria for an ODD diagnosis in the future.

As studies of risk factors associated with ODD symptoms have accrued, our understanding of the structure of ODD has changed. It is now recognized that ODD is not a unitary construct; rather, there are multiple dimensions to ODD. Presently, DSM-5 recognizes a three-dimensional structure of ODD, with dimensions of angry/irritable mood, argumentative/defiant behavior, and vindictiveness. A recent study comparing various proposed structures suggestedthat a two-factor structure (ODDbehavior, similar to the DSM-5 argumentative/defiant factor, and negative affect, similar to the DSM-5 dimension of angry/irritable mood) may more appropriately characterize the structure of ODD symptoms in young children(Lavigne, Bryant, Hopkins, & Gouze, 2015). Studies have examined whether specific dimensions are better predictors of subsequent disorders (Lavigne, Gouze, Bryant, & Hopkins, 2014; Stringaris & Goodman, 2009), and specific dimensions may be better predictors of subsequent ODD diagnosis than overall disorder alone.

A major limitation to developing practical applications for the growing information about the risk factors and structure for ODD is that the results of inferential statistical approaches used to establish associations between predictors and outcomes (e.g., correlations, effect sizes, regression coefficients) do not readily translate into a form useful for identifying individual children at risk for disorder. Clinicians often find research results to be of limited clinical utility(Garland, Hurlburt, & Hawley, 2006), but they may find that results identifying a specific percentage of children with a particular set of attributes (i.e., likely to benefit from treatment, develop a disorder, etc.) are more useful than traditional presentations of statistical analyses.An alternative approach is needed that can: (a) identify cutoff scores useful for predicting who will later have a disorder, and (b) examinehow multiple risk factors mightbest be combined to predict future disorder. The purpose of the present study is to determine how well multi-domain risk factors for ODD assessed in,a preschool or primary care pediatric settingcan be used to identify which children are either likely or unlikely to meet criteria for an ODD diagnosis at age 6, when most children are enrolled in, or about to enter, grade 1. The classification accuracy of two models for

predicting which children would meet criteria for an ODD disorder at age 6 were compared to one another, to chance, and to classification accuracy based solely on the presence of ODD symptoms at age 4. These two classification models included multi-domain risk factors and either(a) a measure of overall ODD symptoms or (b) measures of dimensions of ODD.

Method

Participants

This report is part of a longitudinal study of factors related to symptoms of ODD in young children(Lavigne, Gouze, et al., 2015). Children and their parents were followed from ages 4 to 6 years. For this report, predictor measures were completed when the child was age 4 years, and the diagnosis of ODD was based on assessment at age 6 years, when most children were in grade 1. Children (N = 796) and their parents were recruited from 13 Chicago Public School preschool programs and 23 primary care pediatric clinics in Cook County, Illinois. All children included in the study were: (a) age 4 at enrollment; (b) English- or Spanish-speaking; (c)living with the parent(s) for the previous 6 months or longer; (d) obtained a standard score on the Peabody Picture Vocabulary Test ⩾ 70(Dunn & Dunn, 1997); and (e) did not meet criteria for Autism Spectrum Disorder.

During recruitment, 1,738 families expressed interest in the study and 827 (47.5%) agreed to participate. Of those families,31were ineligible, resulting in a final sample of 796 children (183 were recruited through the schools and 613 from pediatric practices). The mean age was 4.44 years, and 391 (49.1%) were boys. Members of all social classes (Hollingshead, 1975) participated, with 303 (38.1%) children in the highest class (Class I), 290 (36.4%) in Class II, 79 (9.9%) in Class III, 63 (7.9%) in Class IV, and 61 (7.7%) in Class V. We sought a racially/ethnically diverse sample similar to the diversity of Cook County, Illinois, wherethe sample was recruited. Parent-reported racial/ethnicity for the sample was: 433 (54.4%) White, non-Hispanic; 162 (20.4%) Hispanic; 133 (16.7%) African American; 19 (2.4%) Asian; and 35 (4.4%) self-described as multi-racial or other. Race/ethnicity was not reported by 14 (1.8%) parents. For 31 families (3.9%), a primary caretaker father was the participating parent.

A total of 626 children and families (78.6%) participated in all waves of data collection. Families completing the first and third waves differed from those who did not with respect to: (a) being lower SES, $\chi^2(4, N = (796) = 69.61, p = .001$; (b) race, with a greater proportion of dropouts among minority participants, $\chi^2(5, N(796) = 77.7, p = .001$; and (c) younger children, 25 days older at study entry, $t(773) = 2.41, p = .02$. Because listwise deletion is more likely to introduce biases than multiple imputation procedures (Graham, 2009), missing data were multiply imputed (see data analysis section, below). The final sample N was 796.

Measures

A multi-informant approach to measurementincluded observer ratings for risk factors of child attachment and parent scaffolding, parent reports and interviews to assess outcomes

and specific risk factors, and performance-based measures of inhibitory control and child verbal skills. A parent-reported diagnosis based on a structured interview at age 6 was used because: (a) parentreport typically forms the basis of clinical evaluations, so a structured parent interview was clinically relevant; (b) there are no structured child-reported interviews for six-year-olds; (c) significant evidence has emerged indicating that teacher reports are substantially different from parent reports and that parent and teacher symptom reports cannot serve as proxies for one another(Drabick, Gadow, & Loney, 2007; Lavigne, Dahl, Gouze, LeBailly, & Hopkins, 2015); (d) there are no suitable teacher interview or other observer-based approaches for assigning a DSM diagnosis for young children.

Because the primary study (Lavigne,Gouze, et al., 2015) was designed to examine mediating processes, multiple measures were obtained to estimate latent factors. In the present report, individual scalesuseful in clinical settings were used to measure study variables. There was one exception: a composite measure of family conflict based on multiple scales was used because the internal consistency for that composite was better than that for any individual scale. Internal reliability coefficients reported were obtained at age 4. Additional information on reliability and validity are available in an on-line supplement.

Contextual measures

SES and demographics.—Parent-reported demographic information was obtained on the child's race, age, sex, parent's education, and parent's employment status. SES was coded using the Hollingshead Four-Factor Index of Social Status (Hollingshead, 1975).

Family conflict.—The composite measure of parent-reported family conflict (coefficient alpha =.71) included: (a) Family Environment Scale (FES) conflict scale(Moos & Moos, 1986). Sample items include, "Family members criticize each other often;" "Fight a lot in family;"(b) McCubbin Family Distress Index(H. I. McCubbin, Thompson, & Elver, 1996). Sample items include," Increase in conflict between adults in the house;" "Increased disagreement about a member's friends or activities;" (c) Family Problem Solving/ Communication scale(M. A. McCubbin, McCubbin, & Thompson, 1996). Sample items include, "We talk things through till we reach a solution;" "We make matters more difficult by fighting and bringing up old matters."

Life Stress.—The Perceived Stress Scale(Cohen, Kamarck, & Mermelstein, 1983) is a 14-item measure in which respondents describe the degree to which their lives appear be uncontrollable, and overwhelming. Sample items include,"In the last month, how often have you been upset because of something that happened unexpectedly?;" "In the last month, how often have you been upset because of something that happened unexpectedly?" Coefficient alpha was .87.

Parental depression.—The Beck Depression Inventory (BDI) (Beck, Ward, Mendelson, Mock, & Erbaugh, 1961) is a widely-used, 21-item self-report measure of symptoms of adult depression. Sample items include, "Suicidal thoughts or wishes;" "Self-dislike;" "Guilty feelings." In this study, coefficient alpha was .86.

Parenting measures

Parent support and hostility.—The Parent Behavior Inventory (PBI) (Lovejoy, Weis, O'Hare, & Rubin, 1999) is a parent-report measure of parenting behavior with preschool and school-age children. The PBI has two factor-analytically-derived subscales, Support/ Engagement and Hostility/Coercion. Sample items for support/engagement include "Pleasant conversations with child," and "Listen and try to understand child's feelings." Sample items for hostility/coercion are, "Lose temper when child doesn't do something requested;" "Demand child does something right away." Coefficient alpha was .85 for support and .73 for coercion.

Scaffolding.—A 15-minute videotaped, semi-structured parent-child interaction procedure, the NICHD Three Boxes Paradigm (NICHD Early Childhood Research Network, 1999), was used to assess scaffolding. Scaffolding refers to the skill with which parents assist their child in managing frustration and developing skills needed to succeed on a given task (Baker, Fenning, Crnic, Baker, & Blacher, 2007). The task includes two activities designed to be too difficult for the child to accomplish without assistance, and a free-play activity. Parental scaffolding was a composite score of ratingsby trained research assistants on 7-point qualitative scales for supportive presence, respect for autonomy, quality of assistance, cognitive stimulation, confidence, and hostility. Inter-rater reliability for these scales wasgood to excellent(NICHD Early Childhood Research Network, 1999), and composite ratings of the scales significantly predict children's attachment security and social competencies(NICHD Early Childhood Research Network, 1999).Study coders were trained to a criterion of 80% reliability compared with two master coderstrained by a doctoral-level NICHD study member. A random sample of 20% of the tapes was double-coded to assess inter-rater reliability. Reliabilities (intraclass correlation coefficients) in the present study ranged from .80 for quality of assistance to .69 (likely deflated by a low base rate) for maternal hostility, with a mean reliability of .74. To create a composite measure of parenting behavior, the six parent ratings were factor-analyzed using a maximum likelihood method. A one-factor solution provided the best fit to the data. As a result, a composite measure of scaffolding was calculated by summing the ratings for supportive presence, respect for autonomy, cognitive stimulation, quality of assistance and confidence and subtracting the hostility rating. The mean inter-rater reliability score calculated using intra-class correlation, was .74.

Child factors

Child negative affect (NA) and effortful control (EC).—Rothbart, Ahadi, Hershy, and Fisher's(2001) Children's Behavior Questionnaire (CBQ) is a widely-used, parent-report measure of temperament yielding a measure of NA (alpha = .62) and EC(alpha =.72). Sample items for NA are, "Is very frightened by nightmares," and "tends to feel down at the end of the day." Sample items for EC include, "Can lower his voice when asked to do so," and "Has an easy time leaving play to come to dinner."

Child sensory regulation (SR) ability.—A 38-item parent-report questionnaire, the Short Sensory Profile (SSP; study alpha = .81) yielded a single total score of SR ability(McIntosh, Miller, Shyu, & Hagerman, 1999). Higher scores reflect better SR ability.

Sample items include, “Avoids going barefoot, especially in sand or grass;” and “Becomes anxious or distressed when feel leave the ground.”

Attachment.—The Attachment Q-Sort (AQS) (Waters, 1987) is a continuous measure of attachment security exhibiting good convergent validity with the Strange Situation Paradigm(van IJzendoorn, Vereijken, Bakermans-Kranenburg, & Riksen-Walraven, 2004). A research assistant observed the mother and child during the home visit for 120 minutes and then completed the Q-Sort. A second observer rated a 20% random sample of visits. Interrater reliability was .77.

Receptive language.—The Peabody Picture Vocabulary Test (PPVT), alpha = .94(Dunn & Dunn, 1997), a measure of single-word receptive language, was used as a measure of receptive language skill. Supporting the validity of the PPVT as a measure of language ability,the PPVT correlates .85 with the Stanford-Binet Intelligence test (Ollendick, Finch, & Ginn, 1974).

Child inhibitory control (IC).—The Statue subtest from the Developmental Neuropsychological Assessment (NEPSY) (Korkman, Kirk, & Kemp, 1997), study alpha = . 91, assesses the child’s ability to inhibit prepotent responses to a stimulus. The child is asked to maintain a specific body position, with closed eyes, for 75seconds while an assistant makes sounds to elicit the child’s visual attention. Both the Statue task and the EC measure, in part, assess the child’s ability to inhibit certain responses. The tasks are considered different because the Statue task assesses inhibitory control on “cool” tasks, while the CBQ ECsubscale assesses inhibition of negative emotions (sample CBQ items: able to resist temptations, approaches dangerous places cautiously)(Zelazo & Carlson, 2012). Statue provides a single indicator of IC.

Child psychopathology.—TheChild Symptom Inventory (CSI)(Gadow & Sprafkin, 1997)is a parent-reported,DSM-IV-keyed problem behavior checklist with symptom occurrence rated from “never” (0) to “very often” (3). Parent-reported scale scores at age 4 from the following scales were used: ODD (study alpha, .86); depression (study alpha, .68), generalized anxiety (GAD) (study alpha, .70), separation anxiety (study alpha, .77), ADHD Inattention (study alpha, .88), and ADHD hyperactivity-impulsivity (study alpha, . 88).Typically, scores for each item on a scale are summed together and converted to T-scores. DSM-5, however, considers a symptom to be present if it occurs “often” or “very often.”To increase the generalizability of results, for each item, scores of “never” (0) or “sometimes” (1) were scored as zero (i.e., symptom not present) and scores of “often” (2) or “very often” (3) were scored as 1 (i.e., symptom present).Scores were summed to create the scale total score.

DSM-5(American Psychiatric Association, 2013) recognizes that there are different dimensions of the ODD diagnostic construct.Prior studies have shown that a two- factor structure(Burke, Loeber, Lahey, & Rathouz, 2005) is preferable to the three-factor structure of ODD incorporated into DSM-5 for young children. In the present study, a two-factor structure that was invariant for age and gender in young children (Lavigne, Bryant, et al., 2015) was used in analyses involving ODD dimensions.The two factors are ODD-Behavior

(ODD-B, loading on items for argues, defies, and temper tantrums) and ODD-negative affect (ODD-NA, loading on items touchy, angry, spiteful/vindictive). For the CSI ODD-B dimension, alpha = .82 at age 4; for NA, .76 at age 4.Sample CSI items include: “Loses temper” for ODD-B, and “touchy” for ODD-NA;” “Is sad for most of the day (depression); “Worries more than other children (generalized anxiety);” “Worries about being left at home alone or with a sitter (separation anxiety);” “Has difficulty organizing tasks or activities (ADHD-inattentive);” “Is on the go or acts as if driven by a motor (ADHD-hyperactivity).”

TheDiagnostic Interview Schedule for Children-Parent Scale - Young child version (DISC-YC).—The DISC-YC(Fisher & Lucas, 2006)is a developmentally-appropriate adaptation of the DISC-P, a DSM-IV-based structured parent interview that yields information about the presence or absence of a psychiatric disorder, and the type of diagnosis.The DISC-YC is a “fully structured” interview (Costello, Egger, & Angold, 2005) in which parents answer “yes” or “no” to questions about symptoms, contingent questions assess details about the symptom, and the computer program skips to the next question contingent upon parental response. Unlike semi-structured, interviewer-based procedures, the role of clinical judgment in the DISC-YC is essentially eliminated.As a result, the interview can be conducted by trained research assistants without extensive clinical experience.The research assistants administering the DISC-YC were either clinical psychology graduate students or,in one instance, a B.A.-level graduate applying to graduate school. Trained by DISC-trained staff, the interviewers practiced conducting interviews until they were able to reproduce the skip patterns used by the DISC-trained interviewers. This follows procedures in Shaffer, Fisher, Lucas, Dulcan, and Schwab-Stone (2000).Children meeting criteria for an ODD diagnosis at age 6 exhibited impairment level A, defined by the test developers as showing an intermediate or severe level of impairment on at least one symptom.

Procedure

Parents of 4-year-olds were approached by research assistants in pediatric offices and at drop-offand pickup times at preschools. Subsequently, a home visit was arranged with interested families whowere contacted by telephone, andstudy measures were completed at the home visit. Parents were re-contactedfor another visit when their children were age 6 years and study measures were re-administered. This study was approved by the authors’ Institutional Review Boards; written consent was obtained. Families were paid for participation to compensate for time spent in data collection.

Statistical analysis.—Univariate optimal data analysis (uniODA)(Yarnold & Soltysik, 2005)was used to identify optimal cutoff scores for age 4 predictor variables, using an exact permutation probability requiring no distributional assumptions. Each predictor variable is analyzed individually (hence, “uniODA” analyses) to assess accuracy in predicting the target variable (i.e., presence or absence of ODD diagnosis at age 6 years). After determining the value, statistical significance, and effect strength of theoptimal cut-score on the predictor that maximizes its predictive accuracy, the expected cross-sample stability of the ODA model is assessed. In this process, “jack-knife” leave-one-out (LOO) analyses are conducted to determine the likelihood of replicating the analysis of each predictor in an independent

sample. In the LOO analysis, each observation is removed one at a time from the sample, the uniODA analysis is repeated using the remaining observations, the resulting ODA model is used to classify the single held-out observation, and overall LOO classification accuracy is computed across all of the held-out observations. Results are considered LOO stable for the given predictor variable if the overall classification accuracy of the LOO analysis does not fall below the classification accuracy of the initial uniODA model. To maximize the expected cross-sample generalizability of results of the present study, only statistically significant uniODA analyses that demonstrated stability in the LOO analyses were used.

Subsequently, automated classification tree analysis (CTA) software(Soltysik & Yarnold, 2010)was used to: (a) determine the combination of variables maximizing classification accuracy in predicting age 6 diagnostic status;and (b) apply a sequentially-rejective Bonferroni-type adjustment(Yarnold & Soltysik, 2005)to the *p* valueassociated witheach predictor in the CTA model. In the present study, the minimum samplesize at each endpoint in the CTAmodel was set at 20; predictors not significant at the alpha-adjusted .05 level were eliminated.Using this procedure, all possible combinations of attributes are tested across the first 3 levels of analyses. After determining which variables warranted inclusion, an automatic “pruning” process deconstructs the initial “Bonferroni-pruned” model into all possible nested sub-branches to identify the sub-branch maximizing mean sensitivity, eliminatingsub-branches not improvingeffect strength for sensitivity of the classification tree model to avoid over-fitting (Yarnold & Soltysik, 2010).

To enable comparisons of the effect strength of each predictor in the uniODA and CTA analyses, effect strength for sensitivity (ESS) is reported. ESS summarizes the overall accuracy of the prediction in terms of a chance-corrected estimate of classification accuracy for which 0% represents the expected chance effect, and 100% represents perfect classification accuracy. Thus, an ESS of 50% indicates that the model’s classification accuracy is halfway toward maximum attainable classification accuracy. Yarnold and Soltysik (2005) suggest the following guidelines for interpreting ESS values: < 25%, weak; 25–50%, moderate; 50–75%, relatively strong; 75–90%, strong; > 90%, very strong. ESS is also useful for comparing alternative models. In thisstudy, the ESS for each classification tree also was compared to chance, and to the ESS for a parsimonious model employing only thebest single-variable predictor of age 6 ODDdiagnosis.Findings regarding specific indices of classification accuracy (sensitivity, etc.) for each predictor and classification tree are included in the supplemental, on-line material.

In the initial uniODA analyses conducted for each specific predictor, a large number of *p* values were computed. Corrections to adjust for Type I experiment-wise error rate, however, were not reported for each uniODA model, because theseindividual models were considered to be preliminary analyses designed to determine the relative strength of each variable as an individual predictor. Each predictor that was statistically significant in the uniODA analyses was not necessarily expected to be useful when combined with the other predictors as they would be in the optimal multivariate CTA model. In contrast to the uniODA analyses, the CTA model, whichwas designed todetermine the combination of predictors thatmaximized predictive accuracy, was pruned so that only variables adding to overall prediction at least at .05 were retained.

Results

Percent of children meeting age 6 ODD diagnostic criteria

At age 6, 11.7% (n = 93) of the 796 children met criteria for an ODD diagnosis with some impairment.

Single variable (uniODA) predictors

Table 1 lists the individual age 4 predictors of an ODD diagnosis based on the DISC-YC at age 6, along with ESS values, specific cutoff scores, and directions of effect. Of the 27 individual predictors, 16(59.2%) were both statistically significant and stable in the LOO analyses. Of thesesignificant, stableindividual predictors: (a)7(43.8%), specifically marital status, race, caretaker depression, NA, separation anxiety, conflict, and the ODD NA dimensionhad weak predictive effects(i.e., ESS < 25%); (b)9 (56.2%),including parental stress, child attachment security, SR ability, child EC, and symptoms of overall oppositional behavior,generalized anxiety,ADHD-inattentive type, ADHD-hyperactive impulsive type, and the ODD behavior dimensionhad moderate predictive effects (i.e., ESS = 25–50%).

Demographic variables.—Marital status, specifically having parents who were married had a weak effect predicting ODD diagnosis at age 6. Race had a significant, but weak,predictive effect, with being White or self-identified as “other” for race predicting an age 6 ODD diagnosis. Sex and SES were not significant single variable predictors of age 6 ODD diagnosis.

Contextual variables.—Age 4 conflict and stress were both significant predictors of age 6 ODD diagnosis. The overall strength of the predictive relationship was weak for conflict but moderate for stress. Higher conflict and stress levels were associated with age 6 ODD diagnoses.

Parent and parenting variables.—Parental depression was a weak but significant predictor, with higher levels of age 4 parental depression predicting ODD diagnosis at age 6. The three parenting variables, support, engagement and scaffolding were either (a) not significant, or (b) significant but not LOO stable, predictors of age 6 ODD diagnosis.

Child factors.—Several age 4 child factors were associated with age 6 ODD diagnosis. Lower levels of EC or SR ability at age 4 were moderately strong predictors of an ODD diagnosis at age 6. Secure attachment was alsoa moderately strong predictor, with less securely attached children at age 4 more likely to have an ODD diagnosis at age 6. NA was a weak but significant predictor, with higher levels of NA at age 4 associated with an ODD diagnosis at age 6. Measures of behavioral inhibition (Statue) and receptive vocabulary were not significant predictors.

Child psychopathology.—Higher levels of ODD symptoms, GAD, ADHD-H, ADHD-I,and ODD-Bat age 4 were moderately strong predictors of an ODD diagnosis at age 6. Separation anxiety andODD-NA were weak predictors, while symptoms of child depression were not a significant predictor of an age 6 ODD diagnosis.

Combining potential predictors: CTAs

The next set of analyses examined the classification accuracy of combinations of predictors of age 6 ODD diagnostic status. The first model (Figure 1) included predictors from all fourrisk domains and demographic characteristics, as well as total score on the ODD scale and each measure of psychopathology. The second model (Figure 2) examined the two ODD dimensions rather than overall ODD, along with the other predictors.

The enumerated CTA analysis examines all possible combinations of predictors, including predictors not significant in uniODA,to identify the multivariable model with the best global predictive ability. In interpreting the results, ESS value provides useful information about the predictive ability of the overall CTA model compared to chance.Following customary practice in diagramming CTA models, right branches predict case status, and left branches predict non-case status. Terminal endpoints for right branches are presented as percent values meeting criteria for case status, whereas left branches are presented as percent values meeting criteria for non-case status. Significant cutoff scores are presented for each branch, along with sequentially rejective Bonferroni-adjusted, two-tailed p values.

Model 1: overall ODD scale as a predictor.

In the model with overall ODD scores entered as a predictor (see Figure 1), the variable appearing atthe first node is age 4 total CSI ODD score. The presence of any ODD symptom rated “often” or “very often” (i.e., the right-hand arrow) predictsthe presence of an age 6 ODD diagnosis. Individuals with ⩾1 on the symptom count scale were more likely to be a case, with 32.54% of those with⩾1 symptom being cases at age 6. Among those with no age 4 symptoms, 93.94% were not ODD cases at age 6.

There were three specific paths to case status: (a) for children with no CSI ODD symptoms, who were self-classified as White or “other” race, and who had poorer SRability (SR score ⩽ 171.32), 21.35% met criteria for an ODD diagnosis at age 6; (b) for children with >1 ODD symptoms, 44.00% met criteria for an ODD diagnoses at age 6; (c) for those with 1 CSI ODD symptoms who were ⩽ 4.18 years (approximately 4 years, 2 months) when tested, 34.78% were diagnosable with ODD at age 6.

None of the branches in the total classification tree were verystrong predictors of the presence of ODD at age 6, i.e., no branch correctly classified children with ODD at age 6 at a rate higher than 44%. In contrast, four branches were strong predictors of the absence of ODD diagnoses at age 6: (a) the branch including children with no age 4 ODD symptoms who were Black, Latino, or Asian correctly predicted the absence of ODD diagnoses for 98.77% of children; (b) for children with no age 4 ODD symptoms who were White or of “other” race and who showed no age 4 symptoms of generalized anxiety, 93.89% did not meet age 6 ODD criteria; (c)for children with no age 4 ODD symptoms, of White or “other” race, ⩾ 1 GAD symptom at age 4, and good SR ability, 100% did not meet age 6 criteria for ODD;(d) for children with one age 4 CSI ODD symptom who were > 4.18 months of age when evaluated, 93.48% did not meet criteria for ODD at age 6 years.

The ESS for model 1 overall was56.29%, which is a relatively large effect. Clearly, the model 1 classification tree is an improvement over chance prediction of an age 6 ODD

diagnosis.When compared to a parsimonious model including only the best single-variable predictor, the age 4 ODD symptom scalefor which the ESS was 42.92%, there was a notable 13.37%increase in ESS for the Model 1 classification tree.When other measures of classification accuracy are considered, improvements in prediction for the classification tree approach are also apparent. First, the sensitivity (SE) of the CSI ODD scale alone was 59.14%; for the CTA, SE increased by 17.2%, to 76.34% (see Table 2 for classification accuracy statistics for the CTAs). Generally, when SE improves, specificity (SP) of a measure declines. For these two measures, the relatively large increase in SE for the CTA model was accompanied by a small decrease in specificity (from 83.78% for the CSI ODD scale alone, to 79.94% for the CTA model). This result is an increased ability to identify cases; the CTA approach leads to better identification of true cases at age 6, with a small decline in the ability to detect true noncases, compared to prediction using the overall ODD scale score alone.

There would also be apotential advantage to using the CTA approach over the more parsimonioususe of the CSI ODD scale alone if the CTA approach improved the ability to predict specific groups of individuals who are either likely or unlikely to develop ODD. The overall positive predictive value (the number of true positives correctly predicted by a high score on the ODD scale) of the ODD scale is 32.54%.The ability of the classification tree to predict true cases of ODD is better than 33.01% for two branches, for which the 44.0% and 34.78% were correctly identified, and worse (21.35%)for one other. The negative predictive value (NPV, the number of true negatives correctly identified by a low score on the test) for the ODD scale alone is 93.94%.Because the NPV for the ODD scale is so high, the CTA approach is unlikely to produce amarkedly stronger NPV. Indeed,the NPV for the four branches of the CTA that predicted no disorder were only slightly better for two branches (98.77%, 100.0%) and slightly worse (93.89%, 93.48%) for the others. Thus, the CTA approach provides an overall advantage in classification accuracy compared to the use of the CSI ODD scale alone, and is better for classifying the presence of age 6 ODD for specific subgroups. It shows no significiant advantage, however,in predicting the absence of an ODD diagnosis compared to the CSI ODD scale at age 4 alone.

Model 2:ODD dimensions as a predictor.—In model 2, the two dimensions of ODD were used as predictors rather than the overall CSI ODD scale. The CTA approach resulted in a classification scheme that included the child'sCSI GAD level, SES, ODD-B symptom level,effortful control, and temperamental negative affect. The ODD-NA dimension did not improve classification accuracy in the CTA model.

In the second model, the presence of any (⩾1.0) CSI-GAD symptoms at age 4 predicted an age 6 ODD diagnosis correctly for 20.75% of children. For children with at least 1 GAD symptom at age 4, two branches predicted age 6 ODD: (a) for those with at least one GAD symptom and one ODD-B symptom at age 4, 38.64% met criteria for an ODD diagnosis at age 6; for those with at least 1 GAD symptom at age 4, no ODD-B symptoms at age 4, but high levels of NA, 20.37% met criteria for an ODD diagnosis at age 6. Among children not exhibiting anyGAD symptoms at age 4, an age 6 ODD diagnosis was predicted correctly by being in the highest (Class 1) SES group and exhibiting lower (⩽ 4.95) levels of EC for 32.56% of such children. The absence of an ODD diagnosis was predicted correctly by: (a)

an absence of age 4 GAD symptoms and SES below Class 1 for 97.50% of children in that branch; (b) no GAD symptoms, SES below Class 1, and high levels of EC for 96.13% of such children; (c) at least one age 4 GAD symptom, no age 4 ODDB symptoms, and low NA, for 91.80% of such children.

The overall ESS for this secondCTA model was 51.23% (SE = 75.27%; SP = 75.96%). While this is an improvement compared to either the ESS for ODD-B (35.00%) or ODD-NA (23.63%), and to the ESS of 42.92% for the CSI ODD scale alone, it is smaller than the ESS for model 1 (56.29%).For these reasons, model 1 is a more effective predictive model.

Sex- and race/ethnicity-specific CTAs.—Readers accustomed to seeing demographic differences in standardized scores for tests might prefer to use classification trees that differed across demographic groups.Because the enumeration process used in the automated CTA program tested every possible combination of predictors at each node in the first 3 levels of the classification trees, it is clear that no tree in which a demographic variable was tested in the first node was superior to the classification trees in Figures 1 and 2. Completely separate trees, however, could yield results in which classification accuracy was superior for one particular demographic group compared to a classification tree that was not specific to a single demographic group. As a result, automatedCTAs were conducted for each of the two sexes separately, and for Whites and for combined minority groups separately (sample sizes were too small to allow for analyses of each race/ethnic group). The overall classification accuracy values (ESSs) were then compared across groups.

The classification tree in Figure 1 included the variable of overall ODD symptoms; for that model the ESS was 56.29%.The ESS for the girls-only model that included the overall ODD symptom variable was slightly lower (53.20%) and lower (48.26%) for boys.When separate classification trees were examined for Whites and minority groups, the ESS for Whites of 39.35% was considerably lower than the ESS of 56.29% for the total sample. The only demographic group for which the ESS was better than that for the total sample was that for minority group members, for which the ESS was 63.97%. For that group, the total CSI ODD score was the only significant predictor.Overall, it appears that there is little advantage for separate classification trees for 3 of the 4 demographic groups.

In the second model, the two ODD dimensions were included as predictorsrather than the overall ODD score.For the total sample, the ESS for this second model was 51.23%. For separate demographic groups, there was an improvement in ESS for the minority groups (ESS = 64.71%), a slight improvement for girls (53.20%), but the ESS was lower for boys (47.10%) and considerably lower for Whites (39.55%). When compared to the overall ESS for the total sample of 56.29%,there was improvement when ODD dimensions were predictors only for minority groups.Overall, there appears to be little advantage to using classification trees that differ based on demographic subgroups. See on-line supplementary material for detailed results of the CTA models for the subgroups.

Discussion

There have been considerable advances in our understanding of the structure of ODD symptoms(Ezpeleta, Granero, de la Osa, Penelo, & Domenech, 2012; J V Lavigne, F B Bryant, et al., 2015)and the multi-domain risk factors for this disorder(Lavigne, Gouze, et al., 2015; Lavigne et al., 2012). Nevertheless, the statistical procedures used in such studies do not lend themselves to determining how well we can predict which preschool-age children will meet criteria for this disorder in the future. Because early onset ODD carries risk for future academic and mental health problems, and there are effective interventions for this disorder, it is important tooptimize prediction of which young children are most likely to meet criteria for ODD in order to determine whom to target for early intervention.

The aim of this study, therefore, wastwo-fold: (a) to determine if, at age 4, optimal cut-scores on variables in multiple domainscould be identified that predictODD diagnoses at age 6 in a highly diverse sample; and (b) to determine if classification tree analyses incorporating multiple predictors could improve on the identification of children with and without an ODD diagnosis at age 6. First, we examined single risk factors to determine if cut-scores on specific variables predicted ODD diagnosis at age 6. We then we used automated CTA to determine which combination of variables provided the greatest classification accuracyin predicting ODD status at age 6.

In the single predictor analyses, age 4 risk factors from each domain,except for the parenting domain,yielded statistically significant predictions of an ODD diagnosis at age 6 that had strong, expected cross-sample generalizability. Specifically, being white and having married parents were significant, but weak predictors of an ODD diagnosis at age 6. The contextual factor of conflict (score > 2.66) was a weak predictor, while stress (score > 20.86) was amoderately strong predictor of an ODD diagnosis at age 6. Even relatively low levels of caregiver depression (BDI score > 5.35) had a small effect sizein predicting the likelihood of an ODD diagnosis at age 6. Child characteristics of attachment security, effortful control,SR ability had moderate effect sizesin predicting an ODD diagnosis at age 6, and NA a small effect size.

Given that ODD symptoms are fairly stable from preschool through the early school years(Campbell, 1990; Lavigne et al., 1998), it is not surprising that even relatively low levels (score ⩾1 on a scale of 0 – 8 symptoms) of ODD symptoms at age 4 had the strongest effect on the accuracy of predicting ODD diagnosis at age 6. In addition tohomotypic symptoms predicting the likelihood of an age 6 ODD diagnosis, both generalized and separation anxiety were small, but statistically significant, predictors of an ODDdiagnosis at age 6.

After examining the individual risk factors, we compared two classification tree analyses to determine which combination of variables provided the greatest classification accuracy in predicting age 6 diagnostic status.The two models differed in one important respect: Model 1 used the original age 4 ODD scale score as a predictor, and Model 2 used the two age 4 ODD dimension scores as predictors. The first model yielded an overall classification accuracy of 56.29%, with a specificity of 79.94% and a sensitivity of 76.34%. Model 2

yielded an overall classification rate of 51.23%, with a specificity of 75.96% and a sensitivity of 75.27%. While both models improved predictive accuracy compared to chance, the ESS for Model 1 (total ODD scale) was higher than that for Model 2 (ODD dimensions); for that reason Model 1 is preferable.

When the CTA model 1is compared to the use of the age 4 ODD symptom scale alone, there is an improvement of over 13% in overall classification accuracy and an improvement in SE of over 17% at the expense of a slight, approximately 4%, decrease in SP. This is a notable improvement in classification accuracy for the CTA.

While the CTA approach improved overall classification accuracy, our ability to predict who will meet criteria for an ODD diagnosis at age 6 is still limited, and it is easier to predict who will not meet criteria for that disorder.So few efforts have been made to report on predictive classification accuracy for other disorders in young children that it is difficult to know how the degree of accuracy in predicting ODD compares with that of other disorders. Given the interest in trying to determine if dimensions of ODD differ in predictive value for various disorders, it is of note that analyzing separate dimensions of ODD did not materially improve results compared to using a global measure of ODD. ODD-NA was a statistically significant but weak solo predictor, but was not a good predictor for any of the subgroups included in the classification tree when entered into the multivariable CTA analyses. ODD-B did contribute to the overall predictive accuracy ofModel 2, but that model was less effective overall than the model that included the global CSI ODD scale.

The clinical utility of the use of classification tree approaches and, of this study in particular,will be dependent, in large part, on two factors: the availability of community resources and whether the clinician or clinical administrator is concerned with population health or mental health. In the typical community or tertiary care mental health setting where treatment is provided for children and adolescents, it is unlikely that the considerations addressed in this report would influence decisions about who receives treatment. In such settings, a parent is seeking treatment for a perceived problem, increasing the likelihood that some impairment is present. If the child’s behavior meets criteria for ODD, treatment would be implemented, and this would occur for all children meeting the diagnostic criteria. Concerns about setting priorities for treatment come into play when mental health problems are being addressed at a population level, in schools or pediatric practices.Since screening is recommended in primary care and each pediatric practitioner will see between 1,000 and 1,700 children under age 18 in a primary care practice(Committee on Careers and Opportunities, 1996), and visit frequency is highest among preschoolers, between dozens and hundreds of young children would be screened annually. Given the prevalence of ODD in that age group at 8–10% (with impairment), many children would be eligible for referral or in-setting treatment, potentially more than can readily be treated in busy practices or communities with limited resources.

In such settings, there are two significant methodological problems that enter into deciding whom to treat. The first has to do with limitations in identifying children with behavior problems via screening. The prevalence of ODD at 8–10 percent is high enough to warrant attention as a matter of concern for population health. However, given the classification

accuracy (Lavigne, Meyers, & Feldman, 2016) of even the best available screening instruments, the false positive rate will be very high for disorders with that prevalence (Lavigne, Feldman, & Meyers, 2016).To reduce the impact of this problem, sequential screening has been recommended(Lavigne, Feldman, et al., 2016). Once the preschool children with ODD have been identified, the issues concerning who to prioritize for treatment discussed in the present report become relevant. If resources allow, all children with high levels of ODD symptoms could be treated; there would be no need to be concerned about how likely the ODD symptoms are to be present in two years. Other factors (e.g., GAD and ADHD symptoms) might need to be considered, but some groups of children with a lower likelihood of exhibiting ODD at age 6 (e.g., children with only 1 ODD symptom who were screened later at age 4) might be assigned to a watchful waiting group to be monitored at future visits.For a test or classification tree with high negative predictive value and high specificity, confidence that neither intervention nor watchful waiting is needed increases, and parents can be informed with greater confidence that their child is likely to exhibit ODD at a later date.

This approach could also play a role in evidence-based assessment approaches such as that recommended by Youngstrom, Choukas-Bradley, and Calhoun(2015)thatincludeconsiderations of base rates, psychopathology risk factors, rating scales to supplement and potentially streamline evaluations. Information derived from classification trees can also be important in informing parent consumers who are deciding whether to pursue treatment in view of other family, child, and work considerations in concrete terms about the likelihood of remission so that they can estimate what might be gained from initiating treatment.

Limitations

When considering the CTA results, it is important to keep in mind that this statistical approach is designed to optimize the predictive value of combinations of variables, and not to explicate causal relationships between variables.Although the pruning process for reducing the branches of the tree that do not add significant value to prediction reduces the likelihood of chance findings, this analytic approach cannot articulate why the combinations of factors lead to the results they do, e.g., why there are very different likelihoods of meeting diagnostic criteria for ODD at age 6 for boys and girls who are low in ODD symptoms, high on EC, and of self-described White or Other race. Finally, this study is based on parental reports and results may differ if early childhood predictors included both parent and teacher reports.

Despite these limitations, the present findings provide useful information for determining which children to target for treatment.Future studies are needed to try to improve on the classification accuracy of these variables, and to determine whether these results can be replicated.There is general concern about the limitations of the utility of research findings for clinicians, and an approach such as the one we adoptedhere holds some promise for increasing the clinical utility of findings on risk factors and processes by which disorders may develop. In addition, little is known about the ability to predict other disorders among

young children, and studies of disorders that might be present in early grammar school are needed.

Acknowledgments

FUNDING

This research was supported by National Institute of Mental Health grant MH 063665.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.) Washington, DC: American Psychiatric Association.

Angold A, Costello EJ, & Erkanli A (1999). Comorbidity. Journal of Child Psychology and Psychiatry, 40, 57–87. [PubMed: 10102726]

Baker JK, Fenning RM, Crnic KA, Baker BL, & Blacher J (2007). Prediction of social skills in 6-year-old children with and without developmental delays: contributions of early regulation and maternal scaffoldng. American Journal of Mental Retardation, 112, 375–391. [PubMed: 17676961]

Barlow J, Bergman H, Konner H, Wei Y, & Bennett C (2016). Group-based parent training programmes for improving emotional and behavioral adjustment in young children (review). Cochrane Database of Systematic Reviews, Art. No. CD0003680. doi: 10.1002/14561858.CD0003680

Beck AT, Ward CH, Mendelson M, Mock J, & Erbaugh J (1961). An inventory for measuring depression. Archives of General Psychiatry, 4, 561–571. [PubMed: 13688369]

Burke JD, & Loeber R (2010). Oppositional Defiant Disorder and the explanation of the comorbidity between behavioral disorders and depression. Clinical Psychology: Science and Practice, 17, 319–326.

Burke JD, Loeber R, Lahey BB, & Rathouz PJ (2005). Developmental transitions among affective and behavioral disorders in adolescent boys. Journal of Child Psychology and Psychiatry, 46, 1200–1210. [PubMed: 16238667]

Campbell SB (1990). Behavior problems in preschool children: Clinical and developmental issues. New York: Guilford.

Campbell SB (1995). Behavior problems in preschool children: A review of recent research. Journal of Child Psychology and Psychiatry, 36(9), 113–114. [PubMed: 7714027]

Cohen S, Kamarck T, & Mermelstein R (1983). A global measure of perceived stress. Journal of Health and Social Behavior, 24 385–396. [PubMed: 6668417]

Committee on Careers and Opportunities. (1996). Committee report: Population-to-pediatrician ratio estimates: a subject review. Pediatrics, 97, 597–600. [PubMed: 8632955]

Costello EJ, Egger HE, & Angold A (2005). 10-year research update review: The epidemiology of child and adolescent psychiatric disorders: I. Methods and public health burden. Journal of the American Academy of Child and Adolescent Psychiatry, 44(10), 972–986. [PubMed: 16175102]

DeVito C, & Hopkins J (2001). Attachment, parenting, and marital dissatisfaction as predictors of disruptive behavior in preschoolers Developmental Psychopathology, 13, 215–231.

Drabick DAG, Gadow KD, & Loney J (2007). Source-specific oppositional defiant disorder: Comorbidity and risk factors in referred elementary schoolboys. Journal of the American Academy of Child and Adolescent Psychiatry, 46(1), 92–101. [PubMed: 17195734]

Drabick DAG, Ollendick TH, & Bubier JL (2010). Co-occurrence of ODD and anxiety: Shared risk processes and evidence for a dual-pathway model. Clinical Psychology: Science and Practice, 17, 307–318. doi: 10.1111/j.1468-2850.2010.01222x [PubMed: 21442035]

Dunn L, & Dunn L (1997). The Peabody Picture Vocabulary Test (3rd Edition). Circle Pines, MN: American Guidance Service.

Egger HE, & Angold A (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. Journal of Child Psychology and Psychiatry, 47, 313–337. [PubMed: 16492262]

Eisenberg N, Valiente C, Spinrad TL, Cumberland A, Liew J, Reiser M, Losoya SH (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. Developmental Psychology, 45, 988–1008. doi: 10.1037/a0016213 [PubMed: 19586175]

Evans GW (2004). The environment of childhood poverty. American Psychologist, 59, 77–92. [PubMed: 14992634]

Ezpeleta L, Granero R, de la Osa N, Penelo E, & Domenech JM (2012). Dimensions of oppositional defiant disorder in 3-year-old preschoolers. Journal of Child Psychology and Psychiatry, 53, 1128–1138. doi: 10.1111/j.1469-7610.2012.02545.x [PubMed: 22409287]

Fisher P, & Lucas C (2006). Diagnostic Interview Schedule For Children (DISC-IV)-Young Child. New York: Columbia University.

Foy JM, Kelleher KJ, & Laraque D (2010). Appendix S4: The case for routine mental health screening. Pediatrics, 125, S133–S139. doi: 10.1542/peds.2010-0788J

Gadow KD, & Sprafkin J (1997). Child Symptom Inventory 4, Norms Manual. Stonybrook, NY: Checkmate Plus.

Garland A, Hurlburt M, & Hawley K (2006). Examining psychotherapy processes in a services research context. Clinical Psychology: Science and Practice, 13 30–46.

Goodman SH, Rouse MH, Connell AM, Broth MR, Hall CM, & Heyward D (2011). Maternal depression and child psychopathology: a meta-analytic review. Clinical Child and Family Psychology Review, 14, 1–27. doi: 10.1003/sl10567-010-0080-1 [PubMed: 21052833]

Gouze KR, Hopkins J, LeBailly SA, & Lavigne JV (2009). Re-examining the epidemiology of sensory regulation dysfunction and comorbid psychopathology. Journal of Abnormal Child Psychology, 37, 1077–1087. doi: 10.1007/s10802-009-9333-1 [PubMed: 19543968]

Graham JW (2009). Missing data analysis: Making it work in the real world. Annual Review of Psychology, 60, 549–576. doi: 10.1146/annurev.psych.58.110405.085530

Grant KE, Compas BE, Thurm AE, McMahon SD, Gipson PY, Campbell AJ, Weserholm RI (2006). Stressors and child and adolescent psychopathology: Evidence of moderating and mediating effects. Clinical Psychology Review, 26, 257–283. doi: 10.1016/j.cpr.2005.06.011 [PubMed: 16364522]

Hipwell A, Keenan K, Kasza K, R, L., Stouthamer-Loeber M, & Bean T (2008). Reciprocal influences between girls' conduct problems and depression, and parental punishment and warmth: a six year prospective analysis. Journal of Abnormal Child Psychology, 36, 663–677. doi: 10.1007/s10802-007-9206-4 [PubMed: 18172753]

Hollingshead AB (1975). Four-factor Index of Social Position. Yale University Department of Sociology. New Haven, CT.

Korkman M, Kirk U, & Kemp S (1997). NEPSY: A developmental neuropsychological assessment manual. San Antonio: Psychological Corporation.

Lavigne JV, Arend R, Rosenbaum D, Binns HJ, Christoffel KK, & Gibbons RD (1998). Psychiatric disorders with onset in the preschool years: I. Stability of diagnosis. Journal of the American Academy of Child and Adolescent Psychiatry, 37(12), 1246–1254. [PubMed: 9847496]

Lavigne JV, Bryant FB, Hopkins J, & Gouze KR (2015). Dimensions of Oppositional Defiant Disorder in young children: Model comparisons, gender and longitudinal invariance. Journal of Abnormal Child Psychology, 43, 423–439. doi: 10.1007/s10802-014-9919-0 [PubMed: 25117577]

Lavigne JV, Cicchetti C, Gibbons RD, Binns HJ, Larsen L, & DeVito C (2001). Oppositional Defiant Disorder with onset in preschool years: Longitudinal stability and pathways to other disorders. Journal of the American Academy of Child and Adolescent Psychiatry, 40, 1393–1400. [PubMed: 11765284]

Lavigne JV, Dahl KP, Gouze KR, LeBailly SA, & Hopkins J (2015). Multi-domain predictors of Oppositional Defiant Disorder symptoms in preschool children: Cross-informant differences. Child Psychiatry and Human Development, 46, 308–319. doi: 10.1007/s10578-014-0472-4 [PubMed: 24997089]

Lavigne JV, Feldman M, & Meyers KM (2016). Screening for mental health problems: addressing the base rate fallacy for a sustainable screening program in integrated primary care. Journal of Pediatric Psychology. doi: 10.1093/jpepsy/jsw048

Lavigne JV, Gouze KR, Bryant FB, & Hopkins J (2014). Dimensions of Oppositional Defiant Disorder in Young Children: Heterotypic Continuity with Anxiety and Depression. Journal of Abnormal Child Psychology, 42, 937–951. doi: 10.1007/s10802-014-9853-1 [PubMed: 24497230]

Lavigne JV, Gouze KR, Hopkins J, & Bryant FB (2015). A multi-domain cascade model of early childhood risk factors associated wit ODD symptoms on a community sample of -year-olds. Development and Psychopathology. doi: 10.1037/S0954579415001194

Lavigne JV, Gouze KR, Hopkins J, & Bryant F (2015). A multi-domain cascade model of early childhood risk factors associated with ODD symptoms in a community sample of 6-year-olds. Development and Psychopathology. doi: 10.1017/S9054579415001194

Lavigne JV, Gouze KR, Hopkins J, Bryant FB, & LeBailly SA (2012). A multi-domain model of risk factors for ODD symptoms in a community sample of 4-year-olds. Journal of Abnormal Child Psychology, 40 (5), 741–757. doi: 10.1007/s10802-011-9603-6 [PubMed: 22200893]

Lavigne JV, LeBailly SA, Hopkins J, Gouze KR, & Binns HJ (2009). The prevalence of ADHD, ODD, depression and anxiety in a community sample of 4-year-olds. Journal of Clinical Child and Adolescent Psychology, 38, 315–328. doi: 10.1080/15374410902851382 [PubMed: 19437293]

Lavigne JV, Meyers KM, & Feldman M (2016). Systematic review: classification accuracy of behavioral screening measures for use in integrated primary care settings. Journal of Pediatric Psychology. doi: 10.1093/jpepsy/jsw049

Lovejoy MC, Weis R, O'Hare E, & Rubin EC (1999). Development and initial validation of the Parent Behavior Inventory. Psychological Assessment, 11(4), 1–12. doi: 10.1037/1040-3590.11.4.534

McClelland MM, Cameron CE, Connor CM, Farris CL, Jewkes AM, & Morrison FJ (2007). Links between behavioral regulation and preschoolers' literacy, vocabulary, and math skills. Developmental Psychology, 43(4), 947–959. doi: 10.1037/0012-1649.43.4.947 [PubMed: 17605527]

McCubbin HI, Thompson A, & Elver KM (1996). Family Distress Index In McCubbin HI, Thompson A & McCubbin MA (Eds.), Family assessment: Resiliency, coping, and adaptation-Inventories for research and practice (pp. 783–788). Madison: University of Wisconsin.

McCubbin MA, McCubbin HI, & Thompson A (1996). Family Problem Solving and Communication In McCubbin HI, Thompson A & McCubbin MA (Eds.), Family assessment: Resiliency, coping,and adaptation-Inventories for research and practice (pp. 639–686). Madison: University of Wisconsin.

McIntosh DM, Miller LJ, Shyu V, & Hagerman RJ (1999). Sensory-modulation disruption, electrodermal responses, and functional behaviors. Developmental Medicine and Child Neurology, 41, 608–615. doi: 10.1111/j.1469-8749.1999.tb00664.x [PubMed: 10503919]

McMahon SD, Grant KE, Compas BE, Thurm AE, & Ey S (2003). Stress and psychopathology in children and adolescents: is there evidence of specificity? Journal of Child Psychology and Psychiatry, 44(1), 107–133. doi: 10.1111/1469-7610.00105 [PubMed: 12553415]

Moffitt TE (1993). Adolescent-limited and life-course-persistent antisocial behavior: A developmental taxonomy. Psychological Review, 100, 674–701. [PubMed: 8255953]

Moos RH, & Moos BS (1986). Family Environment Scale Manual. Palo Alto, CA: Consulting Psychologists Press.

NICHD Early Childhood Research Network. (1999). Child care and mother-child interaction in the first three years of life. Developmental Psychology, 35, 1399–1413. [PubMed: 10563730]

Ollendick TH, Finch AJ, & Ginn FW (1974). Comparison of Peabody, Leiter, WISC, and academic achievement scores among emotionally disturbed children. Journal of Abnormal Child Psychology, 2(1), 47–51. [PubMed: 4448879]

Ponitz CC, McClelland MM, Matthews JS, & Morrison FJ (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. Developmental Psychology, 45, 606–619. doi: 10.1037/a0015365

Rothbart MK, Ahadi SA, Hershy KL, & Fisher P (2001). Investigations of temperament at three to seven years: The Children's Behavior Questionnaire. Child Development, 72(5), 1394–1408. [PubMed: 11699677]

Shaffer D, Fisher P, Lucas CP, Dulcan MK, & Schwab-Stone ME (2000). NIMH Diagnostic Interview Schedule for Children Version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses Journal of the American Academy of Child and Adolescent Psychiatry, 39, 28–38. [PubMed: 10638065]

Soltysik RC, & Yarnold PR (2010). Automated CTA software: fundmental concepts and control commands. Optimal Data Analysis, 1, 144–160.

Stringaris A, & Goodman R (2009). Three dimensions of oppositionality in youth. Journal of Child Psychology and Psychiatry, 50, 216–233. [PubMed: 19166573]

van IJzendoorn M, Vereijken C, Bakermans-Kranenburg M, & Riksen-Walraven J (2004). Assessing attachment security with the Attachment Q-sort: Meta-Analytic evidence for the validity of the observer AQS. Child Development, 75, 1188–1213. doi: 10.1111/j.1467-8624.2004.00733.x [PubMed: 15260872]

Waters E (1987). Attachment Q-set (version 3.0). Stony Brook, NY: State University of New York at Stony Brook.

Yarnold PR, & Soltysik RC (2005). Optimal Data Analysis: A guidebook with software for Windows. Washington, DC: American Psychological Association.

Yarnold PR, & Soltysik RC (2010). Maximizing accuracy of classification trees by optimal pruning. Optimal Data Analysis, 1, 23–29.

Youngstrom EA, Choukas-Bradley S, & Calhoun CD (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. Cognitive and Behavioral Practice, 22, 20–35. doi: 10.1016/j.cbpra.2013.12.005

Zelazo PD, & Carlson SM (2012). Hot and cool executive function in childhood and adolescence: development and plasticity. Child Development Perspectives, 6, 354–360. doi: 10.1111/j.1750-8606.2012.00246x

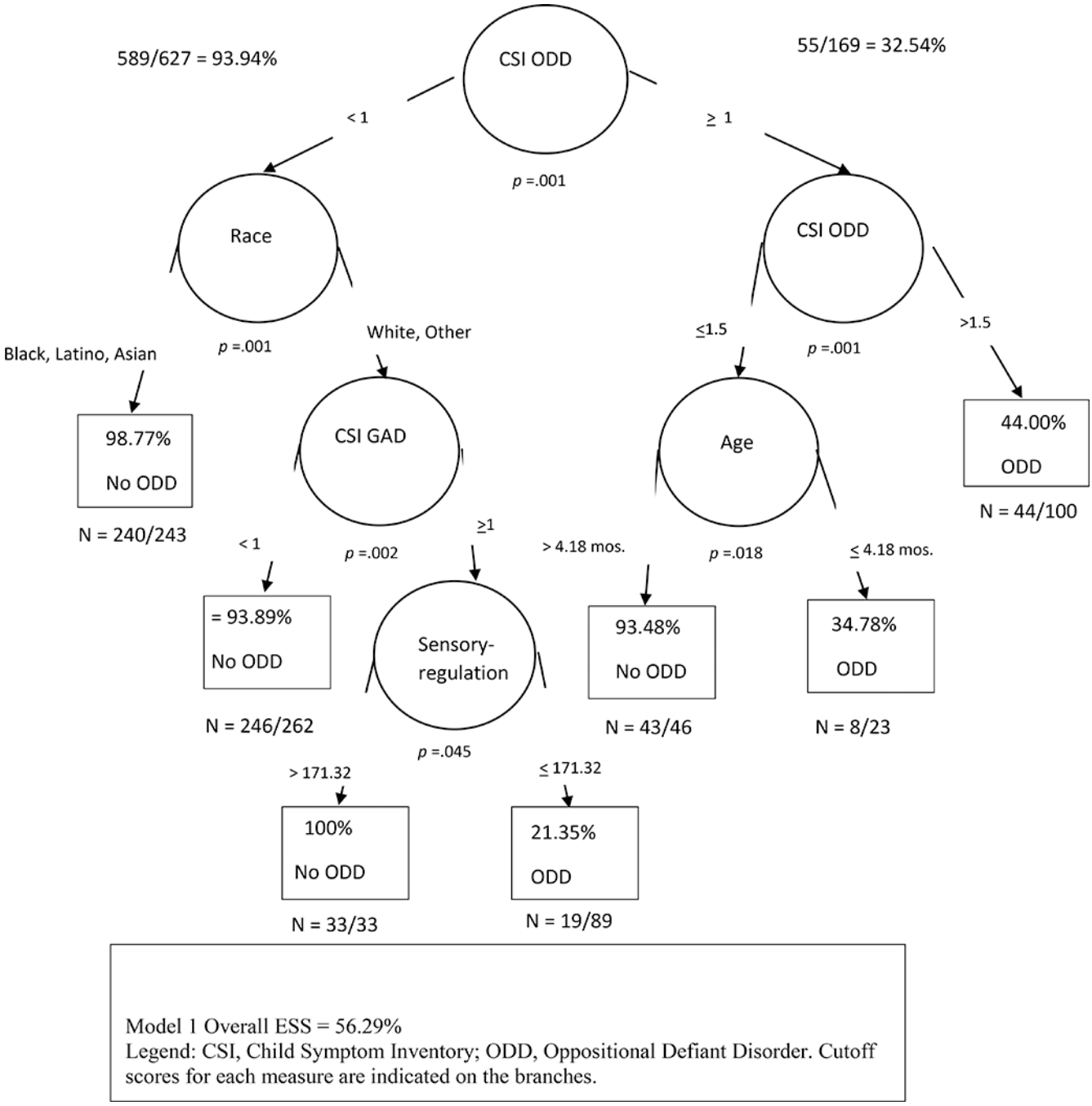

Figure 1:
Pruned tree for age 6 ODD diagnosis with overall ODD scale as an age 4 predictor(Model 1)

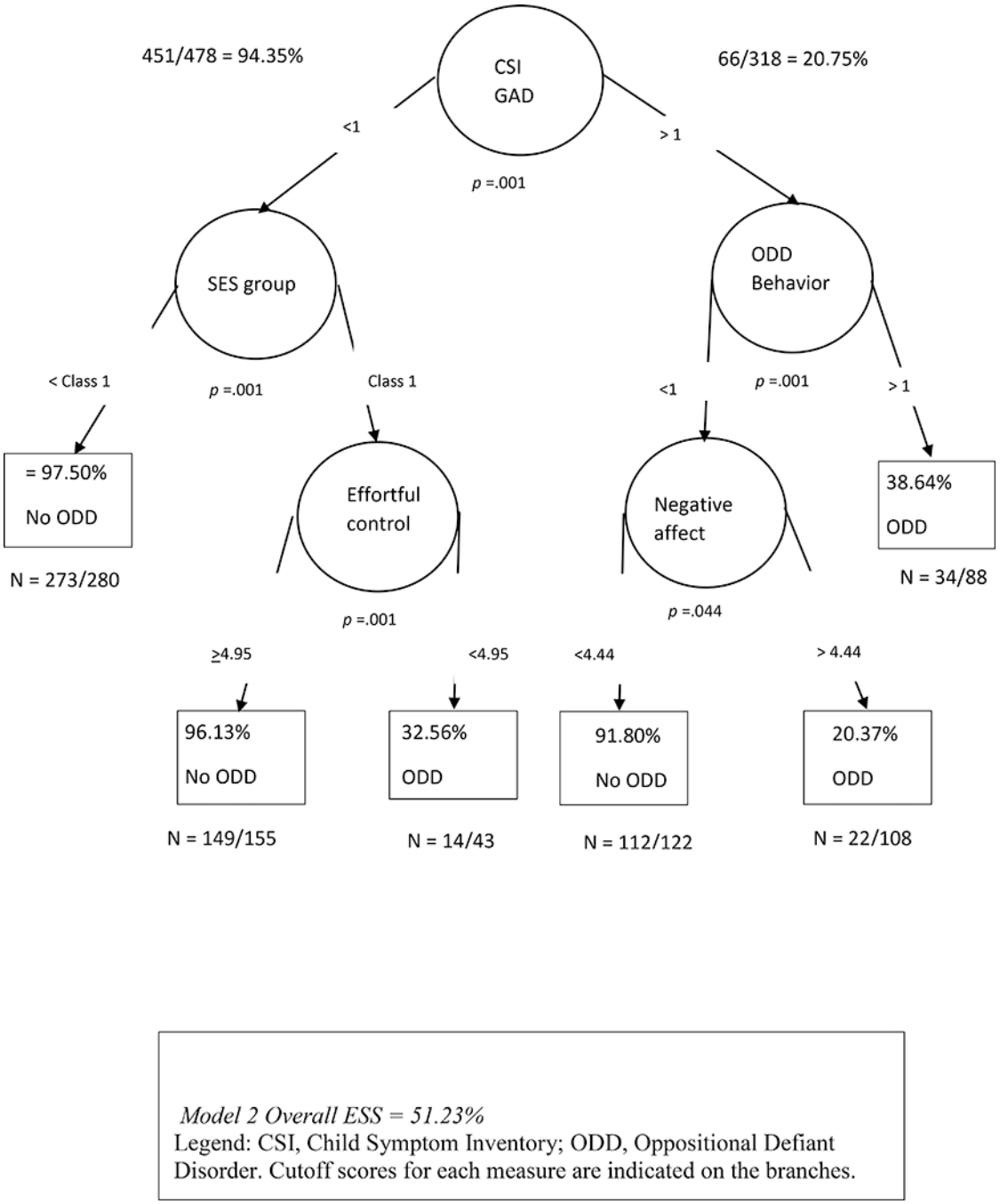

Figure 2:
Pruned tree for age 6 ODD diagnosis with ODD dimensions as predictors (Model 2)

J Clin Child Adolesc Psychol. Author manuscript; available in PMC 2020 January 01.

Table 1.

Individual age 4 predictors of age 6 diagnosis

Domain (age 4)	Age 4 predictors	Mean (SD)	Range	p value (stability)[1]	ESS[2]	Direction of significant effects and cutoff scores (range for continuous measures)
Demographics	Sex	-	-	1		
	Age (wave 1)	3.87 (0.32)	3.87–5.14	.18		
	Marital	-	-	.033 (stable)	10.14%	Married predicts ODD present
	SES group	-	-	.1		
	Race	-	-	.049 (stable)	12.93%	White and Other predicts ODD diagnosis
Context	Conflict (FES)	2.45 (1.86)	0 – 9	.0002 (stable)	22.92%	Conflict > 2.66 predicts ODD present
	Stress (Cohen)	21.64 (7.4)	3 – 52	.0001 (stable)	25.99%	Stress > 20.86 predicts ODD present
Parent	Depression: BDI	5.62 (6.66)	0 – 46.97	.005 (stable)	17.47%	BDI > 5.35 predicts ODD present
Parenting	Support-engagement	44.53 (6.17)	0 – 50	.076		
	Hostile-overall	14.05 (6.0)	0 – 45	.0004 (not stable)		
	Scaffold Overall	0 (2.65)	−12.21 – 5.67	.13		
Child-personality	Attachment security	.40 (.19)	−.44 - .77	.00000 (stable)	28.69%	less secure attachment (<.38) predicts ODD diagnosis
	Sensory regulation total	161.63 (16.2)	73.2 – 190.0	.0000000 (stable)	32.12%	Poorer sensory regulation ability (< 159.82) predicts ODD diagnosis
	Negative affect (NA)	0 (3.05)	−11.9 – 11.0	.004 (stable)	18.64%	more negative temperament (> 4.67) predicts ODD diagnosis
	Effortful Control (EC)	5.16 (.60)	2.96 – 7.0	.000000 (stable)	33.85%	EC <= 4.95 predicts ODD present
	Inhibitory control(Statue)	21.02 (7.49)	0 – 21.05	.001 (not stable)		
Child-cognitive	Verbal (PPVT)	105.03 (15.69)	60 – 148	.58		
Child Psycho-pathology	Oppositional-CSI	0.54 (1.37)	0 – 8	.000000 (stable)	42.92%	ODD ⩾1.0 predicts ODD present
	Anxiety-CSI general	0.65 (1.06)	0 – 10	.000000 (stable)	35.12%	GAD ⩾1.0 predicts ODD present
	Anxiety-Separation anxiety	0.33 (0.89)	0 – 8	.0001 (stable)	20.99%	Separation anxiety⩾1.0 predicts ODD present
	Major depression-CSI	0.04 (0.29)	0 – 4	.20		
	ADHD-Inattentive CSI	0.75 (1.62)	0 – 9	.0001 (stable)	30.09%	ADHDi >⩾1.0 predicts ODD present
	ADHD-Hyperactive-impulsive CSI	1.30 (2.08)	0 – 9	.000000 (stable)	33.29%	ADHDH >⩾1.0 predicts ODD present
	ODD behavior dimension	0.25 (0.65)	0 – 3	.000000 (stable)	35.00%	ODD-B >⩾1.0 predicts ODD present
	ODD negative affect dimension	0.19 (0.45)	0 –3	.000000 (stable)	23.63%	ODD-NA >⩾1.0 predicts ODD present

[1] *p* value refers to the finding of a significant cutoff score. Stability refers to presence or absence of stable findings in leave-one-out (LOO) analyses concerning cross-sample stability of classification results.

[2]ESS = effect strength for sensitivity; ESS summarizes the overall prediction accuracy (0% = chance effect, 100% = perfect classification accuracy). FES-Family Environment Scale; BDI-Beck Depression Inventory; CSI-Child Symptom Inventory; PPVT-Peabody Picture Vocabulary Scale. For scales with mean scores of zero, standard scores were used.

Table 2.

Classification accuracy statistics: Classification trees

Domain (age 4)	Age 4 predictors	ESS[2]	SE	SP	PPV	NPV
	Model 1 (Overall ODD)	56.29%	76.34%	79.94%	33.49%	96.23%
	Model 2 (ODD dimensions)	51.23%	75.27%	75.96%	29.29%	95.87%
Subgroups: overall ODD	Boys	48.26%	67.39%	80.87%	31.96%	94.90%
	Girls	53.20%	80.85%	72.35%	27.74%	96.94%
	White	39.35%	58.33%	81.22%	33.02%	92.47%
	Minorities	63.97%	81.82%	82.15%	31.76%	97.80%
Subroups: ODD dimensions	Boys	47.10%	80.43%	66.67%	24.34%	96.23%
	Girls	53.20%	80.85%	72.35%	27.74%	96.94%
	White	39.55%	58.33%	81.22%	33.02%	92.47%
	Minorities	64.71%	93.94%	70.77%	24.60%	99.14%

Note: ESS, effect size for sensitivity; SE, sensitivity; SP, specificity; PPV, positive predictive value; NPV, negative predictive value; PPVT, Peabody Picture Vocabulary Test