Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2009 November ; 48(11): 1085–1093. doi:10.1097/CHI.0b013e3181ba3dbb.

Maternal ratings of attention problems in ADHD: Evidence for the existence of a continuum

Gitta H. Lubke, PhD¹, James J. Hudziak, MD², Eske M Derks, PhD³, Toos C.E.M. van Bijsterveldt, PhD⁴, and Dorret I. Boomsma, PhD.⁴

¹ University of Notre Dame, IN

² University of Vermont

³ University of Utrecht

⁴ VU University Amsterdam

Abstract

Objective—To investigate whether items assessing attention problems provide evidence of quantitative differences or categorically distinct subtypes of attention problems (AP), and to investigate the relation of empirically derived latent classes to DSM-IV diagnoses of subtypes of attention deficit hyperactivity/impulsivity disorder (ADHD), e.g., combined subtype (CT), Predominantly Inattentive Type (PI), and Predominantly Hyperactive/Impulsive Type (H/I).

Method—Data on attention problems were obtained from maternal ratings on the Child Behavior Check List (CBCL). Latent class models, which assume categorically different subtypes, and factor mixture models, which permit severity differences, are fitted to data obtained from Dutch boys at age 7 (N=8079), 10 (N=5278) and 12 (N=3139). The fit of the different models to the data is compared in order to decide which model, and hence which corresponding interpretation of AP, is most appropriate. Next, ADHD diagnoses are regressed on latent class membership in a subsample of children.

Results—At all 3 ages, models that distinguish between three mainly quantitatively different classes (e.g., mild, moderate, and severe attention problems) provide the best fit to the data. Within each class, the CBCL items measure three correlated continuous factors that can be interpreted in terms of hyperactivity/impulsivity, inattentiveness/dreaminess, and nervous behavior. The AP Severe class contains all subjects diagnosed with ADHD-CT. Some subjects diagnosed with ADHD PI are in the moderate AP class.

Conclusions—Factor Mixture analyses provide evidence that the CBCL AP syndrome varies along a severity continuum of mild, moderate, to severe attention problems. Children affected with ADHD are at the extreme of the continuum. These data are important for clinicians, research scholars, and the framers of the DSM-V as they provide evidence that ADHD diagnoses exist on a continuum rather than as discrete categories.

Keywords

latent class analysis; factor mixture analysis; attention problems; CBCL; DSM

Correspondence to Gitta Lubke, University of Notre Dame, Psychology, 118 Haggard Hall, Notre Dame, IN 46556, glubke@nd.edu.

This article is the subject of an editorial by Dr. Anita Thapar in this issue.

The authors report no conflicts of interest.

Introduction

Is it best to consider Attention deficit/hyperactivity disorder (ADHD) as a categorical disorder or as an extreme of a continuous trait? This is one of the many questions that the DSM-V Child Work group is considering. The question is an important one as the current diagnostic rules as defined in the DSM-IV-TR identify three distinct subtypes (e.g., combined (CT), predominantly inattentive (PI), and predominantly hyperactive/impulsive (H/I)).¹ These subtype diagnoses are based on the presence of at least 6 of 9 inattention items (PI), 6 of 9 hyperactive/impulsive items (H/I) or at least 6 of each (combined type) beginning prior to age 6 and causing impairment in at least two settings. There are no differences in recommendations for diagnostic cut-points for the age or gender of the individual. The DSM-IV items, diagnostic subtypes, and diagnostic rules have been the subject of intense research scrutiny. Critics of this categorical approach point to the fact that the same criteria are applied to girls as to boys, to 6 year olds and to 18 year olds, and set up nonsensical scenarios where a child with academic problems who has 10 of the 18 items does not meet DSM-IV diagnostic criteria for ADHD (e.g. the child has 5 symptoms of Inattention and 5 symptoms of Hyperactivity/Impulsivity) where a child with 6 symptoms (e.g. 6 Inattention symptoms) does meet criteria.² Indeed the members of the DSM-V Disruptive Disorder workgroup are considering how best to recraft the ADHD criteria and are considering both continuous and categorical approaches.²

A variety of research teams have been investigating the categorical-continuum debate. Using latent class approaches, some have argued that ADHD items exist on a severity continuum divided across the attention problem items and the hyperactivity/impulsivity items.³ Some studies conclude that the liability to develop attention problems is continuous, and that clustering of subjects in terms of subtypes neglects variation in severity.^{3–5} Others have argued that the latent classes are replicable across a wide variety of samples, and that they represent a series of genetically discrete disorders.^{6–10} Where prior studies have used LCA or factor analyses in the study of ADHD, in this study we use the Factor Mixture Modeling approach, which combines latent class and factor approaches in order to determine whether ADHD is best conceptualized as a continuous trait or a categorical diagnosis.

To do this work we use the AP syndrome scale of the CBCL,¹¹ which is a widely used instrument to screen for problem behaviors in children. In practice, screening is carried out by summing the item scores in conjunction with established cut-off points (i.e, the minimum sum score to obtain a positive diagnosis). For instance, a cut-off point of 60 when summing the item scores of the CBCL AP scale discriminated well between DSM defined ADHD and non-ADHD patients.¹² In the current analysis, *item level* rather than symptom sum level data are used to permit a more fine grained analysis of specific attention problems. Sum scores treat the scale items uniformly, and in case of differential importance of certain items, or in case subsets of items measure subtype specific symptoms, an item level analysis is likely to capture these differences more adequately. To assess DSM-IV diagnoses and ADHD subtypes we used the Diagnostic Interview Schedule for Children-IV because it had been used by prior groups to determine the relation between categorical and quantitative conceptualizations of attention problems.

The analysis approach of the current study is similar to a recent analysis of the SWAN ADHD data obtained from the Northern Finnish Birth Cohort (NFBC), which showed that in both sexes factor mixture models with severity differences within class provided a clearly superior fit to the data when compared to latent class models.⁴ The results were interpreted as evidence that severity differences are substantial, and that the SWAN ADHD data can best be described in terms of quantitatively ordered classes that differentiate between the unaffected majority and a potentially affected minority. The NFBC data were obtained from

adolescents whereas the current study focuses on data from children 7–12 years old. It has been argued that subtypes of ADHD might be more pronounced closer to latency age. The current study addresses this issue by comparing samples at different ages starting at age 7 and bridging the time interval to early adolescence. Our samples might therefore provide evidence of a decreasing subtype pattern with increasing age.

The central question of whether or not ADHD exists as three discrete diagnostic subtypes or a continuous liability for an attention syndrome that may change across development is largely unanswered. If ADHD exists as a continuous liability, further work will need to be done to determine mediators and moderators of that liability – these include factors such as age of the child, gender of the child, ethnicity and culture, and who the informant is. From a scientific point of view it is difficult to imagine that the same ADHD diagnostic categorical cut point would apply to a 6 year old female from China as would apply to a 18 year old male from the US. Thus contributing to solving the question of whether or not ADHD is best conceptualized as a diagnostic category versus a continuum will have impact on future research, on diagnosis, on treatment, and perhaps on the DSM-V conceptualization of ADHD.

Method

Subjects

The subjects in this study are Dutch male twins whose parents voluntarily registered with the Netherlands Twin Registry (NTR).^{13, 14} The NTR families largely represent the general Dutch population. Based on available data, the average age of the mother and father at birth of the twin is 30.6 and 33.01, respectively. From age 7 to age 12 of the twins, the percentage of married parents decreases from 92.3% to 88.1%. Parents' educational and occupational levels are presented in Table 1.

We use mothers' ratings on the CBCL attention scale at age 7, 10, and 12. Twins are treated as individuals in the current analysis. To account for non-independence of observations, a sandwich type estimator is used to obtain standard errors, see Analysis section. At age 7, N=8079, at age 10 N=5278, and at age 12 N=3139 twins. The smaller N at later ages reflects the longitudinal design of the study (not all children have reached ages 10 and 12). Note that the samples at age 10 and 12 are not exact subsamples of the sample at age 7, but may contain children who were not tested at age 7 (subjects entered the study at a later age or mother reports are missing at earlier ages).

To investigate confounding of changes over time with selection effects in our 3 age samples, we also created a subsample of subjects that was observed at each time point. This longitudinal subsample consists of N=2531 twins. The pattern of results is the same as in the overlapping cross sectional sample. Since the smaller sample size provides less power to detect subtypes if present we focus on the results from the cross sectional sample.¹⁵ The DSM symptoms (see below) were obtained in a selected subsample of N=489. Subtype prevalence in this subset is provided below.

At age 12, we asked if methylphenidate had been prescribed (question available for 86% of the 12-year-olds). This was the case for 96 children (1.4%), and 50 children (0.7%) that had used methylphenidate before but not at present. Data from these children were included in the analyses.

Procedure

Parents were sent the CBCL and were asked to return it by mail. Individuals within the NTR are invited to participate at each wave of data collection, regardless of their previous

participation. Previous work has demonstrated that the responders and non-responders differ only on socioeconomic status, and that the effect size of this difference is negligible.¹⁴ Data were entered, under anonymous IDs, into a phenotype database.¹³ The Diagnostic Interview Schedule for Children (DISC) was administered at age 12 to a subset of the current sample (N=985 total, of which N= 489 males). The selection was based on CBCL scores and aimed at obtaining subjects with high and with low CBCL scores. Since the probability of being selected for the DISC depends only on the CBCL, and since the CBCL data are included in the current analysis, the selection does not induce non-random missingness.¹⁶ Details of the selection procedure are described in Derks et al..¹⁷

Measures

The CBCL is a standardized questionnaire used for parents to respond to 118 problem behaviors exhibited by their child over the previous 6 months. The parent responds along a 3-point scale where 0, 1, and 2 indicating that the behavior is not true, sometimes true, or often true for the child. The psychometric stability of the CBCL has been well established in American and Dutch samples.^{11, 18} The analyses performed here use maternal reports on the 1989 version of the Dutch CBCL. The Attention problems (AP) scale consists of 11 items. The items are shown in abbreviated form in Figure 1.

The DISC is a structured interview that assesses DSM-IV symptoms including those of ADHD. Mothers were asked to indicate whether a symptom was displayed by the child during the last year. The symptoms were aggregated according to the DSM-IV type A criteria to obtain two binary variables indicating presence or absence of an H/I diagnosis and a PI diagnosis, respectively (e.g., 6 symptoms or more). Subjects with both diagnoses belong to the combined subtype (CT). In the cross-sectional samples, the sample size with DSM data at age 7, 10, and 12 are $N_7=449$, $N_{10}=336$, and $N_{12}=331$. In the longitudinal sample $N=284$ subjects have DSM data.

Due to subject overlap, the prevalence of the three subtypes in the larger cross sectional sample is very similar across age and sample type, namely $CT=.065$, $PI=.07$, $H/I=.031$. Note that due to the selection structure of the DSM subsample, these rates only reflect the distribution of subtypes in the subsample with DSM data, but do not necessarily characterize the full samples.

Data Analysis

Models

The majority of previous studies using item level data to investigate the ADHD phenotype relied either on factor analysis (FA), or on latent class analysis (LCA). FA and LCA are latent variable models, which are based on the general concept that observed responses on items of a scale covary because of a small number of underlying latent variables that correspond to the construct of interest (e.g., attention problems). FA uses *continuous* latent variables (i.e., factors), which represent gradual (severity) differences. LCA, on the other hand, uses a *categorical* latent variable with 2 or more categories called latent classes. The latent classes represent different types within a population. Within a type or class, observed items are specified to have zero correlations, such that mean differences between classes explain the overall covariances between observed items. Importantly, fitting either a FA or a LCA model does not test whether the latent variables are continuous or categorical. Only a comparison of FA and LCA within a general statistical framework can assess which model type, and therefore, which type of latent variable, provides a better fit to the data. A better fit of FA models would suggest severity differences, whereas a better fit of LCA models would suggest subtypes.

Factor mixture modeling (FMM) provides such a general framework. FMM extends LCA and FA by combining the two in a single general model. Within each latent class, instead of specifying that variables have zero correlations as in LCA, FMM permits to specify a factor model. The factors within class can capture potential severity differences within class. An FMM with zero factor variance reduces to LCA, and an FMM with a single class reduces to FA. The model was proposed by different authors,^{19–21} and is used in a wide variety of different fields (e.g., developmental psychology,²² addiction,²³ criminology,²⁴ psychiatry⁴).

In the current study, we fit factor models, latent class models, and factor mixture models to CBCL attention items and do so in large samples of boys at ages 7, 10, and 12. Simulation studies have shown that comparing model fit of FA, LCA, and FMM leads to correct model choice in a wide variety of settings.^{15, 25} This is the approach followed here. Since all models are specified within the same general framework, indices of model fit can be used to decide which model is the best fitting model.^{15, 21, 26} It has been shown that the Bayesian Information Criterion (BIC) performs better or equal compared to the Akaike Information Criterion (AIC), the consistent AIC (CAIC), and the adjusted BIC.²⁷ The same study also showed that the BIC clearly outperforms the adjusted likelihood ratio test (aLRT) when comparing FMMs.²⁸ The bootstrapped LRT performs also clearly better than the adjusted LRT, but is not feasible in this study due to computation times. We base our decisions on the BIC, and present the AIC, the CAIC, and the adjusted BIC for completeness.

DSM diagnoses of CT, H/I and PI observed at age 12 are regressed on latent class membership. We estimate the relation between the latent classes and DSM subtype diagnoses in a single analysis. An alternative would be to first assign subjects to the latent classes, and then compute the prevalence of the different diagnoses in each class. However, classification error in assigning subjects to classes can be substantial (e.g., >80% in smaller classes of affected subjects), leading to severely biased prevalence rates.²⁹ The one step approach avoids classification errors, and regressing subtype diagnoses on latent classes results in the estimated proportions of the different diagnoses within each class in a single analysis.

Analysis

All analyses are carried out with Mplus using data from all twins.³⁰ To obtain correct standard errors in the presence of dependent observations, we use a robust “sandwich-type” estimator (MLR estimator).^{31, 32}

An initial exploratory factor analysis showed three eigenvalues larger than 1. The corresponding 3-factor structure has a clear and interpretable loading pattern. The first factor is largely defined by items representing symptoms of hyperactivity/impulsivity (see items 1, 2, 3, 6, 9, and 10 in Figure 1). The second factor explains common variance of indicators of inattentiveness/dreaminess (items 1, 4, 5, and 11). The third factor is defined by the two items related to nervous behaviors (items 7 and 8). The high covariance between these two items as captured by the third factor may be due to similar item wording. The path diagram showing the structure of the 3-factor model is shown in Figure 1.

For the main analysis, we fitted 7 different models to each of the three age groups. Models 1–3 are factor mixture models with 2, 3, and 4 latent classes. Models 4–7 are latent class models with 3, 4, 5, and 6 classes. Based on the factor structure of the initial exploratory factor analysis, the factor mixture models are specified with three correlated factors within class. In the first part, we focus on the structure of AP and fit models without integrating DSM diagnoses. Then the selected best fitting model is fitted again to the data while incorporating logistic regressions of the DSM diagnoses (CT, H/I and PI) on the latent class

variable. Suppose a latent class model is the best fitting model. Since DSM diagnoses are available only in a selected subsample, this part of the analysis will show whether the diagnosed subjects are more likely to belong to a particular latent class (e.g., PI subjects might have a very high probability of belonging to an “inattentive” class, or, all diagnosed subjects might belong to a “severe” class).

Results

Results are compared across the age groups focusing on (i) the best fitting model, (ii) the relative sizes of the latent classes (i.e., class proportions), and (iii) the differences across classes with respect to the response patterns on the 11 CBCL attention items, and the relation of the classes to the DSM diagnoses.

Best fitting models in the three age groups

The general pattern of results is the same for the larger, cross-sectional samples and the smaller longitudinal sample that contains individuals with complete data at all three time points. Due to this similarity, we focus on the results of the larger cross-sectional sample because of the higher power to detect subtypes. In all samples, FMMs have a clearly better fit than the FA or LCA models when considering BIC. Table 2 shows the results for the cross-sectional samples. All information criteria are clearly lower for any of the factor mixture models than for the LCA models with similar model parsimony. To achieve a minimum BIC, LCA requires 8-classes (208 estimated parameters) for the 7-year olds, and 7 classes (183 parameters) for the 10 and 12 year olds. The fact that LCA models have a much higher BIC than FMMs indicates lack of model parsimony.

The model fitting results show that there is substantial variation in attention problems *within* the classes. When comparing the factor mixture models with 2, 3, and 4 classes, it is evident that the power to detect smaller classes decreases with sample size. At age 7, the BIC favors the 3-factor 3-class factor mixture model, at age 10, the BIC does not differ much between the 2 and the 3-class models, and at age 12, the BIC favors the 2-class model. The lack of power to detect the third class is also evident in the smaller sample of subjects with data at all three ages. In the smaller sample, the two-class FMM is the best fitting model based on all information criteria at all three ages.

Since lack of power is the main reason for the better fit of the 2-class model in the sample of 12 year olds compared to the larger samples at age 7 and 10, we base the more detailed comparison below on results corresponding to the 3-class factor mixture model in all age groups.

Relative class sizes of the 3-class factor mixture models compared across age

In the larger overlapping cross sectional samples, the 7-year olds have a larger high scoring class (20.3%) than the 10 and 12 year olds who are very similar with respect to the class proportions (14.5% and 15.6%). The unaffected low scoring majority class has approximately the same size at all ages (62.3%, 63.6%, and 61.5%).

Qualitative vs. quantitative differences

The response patterns on the 11 items are very similar in the larger, cross sectional samples. Figure 2 shows the results for the larger samples since power to detect three classes was sufficient only in the larger samples. As can be seen in the figure, the three classes are mainly quantitatively ordered in all three age groups. Class 1 has a higher probability of scoring “very true” on most items, and class two has higher probabilities than class three. Strictly quantitative differences would be reflected in parallel response profiles on the 11

items, whereas qualitative differences would be reflected in cross-overs with one class scoring high on hyperactivity items but low on inattentiveness items, and another class showing the reverse profile. Cross-overs in response profiles are largely absent in Figure 2. There are three items at age 7 that form an exception. Class two has a slightly higher response probability on items 5 (daydreams) and item 9 (“poor schoolwork”). In addition, class 3 scores slightly higher on item 3 (“can’t sit still”) than class 2, although class 1 has clearly the highest probability of scoring in the highest response category on that item. At age 10, items 5 and 9 show the same tendencies as at age 7. However, none of those differences reach statistical significance in this large sample.

The relation between the CBCL classes and the DSM based diagnoses of H/I, PI, and CT is summarized in Table 3, which provides the proportions of boys with a given subtype diagnosis in the high, moderate and low scoring classes at age 7, 10, and 12. The proportions are derived from the logistic regression of CT, H/I, and PI diagnoses on the latent class variable in the 3-class 3-factor mixture model. It should be noted that the proportions contain the prediction error of the logistic regression. A different approach would be to assign all subjects to their most likely latent class, and then compute the proportions of subtype diagnoses in each class. However, this approach would accumulate the prediction error and the error in class assignment. Class assignment error can be extremely high especially in the smaller classes (e.g. >80% incorrect assignment).²⁹

Based on our analysis we can conclude reliably that the probability of either diagnosis in the low scoring majority class is zero in all three age groups. Furthermore, in all three age groups, the highest scoring class contains all or almost all subjects with a diagnosis of CT or H/I. Depending on age, 30% to 45% of the subjects with a diagnosis of PI belong to the high scoring class, and the moderate scoring class contains the remaining subjects with a diagnosis of PI.

Discussion

The results of the current analysis show quantitative differences in the AP syndrome of the CBCL in children aged 7–12. The FMM analyses of the CBCL data reveal similar results as our earlier findings when we used the same approach with SWAN data obtained in the Finnish adolescents.⁴ The analysis of CBCL AP items shows that the samples consist of three latent classes that are located along correlated continua (“severe”, “moderate”, and “low” scoring AP classes). The severe and moderate classes are small (6–15%, depending on the age group) whereas the ‘low’ scoring class is the largest class (consistent with over 50% of children having low or no attention problems). These findings, using the FMM approach, advance the argument that the AP syndrome exists on a severity continuum, with evidence of a similar class structure across the developmental period of ages 7–12. Especially the sample of 7-year old boys (N=8079) has sufficient power to detect subtypes if present. With a general prevalence of ADHD of 8–12%, approximately 800 of the 7-year olds would be diagnosed, and subtypes within such a large group would be detectable using FMMs.^{15, 25} The current analysis, however, shows that even in younger children AP is best described in terms of severity differences, which matches our conclusion drawn from the analysis of adolescents.⁴

When comparing the different age groups, it is interesting that the two items that most closely map on the DSM H/I subtype, ‘can’t sit still’ and ‘impulsive’ both diminish in intensity with age (e.g. from age 7 to 10 to 12). This finding is consistent with the literature that hyperactivity symptoms diminish with age yet attention problems persist. We observed this pattern in both the larger, overlapping age samples, and in the sample with identical

subjects at the three time points, hence the diminished intensity is not due to changes in the composition of the samples.

The fact that the three CBCL classes are ordered quantitatively is also reflected in the relation between AP and DSM-IV subtype diagnoses in a subsample observed at age 12. All or almost all children with a DSM-IV ADHD CT and H/I are in the 'severe' AP class. Children with DSM-IV PI are divided over the severe and moderate AP classes. Perhaps most important, none of the children with a DSM-IV ADHD subtype are in the low scoring majority class.

As the DSM-V process moves forward it will be important to consider these findings in light of the consideration of including a quantitative axis of diagnostic description. We have argued that a quantitative approach that allows for differences across ages and genders makes sense for both research and clinical work in children who suffer from psychopathologic conditions such as ADHD.³

From this work a clinician will benefit from knowing that attention problems exist on a severity continuum, thus presenting a clear invitation to develop evidenced based interventions that aim towards diminishing the severity of the symptoms within the continuum. In this way the treatment of ADHD is no different than the treatment of hypertension, in which a reasonable evidenced based method can be developed to evaluate and measure the movement from a pathological level (e.g. severe class AP or a diastolic pressure of 100) to a non-pathological level (e.g. AP low class or a diastolic BP of 80) rather than to the absence of attention or the absence of blood pressure. Further, the contention that subtypes of DSM-IV ADHD are not different in their FMM class membership may allow a more general treatment approach toward children with ADHD, which is closer to what happens in most clinics today in any event. For example, most clinicians do not vary their pharmacologic or behavioral treatments based on whether or not the child has DSM-IV ADHD CT, H/I, or PI.

Taken together, these data argue for considering DSM-IV ADHD as existing on a severity continuum rather than as discrete diagnostic categories. Implicit in the continuum argument is the need to identify common mediators of risk. In the area of ADHD it is obvious that the current approach of applying the same criteria to individuals of both gender and all ages is unrealistic. The continuum argument allows for the creation of normative distributions by age, gender, informant and ethnicity. Such advances, which seem so simple to accept, should be considered as key modifications in the DSM-V or subsequent editions of our diagnostic manuals.

Our study has a number of limitations. First, we focus exclusively on AP in boys. Our rationale is that prevalence of attention problems is higher in boys, and that the statistical power to detect subtypes increases with prevalence rates. The increasing sample size in the NTR will permit an analysis of attention problems in girls in the future. A second limitation concerns the fact that we relied on a specific statistical approach to detect subtypes, FMM. Other approaches such as the taxometric procedures developed by Meehl and colleagues have been used for this purpose.³³ However, it has been shown that taxometric procedures have less power to detect classes than FMM.³⁴ Third, we treated twins as individuals, thereby neglecting the genetically informative structure of the sample.³⁵ A twin mixture model has recently been proposed, however, the model decomposes variance *within* class into genetic and environmental components rather than the more interesting decomposition of differences *between* classes.³⁶ In addition, twin mixture models assume that correct estimation of within class variance is unproblematic. However, that this may not be the case, especially when class proportions differ substantially (e.g., small minority classes and large

majority classes).²⁹ Fourth, regarding the utility of factor mixture modeling to support selection of subjects for prevention or treatment, it should be noted that simulation studies have demonstrated high error rates in assigning subjects to classes.²⁹ This clearly limits the potential of mixture analyses for selection purposes. The current study shows that factor mixture analyses may be used to *exclude* subjects that are very unlikely to be affected (i.e., the low scoring majority class). Finally, the current study may be enhanced by including relevant gene candidates to predict class membership. Recent work demonstrates that that substantial sample sizes are needed to reliably detect small gene effects using FMMs.³⁷

Acknowledgments

This research was supported by Spinozapremie (NWO/SPI 56-464-14192; Twin-family database for behavior genetics and genomics studies (NWO 480-04-004); the VU-CNCR (Centre Neurogenetics/Cognition Research), Developmental Study of Attention Problems in Young Twins (NIMH, RO1 MH58799-03).

References

1. American Psychiatric Association (APA). Diagnostic and Statistical Manual of Mental Disorders. 4. Washington, D.C: American Psychiatric Assn; Text Revision (DMS-IV-TR)
2. Hudziak, JJ.; Achenbach, TA.; Althoff, RR.; Pine, DS. A Dimensional Approach to Developmental Psychopathology. In: Helzer, JE.; Kraemer, HC.; Krueger, RF.; Wittchen, HU.; Sirovatka, RJ.; Regier, DA., editors. Dimensional Approaches in Diagnostic Classification Refining the Research Agenda for DSM-V. American Psychiatric Association; 2007. p. 101-14.2008
3. Hudziak JJ, Heath AC, Madden PF, Reich W, Bucholz KK, Slutske W, Bierut LJ, Neuman RJ, Todd RD. Latent class and factor analysis of DSM-IV ADHD: A twin study of female adolescents. *J Am Acad Child Adolesc Psychiatry*. 1998; 37:848–57. [PubMed: 9695447]
4. Lubke GH, Muthen B, Moilanen IK, et al. Subtypes versus severity differences in attention-deficit hyperactivity disorder in the Northern Finnish Birth Cohort. *J Am Acad Child Adolesc Psychiatry*. 2007; 46(12):1584–1593. [PubMed: 18030080]
5. Polderman TJ, Derks EM, Hudziak JJ, Verhulst FC, Posthuma D, Boomsma DI. Across the continuum of attention skills: a twin study of the SWAN ADHD rating scale. *J Child Psychol Psychiatry*. 2007; 48(11):1080–1087. [PubMed: 17995483]
6. Neuman RJ, Sitdhiraksa N, Reich W, et al. Estimation of Prevalence of DSM-IV and Latent Class-Defined ADHD Subtypes in a Population-Based Sample of Child and Adolescent Twins. *Twin Res Hum Genet*. 2005; 8:392–401. [PubMed: 16176725]
7. Rasmussen ER, Neuman RJ, Heath AC, Levy F, Hay DA, Todd RD. Familial clustering of latent class and DSM-IV defined attention-deficit/hyperactivity disorder (ADHD) subtypes. *J Child Psychol Psychiatry*. 2004; 45(3):589–98. [PubMed: 15055377]
8. Rasmussen ER, Neuman RJ, Heath AC, Levy F, Hay DA, Todd RD. Replication of the latent class structure of Attention-Deficit/Hyperactivity Disorder (ADHD) subtypes in a sample of Australian twins. *J Child Psychol Psychiatry*. 2002; 43(8):1018–28. [PubMed: 12455923]
9. Todd RD, Lobos EA. Mutation screening of the dopamine D2 receptor gene in attention-deficit hyperactivity disorder subtypes: preliminary report of a research strategy. *Am J Med Genet*. 2002; 114(1):34–41. [PubMed: 11840503]
10. Todd RD, Sitdhiraksa N, Reich W, Ji TH, Joyner CA, Heath AC, Neuman RJ. Discrimination of DSM-IV and latent class attention-deficit/hyperactivity disorder subtypes by educational and cognitive performance in a population-based sample of child and adolescent twins. *J Am Acad Child Adolesc Psychiatry*. 2002; 41(7):820–8. [PubMed: 12108807]
11. Achenbach, TM. Manual for the Child Behavior Checklist/4–18 and 1991 profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1990.
12. Biederman J, Faraone SV, Doyle A, et al. Convergence of the Child Behavior Checklist with structured interview-based psychiatric diagnoses of ADHD children with and without comorbidity. *J Child Psychol Psychiatry*. 1993; 34(7):1241–51. [PubMed: 8245144]

13. Boomsma DI, de Geus EJ, Vink JM, et al. Netherlands Twin Register: From twins to twin families. *Twin Res Hum Genet.* 2006; 9:849–57. [PubMed: 17254420]
14. Bartels M, van Beijsterveldt TCEM, Derks EM, et al. Young Netherlands Twin Register (Y-NTR): A longitudinal multiple informant study of problem behavior. *Twin Res Hum Genet.* 2007; 10:3–11. [PubMed: 17539360]
15. Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors with categorical outcomes: Class invariance of parameters of factor mixture models. *Multivariate Behavioral Research.* 2008; 43:592–620. [PubMed: 20165736]
16. Little, RJA.; Rubin, DB. *Statistical analysis with missing data.* 2. Wiley and Sons; New York: 2002.
17. Derks EM, Hudziak JJ, Dolan CV, van Beijsterveldt TCEM, Verhulst FC, Boomsma DI. Genetic and Environmental Influences on the Relation Between Attention Problems and Attention Deficit Hyperactivity Disorder. *Behav Genet.* 2008; 38:11–23. [PubMed: 18074222]
18. Verhulst, FC.; van der Ende, J.; Koot, H. *Handleiding voor de CBCL/4–18.* Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie Sophia Kinderziekenhuis, Academisch Ziekenhuis Rotterdam, Erasmus Universiteit Rotterdam; 1996. [Manual for the CBCL/4–18]
19. Yung YF. Finite mixtures in confirmatory factor analysis models. *Psychometrika.* 1997; 62:297–330.
20. Dolan CV, van der Maas HLJ. Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika.* 1998; 63:227–53.
21. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics.* 1999; 55:463–9. [PubMed: 11318201]
22. Nylund K, Bellmore A, Nishina A, Graham S. Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development.* 2007; 78(6):1706–22. [PubMed: 17988316]
23. Neale MC, Aggen SH, Maes HH, Kubarych TS, Schmitt JE. Methodological issues in the assessment of substance use phenotypes. *Addictive Behaviors.* 2006; 31(6):1010–34. [PubMed: 16723188]
24. Nagin DS, Land KC. Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed poisson model. *Criminology.* 1993; 31(3):327–62.
25. Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research.* 2006; 41:499–532. [PubMed: 19169366]
26. Lubke GH, Muthén BO. Investigating population heterogeneity with factor mixture models. *Psychol Methods.* 2005; 10:21–39. [PubMed: 15810867]
27. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling.* 2007; 14:535–69.
28. Lo Y, Mendell N, Rubin DB. Testing the number of components in a normal mixture. *Biometrika.* 2001; 88:767–78.
29. Tueller S, Lubke GH. Evaluation of structural equation mixture models: Parameter estimates and correct class assignment. *Structural Equation Modeling.* 2008; 43:592–620.
30. Muthén, LK.; Muthén, B. *Mplus User's Guide.* 4. Los Angeles, CA: Muthen & Muthen; 1998–2006.
31. Asparouhov T. General multi-level modeling with sampling weights. *Communications in Statistics: Theory and Methods.* 2006; 35:439–60.
32. White H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica.* 1982; 50:1–25.
33. Meehl PE. Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality.* 1992; 60:117–74.
34. Lubke GH, Tueller S. Latent class detection and latent class assignment: A comparison of taxometric and factor mixture modeling approaches. (submitted).
35. Martin N, Boomsma DI, Machin G. A twin-pronged attack on complex traits. *Nat Genet.* 1997; 17:387–91. [PubMed: 9398838]

36. Muthén B, Asparouhov T, Rebollo I. Advances in behavioral genetics modeling using Mplus: Applications of factor mixture modeling to twin data. *Twin Res Hum Genet.* 2006; 9:313–24. [PubMed: 16790142]
37. Lubke GH, Carey G, Lessem J, Hewitt J. Using observed genetic variables to predict latent class membership: A comparison of two methods. *Behav Genet.* 2008; 38:612–53.

Figure 1.
Path diagram representing the factor structure of the 11 CBCL attention items CBCL = Child Behavior Checklist

Figure 2.

Response patterns of the three classes resulting from fitting a 3-factor 3-class factor mixture model to data obtained at age 7, 10, and 12. In all three panels, classes 1, 2, and 3 are high, moderate, and low scoring classes, respectively.

Table 1

Educational level parents and occupational level father in the NTR (in percentages)

educational attainment	mother	father	occupational level	father
less than high school	6.6	7.7	elementary (simple tasks)	6.0
high school	34.0	30.5	low-skilled (no professional training needed)	24.8
professional training	38.7	32.7	intermediate (requires some professional training)	38.4
professional degree	15.6	17.4	higher occupations (professional degree)	21.0
masters/doctoral degree	5.0	11.8	scientific (requires masters/doctoral degree)	9.7

NTR = Netherlands Twin Registry

Table 2

Fit indices and class proportions for 7 models fitted to the 3 samples

fitted models	log likelihood	n par	AIC	BIC	sa BIC	CAIC
T7 F3C2	-48556.463	66	97244.926	97706.730	97496.994	97772.730
T7 F3C3	-48385.052	95	96960.103	97624.820	97322.929	97719.820
T7 F3C4	-48286.035	124	96820.071	97687.702	97293.654	97811.702
T7 LCA 3c	-49619.039	68	99374.077	99849.875	99633.784	99917.875
T7 LCA 4c	-49265.222	91	98712.445	99349.174	99059.994	99440.174
T7 LCA 5c	-49026.820	114	98281.639	99079.300	98717.030	99193.300
T7 LCA 6c	-48834.355	137	97942.710	98901.302	98465.943	99038.302
T10 F3C2	-32433.394	66	64998.788	65432.494	65222.767	65498.494
T10 F3C3	-32312.141	95	64814.281	65438.555	65136.676	65533.555
T10 F3C4	-32238.386	124	64724.771	65539.613	65145.581	65663.613
T10 LCA 3c	-33064.229	68	66264.458	66711.307	66495.225	66779.307
T10 LCA 4c	-32873.421	91	65928.842	66526.831	66237.662	66617.831
T10 LCA 5c	-32709.276	114	65646.553	66395.681	66033.426	66509.682
T10 LCA 6c	-32551.301	137	65376.602	66276.870	65841.529	66413.870
T12 F3C2	-17992.177	66	36116.355	36515.764	36306.055	36581.765
T12 F3C3	-17913.431	95	36016.862	36591.770	36289.915	36686.770
T12 F3C4	-17857.500	124	35962.999	36713.405	36319.405	36837.405
T12 LCA 3c	-18319.787	68	36775.574	37187.087	36971.022	37255.087
T12 LCA 4c	-18165.111	91	36512.221	37062.922	36773.777	37153.922
T12 LCA 5c	-18074.379	114	36376.757	37066.646	36704.421	37180.647
T12 LCA 6c	-17999.636	137	36273.272	37102.349	36667.043	37239.349

Note: The number of estimated parameters is denoted as n par. AIC and BIC are abbreviations for the Akaike Information Criterion and the Bayesian Information Criterion, respectively. The saBIC has a different sample size adjustment, and the CAIC is the consistent AIC. T7 (N=8079), T10 (N=5278), and T12 (N=3139) are the samples containing all available data at ages 7, 10 and 12. Models are abbreviates as, e.g., T7 F3C4 for a 3-factor 4-class models in 7-year olds, or T12 LCA 6c for a 6-class latent class model fitted in 12 year olds.

Table 3

Predicted proportions of DSM-IV ADHD subtypes in each of the three classes of the 3-factor 3-class model at age 7, 10, and 12.

	Class	CT	H/I	PI
7 year old	high	1	1	.439
	moderate	0	0	.561
	low	0	0	0
10 year old	high	1	.831	.285
	moderate	0	.169	.753
	low	0	0	0
12 year old	high	.946	1	.360
	moderate	.054	0	.640
	low	0	0	0

Note: The reported quantities are the predicted proportions of combined, predominantly hyperactive/impulsive, and predominantly impulsive (CT, H/I and PI) subtype diagnosis within each class. The proportions are calculated using estimated regression coefficients of the regression of subtype diagnoses on latent class in conjunction with estimated class proportions, and are therefore subject to prediction error, see text.