Maternal Ratings of Attention Problems in ADHD: Evidence for the Existence of a Continuum

GITTA H. LUBKE, PH.D., JAMES J. HUDZIAK, M.D., ESKE M. DERKS, PH.D., TOOS C.E.M. VAN BIJSTERVELDT, PH.D., AND DORRET I. BOOMSMA, PH.D.

ABSTRACT

Objective: To investigate whether items assessing attention problems provide evidence of quantitative differences or categorically distinct subtypes of attention problems (APs) and to investigate the relation of empirically derived latent classes to *DSM-IV* diagnoses of subtypes of attention-deficit/hyperactivity disorder (ADHD), for example, combined subtype, predominantly inattentive type, and predominantly hyperactive/impulsive type. **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 ($N = 8{,}079$), 10 ($N = 5{,}278$), and 12 years ($N = 3{,}139$). The fit of the different models to the data is compared 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 the three 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 of the subjects diagnosed with ADHD–combined subtype. Some subjects diagnosed with ADHD–predominantly inattentive type are in the moderate AP class. **Conclusions:** Factor mixture analyses provide evidence that the CBCL AP syndrome varies along a severity continuum of mild to 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. *J. Am. Acad. Child Adolesc. Psychiatry,* 2009;48(11):1085–1093. **Key Words:** latent class analysis, factor mixture analysis, attention problems, CBCL, *DSM*.

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 (i.e., combined type [CT], predominantly inattentive [PI], and predominantly hyperactive/impulsive [H/I]).[1] These subtype diagnoses are based on the presence of at least six of nine inattention items (PI), six of nine hyperactive/impulsive items (H/I), or at least six of each (CT) beginning before age 6 years and causing impairment in at least two settings. There are no differences in recommendations for diagnostic cut points for the age or sex 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

Accepted July 22, 2009.

This article was reviewed under and accepted by Ad Hoc Editor Kenneth Towbin, M.D.

Dr. Lubke is with the University of Notre Dame; Dr. Hudziak is with the University of Vermont; Dr. Derks is with the University of Utrecht; Drs. Boomsma and van Bijsterveldt are with the VU University Amsterdam.

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).

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

Correspondence to Gitta Lubke, Ph.D., Department of Psychology, University of Notre Dame, 118 Haggar Hall, Notre Dame, IN 46556; e-mail: glubke@nd.edu.

DOI: 10.1097/CHI.0b013e3181ba3dbb

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 five symptoms of Inattention and five symptoms of Hyperactivity/Impulsivity) where a child with six symptoms (e.g., six Inattention symptoms) does meet criteria.[2] Indeed, the members of the *DSM-V* Disruptive Disorder work group 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 (AP) items and the hyperactivity/impulsivity items.[3] Some studies conclude that the liability to develop APs 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 previous studies have used latent class analysis (LCA) or factor analyses in the study of ADHD, in this study, we use the factor mixture modeling (FMM) approach, which combines latent class and factor approaches 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 Child Behavior Check List (CBCL),[11] which is a widely used instrument to screen for problem behaviors in children. In practice, screening is performed by summing the item scores in conjunction with established cutoff points (i.e., the minimum sum score to obtain a positive diagnosis). For instance, a cutoff 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 APs. 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 (DISC-IV) because it had been used by previous groups to determine the relation between categorical and quantitative conceptualizations of APs.

The analysis approach of the current study is similar to a recent analysis of the Strengths and Weaknesses of ADHD Symptoms and Normal Behavior scale (SWAN) ADHD data obtained from the Northern Finnish Birth Cohort, which showed that, in both sexes, factor mixture models with severity differences within class provided a clearly superior fit to the data when compared with 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 Northern Finnish Birth Cohort data were obtained from adolescents, whereas the current study focuses on data from children aged 7 to 12 years. 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 years 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 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, sex of the child, ethnicity and culture, and the informant. 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 subject from China as would apply to an 18-year-old male subject from the United States. Thus, contributing to solving the question of whether ADHD is best conceptualized as a diagnostic category versus a continuum will have an 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 years, respectively. From age 7 to 12 years 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.

TABLE 1
Educational Level of Parents and Occupational Level of Fathers 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

Note: NTR = Netherlands Twin Registry.

We use mothers' ratings on the CBCL attention scale at ages 7, 10, and 12 years. Twins are treated as individuals in the current analysis. To account for nonindependence of observations, a sandwich-type estimator is used to obtain SEs (see "Analysis"). At age 7 years, N = 8,079, at age 10 years, N = 5,278, and at age 12 years, N = 3,139 twins. The smaller N at later ages reflects the longitudinal design of the study (not all children have reached ages 10 and 12 years). Note that the samples at ages 10 and 12 years are not exact subsamples of the sample at age 7 years but may contain children who were not tested at age 7 years (the 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 three age samples, we also created a subsample of subjects that was observed at each time point. This longitudinal subsample consists of 2,531 twins. The pattern of results is the same as in the overlapping cross-sectional sample. Because 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 489 subjects. Subtype prevalence in this subset is provided below.

At age 12 years, we asked if methylphenidate had been prescribed (question available for 86% of the 12-year-olds). This was the case for 96 (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

The 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 nonresponders 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 DISC-IV was administered at age 12 years to a subset of the current sample (n = 985 total, of which n = 489 male subjects). The selection was based on CBCL scores and aimed at obtaining subjects with high and low CBCL scores. Because the probability of being selected for the DISC depends only on the CBCL, and because the CBCL data are included in the current analysis, the selection does not induce nonrandom 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 for the previous 6 months. The parent responds along a three-point scale where 0, 1, and 2 indicating that the behavior is not true, sometimes true, or often true, respectively, 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 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. The 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 and a PI diagnosis, respectively (e.g., six symptoms or more). The subjects with both diagnoses belong to the CT. In the cross-sectional samples, the sample size with *DSM* data at ages 7, 10, and 12 years are n_7 = 449, n_{10} = 336, and n_{12} = 331. In the longitudinal sample, 284 subjects have *DSM* data.

Because of subject overlap, the prevalence of the three subtypes in the larger cross-sectional sample is similar across age and sample type, namely CT = 0.065, PI = 0.07, and H/I = 0.031. Note that because of 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 LCA. Factor analysis 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., APs). Factor analysis uses continuous latent variables (i.e., factors), which represent gradual (severity) differences. Latent class analysis, on the other hand, uses a categorical latent variable with two 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 an FA or an 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 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

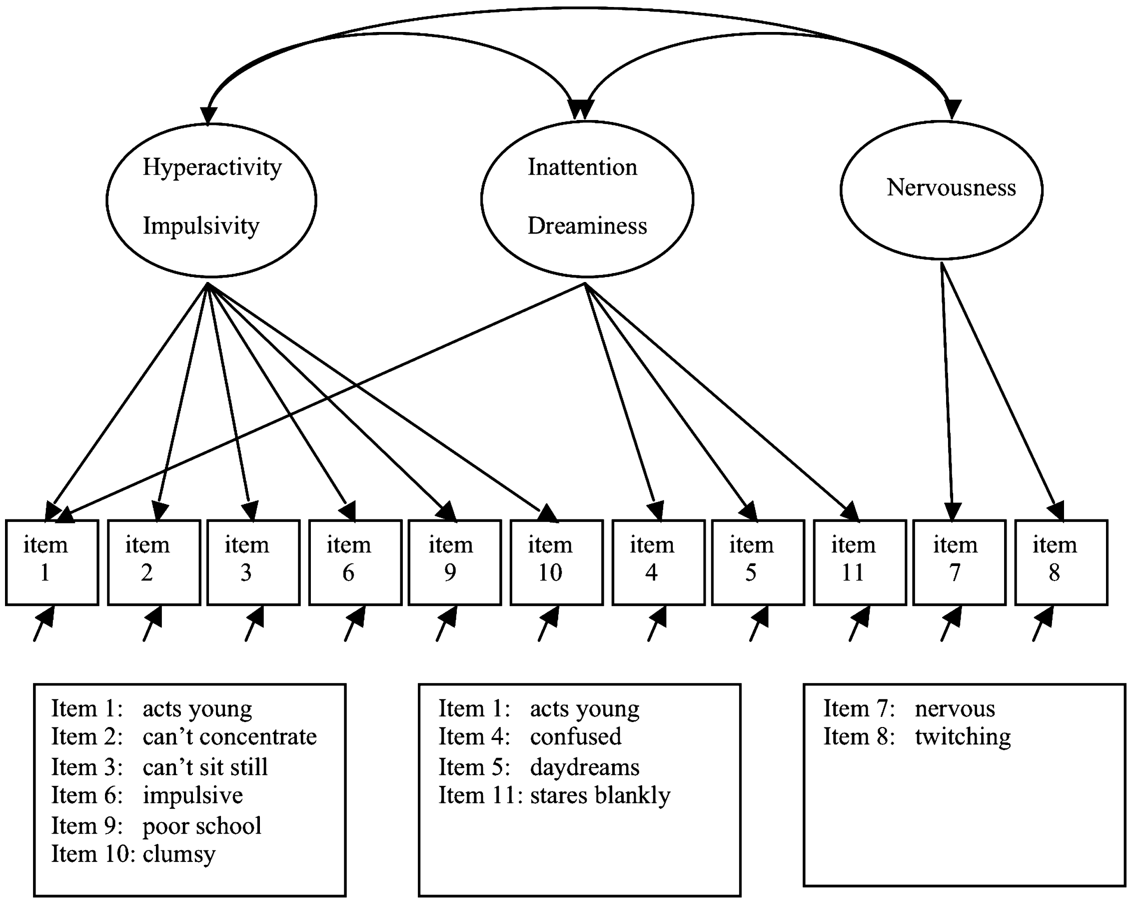

Fig. 1 Path diagram representing the factor structure of the 11 CBCL attention items. CBCL = Child Behavior Checklist.

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] psychiatry[4]).

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 years. 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. Because all models are specified within the same general framework, indices of model fit can be used to decide which one is the best-fitting model.[15,21,26] It has been shown that the Bayesian Information Criterion (BIC) performs better or equally compared with the Akaike Information Criterion (AIC), the consistent AIC, and the adjusted BIC.[27] The same study also showed that the BIC clearly outperforms the adjusted likelihood-ratio test (LRT) when comparing FMMs.[28] The bootstrapped LRT performs also clearly better than the adjusted LRT but is not feasible in this study because of computation times. We base our decisions on the BIC and present the AIC, the consistent AIC, and the adjusted BIC for completeness.

The *DSM* diagnoses of CT, H/I, and PI observed at age 12 years 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 performed with Mplus using data from all twins.[30] To obtain correct SEs in the presence of dependent observations, we use a robust "sandwich-type" estimator (MLR estimator).[31,32]

An initial exploratory FA showed three eigenvalues larger than 1. The corresponding three-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 Fig. 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 three-factor model is shown in Figure 1.

For the main analysis, we fitted seven different models to each of the three age groups. Models 1 to 3 are factor mixture models with two, three, and four latent classes. Models 4 to 7 are latent class models with three, four, five, and six classes. Based on the factor structure of the initial exploratory FA, 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. Because *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., the PI subjects might have a high probability of belonging to an "inattentive" class; all of the diagnosed subjects might belong to a "severe" class).

RESULTS

Results are compared across the age groups focusing on the best-fitting model, the relative sizes of the latent classes (i.e., class proportions), and 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. Because of 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 eight classes (208 estimated parameters) for the 7-year-olds and seven 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 APs within the classes. When comparing the factor mixture models with two, three, and four classes, it is evident that the power to detect smaller classes decreases with sample size. At age 7 years, the BIC favors the three-factor three-class factor mixture model; at age 10 years, the BIC does not differ much between the

TABLE 2
Fit Indices and Class Proportions for Seven Models Fitted to the Three Samples

Fitted Models	Log Likelihood	n par	AIC	BIC	saBIC	CAIC
T7 F3C2	−48,556.463	66	97,244.926	97,706.730	97,496.994	97,772.730
T7 F3C3	−48,385.052	95	96,960.103	97,624.820	97,322.929	97,719.820
T7 F3C4	−48,286.035	124	96,820.071	97,687.702	97,293.654	97,811.702
T7 LCA 3c	−49,619.039	68	99,374.077	99,849.875	99,633.784	99,917.875
T7 LCA 4c	−49,265.222	91	98,712.445	99,349.174	99,059.994	99,440.174
T7 LCA 5c	−49,026.820	114	98,281.639	99,079.300	98,717.030	99,193.300
T7 LCA 6c	−48,834.355	137	97,942.710	98,901.302	98,465.943	99,038.302
T10 F3C2	−32,433.394	66	64,998.788	65,432.494	65,222.767	65,498.494
T10 F3C3	−32,312.141	95	64,814.281	65,438.555	65,136.676	65,533.555
T10 F3C4	−32,238.386	124	64,724.771	65,539.613	65,145.581	65,663.613
T10 LCA 3c	−33,064.229	68	66,264.458	66,711.307	66,495.225	66,779.307
T10 LCA 4c	−32,873.421	91	65,928.842	66,526.831	66,237.662	66,617.831
T10 LCA 5c	−32,709.276	114	65,646.553	66,395.681	66,033.426	66,509.682
T10 LCA 6c	−32,551.301	137	65,376.602	66,276.870	65,841.529	66,413.870
T12 F3C2	−17,992.177	66	36,116.355	36,515.764	36,306.055	36,581.765
T12 F3C3	−17,913.431	95	36,016.862	36,591.770	36,289.915	36,686.770
T12 F3C4	−17,857.500	124	35,962.999	36,713.405	36,319.405	36,837.405
T12 LCA 3c	−18,319.787	68	36,775.574	37,187.087	36,971.022	37,255.087
T12 LCA 4c	−18,165.111	91	36,512.221	37,062.922	36,773.777	37,153.922
T12 LCA 5c	−18,074.379	114	36,376.757	37,066.646	36,704.421	37,180.647
T12 LCA 6c	−17,999.636	137	36,273.272	37,102.349	36,667.043	37,239.349

Note: The number of estimated parameters is denoted as n par. The saBIC has a different sample size adjustment, and the CAIC is the consistent AIC. T7 (N = 8,079), T10 (N = 5,278), and T12 (N = 3,139) are the samples containing all available data at ages 7, 10, and 12 years. Models are abbreviated as, for example, T7 F3C4 for a three-factor four-class model in 7-year-olds or T12 LCA 6c for a six-class latent class model fitted in 12-year-olds. AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; CAIC = consistent AIC; n par = number of estimated parameters; saBIC = sample size adjustment BIC.

two- and the three-class models; and at age 12 years, the BIC favors the two-class model. The lack of power to detect the third class is also evident in the smaller sample of subjects with data at all the three ages. In the smaller sample, the two-class FMM is the best-fitting model based on all information criteria at all the three ages.

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

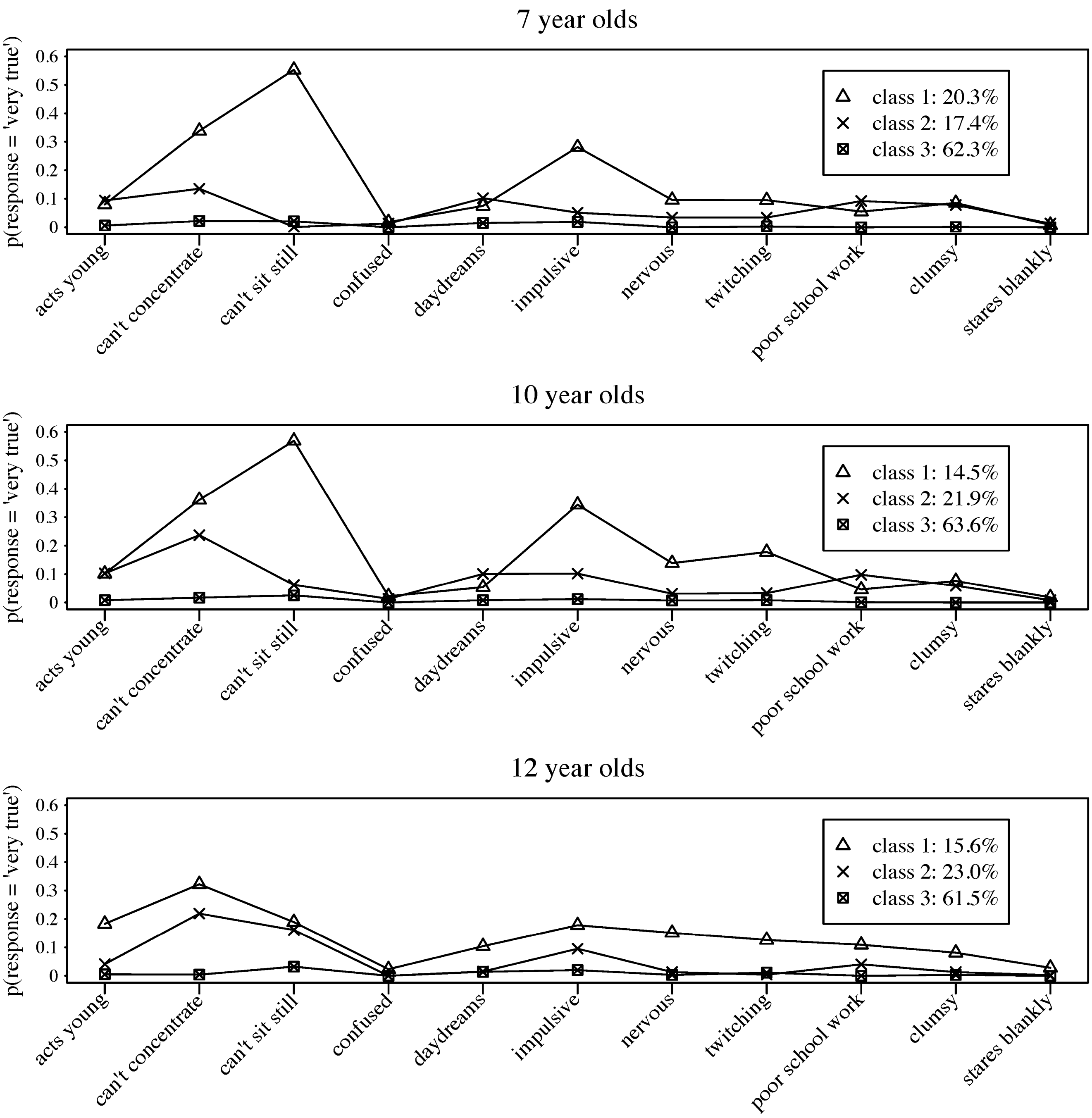

Fig. 2 Response patterns of the three classes resulting from fitting a three-factor three-class factor mixture model to data obtained at age 7, 10, and 12 years. In all three panels, classes 1, 2, and 3 are high-, moderate-, and low-scoring classes, respectively.

Relative Class Sizes of the Three-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 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 Versus Quantitative Differences

The response patterns on the 11 items are similar in the larger cross-sectional samples. Figure 2 shows the results for the larger samples because 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 the three age groups. Class 1 has a higher probability of scoring "very true" on most items, and class 2 has higher probabilities than class 3. Strictly quantitative differences would be reflected in parallel response profiles on the 11 items, whereas qualitative differences would be reflected in crossovers with one class scoring high on hyperactivity items but low on inattentiveness items and another class showing the reverse profile. Crossovers in response profiles are largely absent in Figure 2. There are three items at age 7 years that form an exception. Class 2 has a slightly higher response probability on items 5 (daydreams) and 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 years, items 5 and 9 show the same tendencies as at age 7 years. 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 ages 7, 10, and 12 years. The proportions are derived from the logistic regression of CT, H/I, and PI diagnoses on the latent class variable in the three-class three-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 of the 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]

TABLE 3
Predicted Proportions of *DSM-IV* ADHD Subtypes in Each of the Three Classes of the Three-Factor Three-Class Model at Age 7, 10, and 12 Years

	Class	CT	H/I	PI
7-year-old	High	1	1	0.439
	Moderate	0	0	0.561
	Low	0	0	0
10-year-old	High	1	0.831	0.285
	Moderate	0	0.169	0.753
	Low	0	0	0
12-year-old	High	0.946	1	0.360
	Moderate	0.054	0	0.640
	Low	0	0	0

Note: The reported quantities are the predicted proportions of combined, predominantly hyperactive/impulsive, and predominantly impulsive 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). CT = combined type; H/I = predominantly hyperactive/impulsive; PI = predominantly impulsive.

With our analysis as basis, we can conclude reliably that the probability of either diagnosis in the low-scoring majority class is zero in all the three age groups. Furthermore, in all the three age groups, the highest-scoring class contains all or almost all of the 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 to 12 years. 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 more than 50% of children having low or no APs). 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 to 12 years. Especially, the sample of 7-year-old boys (N = 8,079) has sufficient power to detect subtypes if present. With a general prevalence of ADHD of 8% to 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] However, the current analysis 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, both "can't sit still" and "impulsive" diminish in intensity with age. This finding is consistent with the literature that hyperactivity symptoms diminish with age, yet APs persist. We observed this pattern in both the larger and overlapping age samples, and in the sample with identical subjects at the three time points, 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 years. All or almost all of the children with a *DSM-IV* ADHD CT and H/I are in the severe AP class. The 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 sexes makes sense for both research and clinical work in children who have psychopathologic conditions such as ADHD.[3]

From this work, a clinician will benefit from knowing that APs exist on a severity continuum, thus presenting a clear invitation to develop evidence-based interventions that aim toward 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 evidence-based method can be developed to evaluate and measure the movement from a pathological level (e.g., severe class AP, a diastolic pressure of 100) to a nonpathological level (e.g., low class AP, a diastolic BP of 80) rather than to the absence of attention or the absence of blood pressure. Furthermore, 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 pharmacological or behavioral treatments based on whether 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 sexes and all ages is unrealistic. The continuum argument allows for the creation of normative distributions by age, sex, 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 the prevalence of APs 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 APs 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[33] have been used for this purpose. 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 within-class variance 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, this may not be the case, especially when class proportions differ substantially (e.g., small minority classes, large majority classes).[29] Fourth, regarding the use of FMM 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 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 substantial sample sizes are needed to reliably detect small gene effects using FMMs.[37]

Disclosure: The authors report no conflicts of interest.

REFERENCES

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition Text Revision (DMS-IV-TR)*. Washington, DC: American Psychiatric Association; 2000.
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. *American Psychiatric Association. Dimensional Approaches in Diagnostic Classification Refining the Research Agenda for DSM-V*. Washington, DC: American Psychiatric Association; 2008, pp 101–114.
3. Hudziak JJ, Heath AC, Madden PF et al. Latent class and factor analysis of *DSM-IV* ADHD: a twin study of female adolescents. *J Am Acad Child Adolesc Psychiatry*. 1998;37:848–857.
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: 1584–1593.
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: 1080–1087.
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.
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;45:589–598.
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;43:1018–1028.
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;114:34–41.
10. Todd RD, Sitdhiraksa N, Reich W et al. 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: 820–828.
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;34:1241–1251.
13. Boomsma DI, de Geus EJ, Vink JM et al. Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet*. 2006;9:849–857.
14. Bartels M, van Beijsterveldt CE, 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.
15. Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors with categorical outcomes: class invariance of parameters of factor mixture models. *Multivariate Behav Res*. 2008;43: 592–620.
16. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. New York: John Wiley and Sons; 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.
18. Verhulst FC, van der Ende J, Koot H. *Handleiding voor de CBCL/4-18 [Manual for the CBCL/4-18]*. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie Sophia Kinderziekenhuis, Academisch Ziekenhuis Rotterdam, Erasmus Universiteit Rotterdam; 1996.
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: 227–253.
21. Muthén B, Shedden K. Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*. 1999;55:463–469.
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;78:1706–1722.
23. Neale MC, Aggen SH, Maes HH, Kubarych TS, Schmitt JE. Methodological issues in the assessment of substance use phenotypes. *Addict Behav*. 2006;31:1010–1034.
24. Nagin DS, Land KC. Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric, mixed Poisson model. *Criminology*. 1993;31:327–362.
25. Lubke GH, Neale MC. Distinguishing between latent classes and continuous factors: resolution by maximum likelihood? *Multivariate Behav Res*. 2006;41:499–532.
26. Lubke GH, Muthén BO. Investigating population heterogeneity with factor mixture models. *Psychol Methods*. 2005;10:21–39.
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. *Struct Equation Model*. 2007;14:535–569.
28. Lo Y, Mendell N, Rubin DB. Testing the number of components in a normal mixture. *Biometrika*. 2001;88:767–778.
29. Tueller S, Lubke GH. Evaluation of structural equation mixture models: parameter estimates and correct class assignment. *Struct Equation Model*. 2008;43:592–620.
30. Muthén LK, Muthén B. *Mplus User's Guide*. 4th ed. Los Angeles: Muthen & Muthen; 1998–2006.
31. Asparouhov T. General multi-level modeling with sampling weights. *Commun Stat Theory Methods*. 2006;35:439–460.
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. *J Personality*. 1992;60:117–174.
34. Lubke GH, Tueller S. (submitted). Latent class detection and latent class assignment: a comparison of taxometric and factor mixture modeling approaches.
35. Martin N, Boomsma DI, Machin G. A twin-pronged attack on complex traits. *Nat Genet*. 1997;17:387–391.
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;9:313–324.
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;38:612–653.