

Subtypes Versus Severity Differences in Attention-Deficit/Hyperactivity Disorder in the Northern Finnish Birth Cohort

GITTA H. LUBKE, PH.D., BENGT MUTHÉN, PH.D., IRMA K. MOILANEN, M.D., PH.D.,
 JAMES J. McGOUGH, M.D., SANDRA K. LOO, PH.D., JAMES M. SWANSON, PH.D.,
 MAY H. YANG, M.A., ANJA TAANILA, PH.D., TUULA HURTIG, PH.D.,
 MARJO-RIITTA JÄRVELIN, M.D., AND SUSAN L. SMALLEY, PH.D.

ABSTRACT

Objective: To investigate whether behaviors of inattention, hyperactivity, and impulsivity among adolescents in Northern Finland reflect qualitatively distinct subtypes of ADHD, variants along a single continuum of severity, or of severity differences within subtypes. **Method:** Latent class models, exploratory factor models, and factor mixture models were applied to questionnaire data of ADHD behaviors obtained from the Northern Finland Birth Cohort (NFBC). Latent class models correspond to qualitatively distinct subtypes, factor analysis corresponds to severity differences, and factor mixture analysis allows for both subtypes and severity differences within subtypes. **Results:** A comparison of the different models shows that models that distinguish between a low scoring majority class (unaffecteds) and a high scoring minority class (affecteds), and allow for two factors (inattentive, hyperactive-impulsive) with severity differences provide the best fit. **Conclusions:** The analysis provides support that a high-scoring minority group (8.8% of males and 6.8% of females) likely reflects an ADHD group in the Northern Finland Birth Cohort, whereas the majority of the population falls into a low-scoring group of unaffecteds. Distinct factors composed of items of inattention and hyperactivity-impulsivity are evident for both sexes with considerable variability in severity within each class. *J. Am. Acad. Child Adolesc. Psychiatry*, 2007;46(12):1584–1593. **Key Words:** latent class analysis, factor analysis, factor mixture analysis.

Attention-deficit/hyperactivity disorder (ADHD; American Psychiatric Association, 1994), which is defined by symptoms concerning aspects of inattention, hyperactivity, and impulsivity, is one of the

most common childhood-onset psychiatric disorders, affecting approximately 5% to 10% of children and adolescents (Brown et al., 2001; Faraone and Biederman, 2005). There is an ongoing debate regarding whether phenotypic differences are due to differences in severity of inattentive and hyperactive-impulsive symptoms along an underlying continuum or whether differences reflect qualitatively distinct subtypes with unique etiologies (for a review, see Pickles and Angold, 2003). Studies that focus primarily on the phenotype of ADHD often apply statistical methods that explicitly or implicitly incorporate assumptions concerning the nature of the phenotype of ADHD. Subtypes of ADHD are often defined by using cutoff scores on summed scales drawn from questionnaire data (e.g., Swanson, Nolan, and Pelham questionnaire [SNAP-IV]; Swanson, 1995) or symptoms determined by direct interview and *DSM-IV* (American Psychiatric Association,

Accepted July 3, 2007.

Dr. Lubke is with the University of Notre Dame; Drs. Muthén, McGough, Loo, and Smalley and Ms. Yang are with the University of California, Los Angeles; Drs. Moilanen, Taanila, Järvelin, and Hurtig are with the University of Oulu; Dr. Swanson is with the University of California, Irvine; and Dr. Järvelin is with the Imperial College London.

This research was supported by National Institute of Mental Health grants MH063706 (Smalley, Järvelin) and MH01966 (McGough), and by the Yrjö Johnsson Foundation (Hurtig), the Juselius Foundation, and the Academy of Finland.

Correspondence to Dr. Susan L. Smalley, UCLA Semel Institute, Room 47-438, 760 Westwood Plaza, Los Angeles, CA 90095-1759; e-mail: ssmalley@mednet.ucla.edu.

0890-8567/07/4612-1584©2007 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/chi.0b013e31815750dd

1994). Categorizations based on cutoff values may not be sufficiently fine grained to optimize the definition of the ADHD phenotype.

The present analysis is based on novel statistical developments that allow the comparison of alternative models that correspond to the different conceptualizations of ADHD. We specify models that correspond to a conceptualization in terms of subtypes, models that correspond to a conceptualization in terms of severity differences, and hybrid models that correspond to subtypes but allow for severity differences within subtype. By comparing these alternative models and their fit to collected data, we can make inferences regarding the underlying nature of the ADHD phenotype, as assessed by inattention, hyperactivity, and impulsive behaviors.

The different models used in the present analysis include factor analysis models, latent class models, and factor mixture models. Factor analysis (FA) is based on the assumption differences between individuals are due to differences with respect to one or more continuous underlying traits (i.e., factors). Latent class analysis (LCA) is based on the assumption that differences in symptom endorsements or scores on questionnaire items are due to the fact that individuals belong to different subtypes. The underlying latent class variable is thought to be categorical, and each subtype corresponds to a different latent class. In addition, it is assumed that individuals of a given subtype do not differ systematically; in other words, there are no differences in severity within subtype.

Exploratory factor models and/or latent class models have been used in analyses of ADHD data (Hudziak et al., 2005, 1998, 2000; Neuman et al., 1999; Rasmussen et al., 2002a, 2004, 2002b; Rohde et al., 2001). Applying either one of these two approaches separately precludes the possibility of testing which approach, and hence which set of assumptions, is more adequate to describe the ADHD phenotype. A limitation of the previous work has been the lack of a statistical framework that enables a direct test of factor models versus latent class models. A better fit of latent class models would provide evidence in favor of subtypes, whereas a better fit of factor models would support a continuum. This comparison can be achieved within a general framework that encompasses both types of models.

In the present study, we use a statistical framework called factor mixture modeling (FMM), which combines FA and LCA (Arminger et al., 1999; Dolan and van der Maas, 1996; Heinen, 1996; Jedidi et al., 1997; Muthén and Shedden, 1999; Vermunt and Magidson, 2003; Yung, 1997). FA models and LCA models are submodels of the general FMM. In addition, it is possible to specify hybrid models that have both latent classes and continuous factors within each latent class. This model type would represent a situation in which a population is composed of different subtypes yet in which individuals within a subtype can differ in respect to the severity or symptom patterns of ADHD. The advantage of a general framework is that specific submodels can be compared using well-known measures of goodness of fit (Lubke and Muthén, 2005). A large simulation has shown that, based on measures of goodness of fit, it is generally possible to decide how many classes and/or factors within a class are needed to correctly model the data at hand (Lubke and Neale, 2006; Lubke and Neale, submitted). For overviews and applications of this methodology to diagnostic criteria, see Muthén (2006); Muthén and Asparouhov (2006); Muthén et al. (2006).

In the present study, FMM are applied to ADHD behavioral data collected by a parent questionnaire in the Northern Finland Birth Cohort (NFBC) of originally more than 9,000 adolescents. The current approach differs from multifactorial models of comorbidity described by Neale and Kendler (1995), which are designed to reflect the expected probability of a diagnosis with comorbid disorders under different scenarios. Multifactorial models could be applied to investigate whether, and if so, why diagnoses of inattention and hyperactivity-impulsivity concur. In contrast, the current analysis focuses on the question whether, and if so, why observed inattentive and hyperactive-impulsive behaviors covary. Data on comorbid disorders and subtype diagnosis are only available on a small, nonrandom subset of the NFBC; hence, neither models for diagnoses or comorbidity are covered in this study. Our goal is similar to that of previous analyses (Hudziak et al., 2005, 1998, 2000; Neuman et al., 1999; Rasmussen et al., 2002a, 2004, 2002b; Rohde et al., 2001) in that a clustering of the sample into subtypes is attempted based on ADHD measures alone. The current analysis addresses the question whether there is evidence in favor of clusters

of subjects that differ regarding their response patterns on inattentive and hyperactive-impulsive behaviors. Inclusion of comorbid behaviors would aim at finding clusters reflecting different combinations of ADHD and comorbid behaviors (Neuman et al., 2001, Thapar et al., 2001, Volk et al., 2005).

Due to sex differences in the prevalence of ADHD and potential sex differences regarding subtypes and/or severity, the analyses are conducted for females and males separately. A series of models is fitted to the female and male data including latent class models, exploratory factor models, and models with both classes and factors within class. A comparison of the fit of these models results in choosing a preferred model, which is then tied back to an interpretation regarding the nature of the ADHD phenotype.

METHOD

Subjects

The sample used in the current analysis consists of 6,622 16- to 18-year-olds drawn from the larger NFBC (Järvelin et al., 1993) who completed the 2001–2003 assessment including the postal screening for ADHD and approved of its use in the present study.

Procedure

The Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior scale (SWAN; Swanson et al., 2001b) questionnaire was sent to the families as part of a larger NFBC survey conducted in 2001–2003. Parents were asked to complete the form and informed consent on their children. The study was approved by the Internal Review Board at the University of Oulu in Finland and the Human Subjects Protection Committee at the University of California, Los Angeles (see Smalley et al., 2007).

Measures

The SWAN is a revised version of the SNAP-IV (Swanson, 1995), a behavior rating scale measuring inattentive, hyperactive, and impulsive behaviors. Both the SWAN and SNAP contain the 18 items of the *DSM-IV* diagnostic criteria for ADHD (American Psychiatric Association, 1994). The SNAP-IV is a widely used rating scale that incorporates the *DSM-IV* diagnostic criteria, has norms on large populations, and is sensitive to medication effects. However, it has significant floor effects (i.e., a preponderance of low scores) within the normal population (Swanson et al., 2001a,b), as do most ADHD rating scales used for diagnostic purposes. The SNAP was further developed to create the SWAN in which different wording of the 18 ADHD symptoms and a 7-point (–3 to 3) rating scale was adopted to better capture variation in the normal population (Hay et al., 2007; Swanson et al., 2001a,b). Swanson showed that collapsing the –3 to –1 ratings into the zero category yielded the same distribution of individuals as the SNAP 0–3 rating scale. In the present study, the –3 to –1 ratings are collapsed such that the results can be compared more easily to previous studies

using the SNAP-IV or similar scales that focus on differences between affected and unaffected subjects rather than the variation within the normal population.

Models

We fit three types of models: LCA, FA, and FMM. The LCA model corresponds to a subtype definition of ADHD. Suppose an inattentive subtype exists. Then one would expect that individuals with that subtype have high scores on the inattention items of the SWAN questionnaire while scoring relatively lower on the hyperactivity-impulsivity items. For a hyperactive-impulsive subtype, the expected response pattern would be reversed. Individuals with combined subtype would have high scores on all of the items. If inattentive, hyperactive-impulsive, and combined subtypes exist in addition to unaffected individuals, then an LCA model with four classes should provide a good fit to the observed responses. It is important to note, in case severity differences exist, that it is possible that a model with more than four classes is accepted because additional classes are needed to explain the systematic severity differences (Lubke and Neale, 2006).

FA models are based on the assumption that the population consists of a single type. Suppose that ADHD subtypes do not exist. Then the subjects' observed responses would be expected to differ more with respect to inattention and hyperactivity-impulsivity, and factor models should provide a better fit to the observed responses than LCA models. However, if factor models are fitted to SNAP data and the sample actually consists of more than one subtype, then LCA models should provide the superior fit.

FMMs combine LCA and FA. Similar to LCA, it is possible to fit models with an increasing number of classes. Within each latent class, a factor model with one or more factors is specified that can accommodate systematic differences within a class. This model is therefore a hybrid that allows for subtypes but also for gradual differences in severity within each subtype.

Although LCA and FA models are submodels of the more general FMM, we continue to use the terms LCA model and FA model to allow for an easy distinction. We use the term FMM only for models that have more than one class and at least one factor with nonzero variance within class. Most important, the fit of these different models to the data can be directly compared within this general framework.

Analysis

The prevalence of ADHD is known to be higher in males than in females (American Psychiatric Association, 1994), and in case of subtypes, females and males may differ with respect to the clustering structure. To avoid confounding, the sexes are analyzed separately. The strategy of the present analysis is to evaluate the fit of alternative models by goodness-of-fit indices including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the sample size adjusted Bayesian Information Criterion, which are designed to facilitate the decision as to which model provides a better fit to observed responses. More complex models such as models with more classes have more parameters and therefore can fit the observed data more easily. Fit indices such as the AIC and BIC give penalties to models with more parameters and therefore help protect against unnecessary model complexity.

TABLE 1
Model Comparisons, Fit Indices, and Class Proportions for Females in the NFBC

| Females | No. of Classes (c) or Factors (f) | Log Likelihood | No. of Estimated Parameters | AIC | BIC | Adjusted BIC | Class Proportions |
|-----------|-----------------------------------|----------------|-----------------------------|---------|---------|--------------|---|
| LCA model | 2 c | -11567.116 | 73 | 23280.2 | 23725.8 | 23493.9 | 0.851, 0.129 |
| | 3 c | -11033.243 | 110 | 22286.5 | 22957.9 | 22608.4 | 0.785, 0.184, 0.031 |
| | 4 c | -10883.739 | 147 | 22061.5 | 22958.8 | 22491.7 | 0.777, 0.166, 0.047, 0.010 |
| | 5 c | -10750.224 | 184 | 21868.5 | 22991.6 | 22407.0 | 0.774, 0.130, 0.056, 0.032, 0.009 |
| | 6 c | -10668.427 | 221 | 21778.9 | 23127.9 | 22425.6 | 0.773, 0.122, 0.052, 0.028, 0.018, 0.008 |
| | 7 c | -10618.557 | 258 | 21753.1 | 23328.0 | 22508.2 | 0.775, 0.105, 0.050, 0.040, 0.020, 0.008, 0.005 |
| EFA model | 1 f | 11024.375 | 54 | 22156.8 | 22486.4 | 22314.8 | |
| | 2 f | 10754.024 | 71 | 21650.1 | 22083.4 | 21857.8 | |
| | 3 f | 10688.792 | 87 | 21551.6 | 22082.6 | 21806.2 | |
| FMM | 2 f 1 c | 10787.920 | 55 | 21685.8 | 22021.6 | 21846.8 | |
| | 2 f 2 c | 10680.112 | 95 | 21550.2 | 22130.1 | 21828.3 | 0.932, 0.068 |
| | 2 f 3 c | 10611.565 | 135 | 21493.1 | 22317.2 | 21888.2 | 0.748, 0.210, 0.042 |

Note: The 2f 2c and the 2f 3c FMM are clearly superior to all of the other fitted models. The 2f 2c FMM is the more parsimonious than the 2f 3c FMM. NFBC = Northern Finland Birth Cohort; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LCA = latent class analysis, EFA = exploratory factor analysis; FMM = factor mixture model.

Because the different models correspond to different assumptions about the structure of the phenotype of ADHD (e.g., FA models correspond to continuous liability, LCA models to subtypes, and FMM to a combination of subtypes and variation within subtype), the decision as to which type of model provides the most adequate fit can support corresponding conclusions regarding how best to think of the phenotype of ADHD.

All of the analyses were carried out using the Mplus computer program (Muthén and Muthén, 1998–2006). Twelve models

were fitted to females and males as shown in Tables 1 and 2. Models 2c to 7c are LCA models with two to seven classes, respectively; models 1f to 3f are exploratory FA models with one, two, or three factors; and 2f1c to 2f3c are FMMs with one to three classes and two factors within a class, respectively. The FMMs have two factors (inattentive and hyperactive-impulsive) within each class, and the nine inattentive items and the nine hyperactive-impulsive nine items of the SWAN are allowed to load only on their corresponding factor (i.e., simple structure).

TABLE 2
Model Comparisons, Fit Indices, and Class Proportions for Males in the NFBC

| Males | No. of Classes | Log Likelihood | No. of Estimated Parameters | AIC | BIC | Adjusted BIC | Class Proportions |
|-----------|----------------|----------------|-----------------------------|---------|---------|--------------|---|
| LCA model | 2 c | -16352.978 | 73 | 32852.0 | 33297.7 | 33065.7 | 0.806, 0.194 |
| | 3 c | -15455.019 | 110 | 31130.0 | 31801.7 | 31452.2 | 0.714, 0.232, 0.053 |
| | 4 c | -15227.813 | 147 | 30749.6 | 31647.2 | 31180.1 | 0.726, 0.172, 0.055, 0.047 |
| | 5 c | -15038.945 | 184 | 30445.9 | 31569.4 | 30984.7 | 0.687, 0.195, 0.066, 0.034, 0.019 |
| | 6 c | Not converged | 221 | | | | |
| | 7 c | -14813.933 | 258 | 30143.9 | 31719.2 | 30899.4 | 0.670, 0.141, 0.083, 0.053, 0.027, 0.019, 0.008 |
| EFA model | 1 f | -15395.865 | 54 | 30899.7 | 31229.5 | 31057.9 | |
| | 2 f | -14979.093 | 71 | 30100.2 | 30533.1 | 30308.1 | |
| | 3 f | -14898.652 | 87 | 29971.3 | 30502.5 | 30226.1 | |
| FMM | 2 f 1 c | -15026.736 | 55 | 30163.5 | 30499.3 | 30324.5 | |
| | 2 f 2 c | -14820.454 | 95 | 29830.9 | 30411.0 | 30109.1 | 0.913, 0.087 |
| | 2 f 3 c | -14719.286 | 135 | 29708.6 | 30532.9 | 30103.9 | 0.663, 0.272, 0.065 |

Note: As for the females, the 2f 2c and the 2f 3c FMM are clearly superior to all of the other fitted models. NFBC = Northern Finland Birth Cohort; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; LCA = latent class analysis, EFA = exploratory factor analysis; FMM = factor mixture model.

RESULTS

Goodness-of-fit measures shown in Table 1 (for females) and Table 2 (for males) illustrate the results of fitting the different models. Lower values of AIC and BIC indicate better fit. General results regarding the model comparison and the choice of the preferred model are similar for the two sexes. Because previous analyses have not directly compared LCA and FA models, we organize the discussion of each table by first comparing models within a given model type (e.g., LCA, FA, FMM) to show what the conclusions would have been had the analyses been limited on one type of model. This is followed by an overall comparison across the three model types.

Latent Class Models (Models 2c–7c)

When restricting the comparison to the six fitted latent class models, it can be seen in Tables 1 and 2 that models with a higher number of classes have higher log likelihood values. This is expected because models with more classes have more parameters and can therefore provide a better fit to the data. The information criteria AIC, BIC, and the sample size adjusted BIC have different penalties for an increase in model leniency as measured by the number of estimated parameters and therefore do not necessarily point to the same model as a best fitting model. Among the LCA models, the BIC favors the five-class model in males and a three-class model in the females, whereas the sample size adjusted BIC points to six classes in the females. The AIC is lowest for the seven class solutions for both sexes. Limiting the analyses to LCA, a researcher would proceed to interpret six or seven clusters.

Considering the FA models separately, all of the information criteria point to the three-factor model; however, the third factor in the three-factor model has no clear interpretation. No specific cluster of items loads on this factor, rather, almost all items have minor loadings. When comparing the three- and two-factor models, the two-factor model would be chosen because only two eigenvalues are larger than 1. Furthermore, the inattentive and hyperactive-impulsive behaviors load separately on the two factors, consistent with the design of the SWAN instrument. The conclusion of fitting only FA models would be that there are two correlated continua corresponding to inattentiveness and hyperactivity-impulsivity.

Among the FMMs, the two-class and three-class two-factor models provide a clearly better fit than the single class two-factor model. The AIC favors the model with three classes, whereas the BIC points to the two-class model in both females and males. The sample size adjusted BIC is smaller for the two-class model in the females and slightly smaller for the three-class model in the males.

An overall comparison of all of the models results in a clear preference for the FMMs over either the LCA or exploratory FA models. This result is the same for both sexes. As indicated in the tables, information criteria show that the FMMs provide a superior fit to the data than any of the LCA or exploratory FA models.

As shown in Table 1, the two-class, two-factor FMM for females is the best fitting model. The two classes differ in that the larger class (93.2% of the female sample) scores on average lower on all 18 SWAN items than the smaller class (6.8% of the female sample). For males, the three-class and the two-class model provide a similar fit (Table 2); however, the substantial increase in parameters of the three-class model argues for the simpler two-class, two-factor FMM solution. Under this model, the low-scoring majority class comprises 0.913 of the male population and 0.088 fall in the high-scoring minority class. The male/female ratio in the high scoring (ADHD) class is 0.087:0.068 (ratio = 1.3:1). The structure of the best fitting model is depicted in Figure 1.

The odds of scoring in the highest response category of the 18 recoded SWAN questionnaire items for females and males are shown in Figure 2. It is important to note that the odds pertaining to the inattentiveness items and the odds pertaining to the hyperactivity items are both elevated in the minority class. This shows that the majority and minority classes differ quantitatively rather than qualitatively. Even under the three-class, two-factor solution, there is no clear evidence in favor of qualitatively distinct subtypes. Under the three-class, two-factor solution (which does not fit as well as the two-class, two-factor solution when all of the criteria are considered), the third class is merely adding a quantitatively intermediate between the high- and the low-scoring classes. We see no support of qualitatively distinct ADHD subtypes in these analyses. Specifically, an inattentive subtype would be evident if a latent class showed elevated odds on only the inattentive items, and a hyperactive subtype would be evident if a latent class

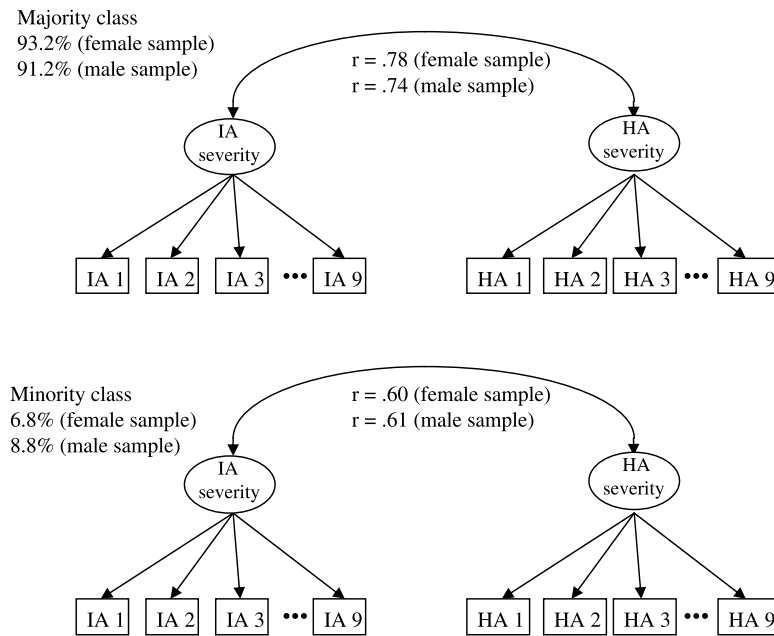


Fig. 1 The two-class two-factor model which provided the best fit to the data for males and females. The rectangular boxes (IA 1 to IA 9 and HA 1 to HA 9) represent the 18 observed Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior scale (SWAN) behaviors, the ovals represent underlying severity factors. For both males and females the correlation between the IA and HA severity factors is strong in the low-scoring majority class and moderate in the high-scoring minority class. IA = inattentiveness; HA = hyperactivity-impulsiveness.

showed elevated odds on only the hyperactivity items. No such distinct classes are observed in the present analysis.

It may be argued that factor mixture analysis is not sufficiently sensitive to detect small classes reflecting qualitatively different subtypes. The high-scoring

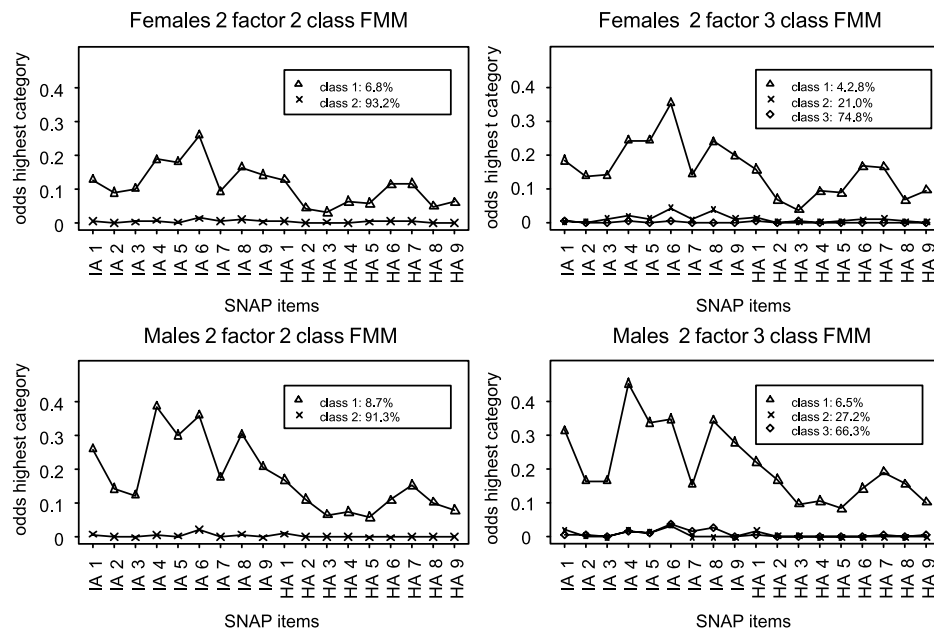


Fig. 2 The probability of scoring in the highest response category of the 18 SNAP items as compared to scoring in any of the other categories for females and males in the two- and three-class factor mixture model. The minority class in all four models (males, females in the two- and three-class models) has elevated probabilities of scoring in the highest response categories of all 18 items. A specific subtype (e.g., inattentive) would be apparent if a class has elevated probabilities only for the corresponding items (e.g., IA items). IA = inattentiveness; HA = hyperactivity-impulsiveness; FMM = factor mixture model.

minority class in our analysis may comprise individuals who would be diagnosed as belonging to the inattentive, hyperactive-impulsive, or combined subtypes. Subtypes are commonly derived by applying cutoff criteria to sum scores computed from symptom endorsements or item responses, which is an arbitrary categorization. The current analysis shows considerable variability on two underlying dimensions represented by the inattentiveness and hyperactivity factors within both classes. The correlation of the inattentiveness and hyperactivity factors is approximately 0.6 in the high-scoring class for both males and females (Fig. 1). This correlation is lower than in the unaffected majority class where the correlation is around 0.75. It is therefore possible that the high-scoring class contains individuals who score low on hyperactivity and high on inattentiveness (i.e., inattentive subtype), or vice versa (i.e., hyperactive subtype), or high on both (i.e., combined subtype). In other words, the extremes of the two continuous dimensions may actually comprise the subjects who would be categorized as belonging to one of the subtypes if sum score cutoff criteria were used.

Because no data on *DSM-IV* subtype diagnosis are available for the full data set, this issue cannot be addressed directly. Instead, we conduct the following post hoc analysis. We categorize inattentiveness and hyperactivity scores to evaluate whether our results match findings concerning the prevalence of subtypes. Table 3 shows a cross-tabulation of females and males in the high-scoring class (225 females, 290 males) after categorizing their inattentiveness and hyperactivity scores in four equidistant intervals. (Note that categorization in four intervals is an arbitrary choice. More intervals would provide a more detailed description of the distribution of factor scores but would not change the conclusions presented here.) A hyperactive-impulsive subtype would be evident in case of a disproportional number of subjects with high scores on the hyperactive and simultaneously low scores on the inattentiveness (see ^acells in Table 3).

Similarly, Inattentive and Combined subtypes would be evident in case of a clustering of subjects in the cells with superscript *b* and *c*, respectively. In addition, there should be no or only few subjects in the cells without superscripts. This, however, is not the pattern that resulted from the 2f 2c FMM for males and females that provided the best fit to the

TABLE 3

Cross-Tabulation of Factor Scores for Females (*n* = 225) and Males (*n* = 290) in the High-Scoring Class

| | HA 1 | HA 2 | HA 3 | HA 4 |
|---------|-----------------|-----------------|-----------------|----------------|
| Females | | | | |
| IA 1 | 7 | 8 | 0 ^a | 0 ^a |
| IA 2 | 63 | 47 | 12 ^a | 0 ^a |
| IA 3 | 20 ^b | 29 ^b | 16 ^c | 1 ^c |
| IA 4 | 2 ^b | 11 ^b | 4 ^c | 5 ^c |
| Males | | | | |
| IA 1 | 16 | 12 | 3 ^a | 0 ^a |
| IA 2 | 58 | 57 | 28 ^a | 0 ^a |
| IA 3 | 12 ^b | 47 ^b | 23 ^c | 3 ^c |
| IA 4 | 1 ^b | 10 ^b | 11 ^c | 9 ^c |

Note: The cross-tabulation is derived by categorizing the inattentiveness (IA) and hyperactivity-impulsiveness (HA) factors into four equally wide categories. Although males have higher proportions in cells with superscripts compared to females, there is no disproportional clustering of subjects in those cells that would indicate the existence of subtypes.

^a Cells that reflect high scores on the hyperactive-impulsive factor and low scores on the inattentive factor representative of the Hyperactive-Impulsive subtype.

^b Cells that reflect high scores on inattentiveness and low scores on hyperactivity-impulsivity representative of the Inattentive subtype.

^c Cells that reflect high scores on both hyperactive-impulsive and inattentive factors representative of the Combined subtype.

NFBC data. Rather, the subjects are more equally distributed over all of the cells. Specifically, there are some individuals with high scores on the inattentive factor with simultaneous low scores on the hyperactivity factor. The reverse is much less likely. This is consistent with the generally lower incidence of the Hyperactive subtype. Few individuals have high scores on both inattentiveness and hyperactivity. The main proportion of individuals in the high-scoring class has intermediate scores on inattentiveness and lower scores on hyperactivity. This pattern of results is similar for females and males (Table 3) and provides evidence of two correlated underlying continua.

DISCUSSION

The statistical approach used in the present study shows that FMMs provide the best fit to the behavioral measures of inattention, hyperactivity, and impulsivity derived from a parent questionnaire obtained on

adolescents in the NFBC. These best fitting models differentiate between the unaffected majority of the cohort and the potentially affected minority, and, in addition, allow for systematic variation in severity. LCA is based on the assumption that individual differences in the observed responses are due to the existence of subtypes, and applying only LCA can therefore not provide evidence in favor of or against the existence of subtypes. A similar argument holds for FA, which is based on the assumption that observed differences are due to gradual severity differences. Only by comparing models that correspond to these competing hypotheses about the ADHD phenotype can we decide which hypothesis makes most sense given our observed responses on the SWAN questionnaire. In our analysis, the hybrid model that separated the unaffected majority from a potentially affected minority while allowing for severity differences in both inattentiveness and hyperactivity-impulsivity provided the best fit to the SWAN data. It would be extremely valuable to apply our approach to previously studied ADHD data sets that used either LCA and extracted multiple classes or FA. Replication of our results would validate the application of FMMs to ADHD behavioral data and perhaps simplify our current understanding of ADHD and putative subtypes in other populations.

The NFBC is a population sample, and the division in a low-scoring majority and a high-scoring minority class is not surprising. The prevalence of the high-scoring minority group in the current analysis by sex (8.8% of males, 6.8% of females) differs from the sex prevalences based on *DSM-IV* criteria in the cohort using *DSM-IV* criteria (see Smalley et al., 2007, this issue). One explanation for these differences is that diagnoses based on *DSM-IV* require additional components beyond the symptom elevation, specifically, age at onset and impairment in two settings. This is an important distinction because the SWAN data may be better viewed as a measure of trait liability to ADHD (based on behaviors), whereas the *DSM-IV* diagnoses may be a better reflection of who needs clinical services. Future work investigating the sensitivity and specificity of SWAN and *DSM-IV* in this sample along with other variables of interest (e.g., comorbidity) are needed to refine the ADHD phenotype and to further the distinction between trait liability and disorder. Based on the current analyses and those in the accompanying article using a smaller subset of the data (Smalley et al.,

2007) combining SWAN and *DSM-IV* in examining prevalences by sex, the population distributions suggest that males and females have similar prevalences in risk, but that males may have more associated impairment or features that result in clinical diagnoses.

The importance of our results concern the fact that the best fitting models support two continuous factors representing severity of inattentiveness and hyperactivity with considerable variability on both dimensions. Perhaps most important is that there is no strong indication for a need to identify subgroups within ADHD in this particular sample. The minority potentially affected class has variability within it along inattentive and hyperactive-impulsive dimensions, but our analyses did not indicate that further subgrouping is necessary in the NFBC.

In our analysis, inattentiveness and hyperactivity are positively correlated within both groups. Note that this correlation does not exclude the possibility that individuals have opposite scores on the two dimensions, thus reflecting mainly inattentive symptoms in the absence of hyperactive symptoms (or vice versa). If the two continua are categorized, then individuals can be assigned to different subtypes. Although such an assignment may be useful in practice (e.g., choice of treatment), based on behaviors of ADHD, there is no indication that categorizing into subgroups would increase homogeneity; rather, it is important to realize that it is derived by categorizing what appear to be two continuous, moderately correlated traits. In the NFBC sample, most of the high-scoring minority have moderately high inattentive symptoms and low to moderate hyperactive-impulsive symptoms. Few individuals score in the severe inattentive range, which may reflect cultural differences in symptom endorsement or qualitative differences in severity. Further work is needed to better understand such possible cross-cultural differences in population distributions based on the SWAN.

The present analysis focuses on defining the ADHD phenotype in a large population sample of adolescents. Factor mixture analysis can easily integrate additional information concerning comorbid behaviors. Analyzing ADHD data and data concerning comorbid conditions jointly would provide information whether certain combinations of a comorbid condition and inattentiveness and/or hyperactivity-impulsivity are more prevalent than others.

Limitations

There are several limitations to the present study. First, the conclusions drawn from the FMM analyses are dependent on the accuracy of the SWAN scale to reflect ADHD risk. Although the SWAN appears sensitive to detect ADHD cases, there were many SWAN-identified cases that did not meet *DSM-IV* criteria (Smalley et al., 2007), suggesting that results based on SWAN scores may differ somewhat from those based on *DSM-IV* criteria. Because the behavioral criteria for ADHD may be somewhat removed from the underlying etiology, conclusions as to underlying subgroups based on the behavioral criteria may also be somewhat removed from etiology. Second, the analyses are based on current ratings by parents of ADHD behaviors, parental ratings may be less accurate in adolescence than in childhood, and replication with both self and/or teacher ratings may yield a different picture. Third, conclusions based on adolescent symptom patterns may differ from those at other times in the life span, so additional analyses in other populations and other age groups are warranted. It is possible, for instance, that subtypes are distinguishable in latency-age individuals. Fourth, the NFBC is an ethnically and geographically restricted sample, which may limit the generalizability of our results to other populations. Fifth, statistical analyses such as LCA, FA, and FMM are based on assumptions regarding the distribution of the data and depend on power. Results should be interpreted accordingly (Lubke and Neale, 2006).

Clinical Implications

The analysis supports the hypothesis that ADHD may be seen along two moderately correlated continua (inattentiveness, hyperactivity-impulsivity) rather than clusters of distinct subgroups. The findings, if replicated in additional samples, support the idea that, at least in adolescence, ADHD may be seen as a single extreme along a continuum in the population, with differences along both dimensions of inattention and hyperactivity-impulsivity. This has important implications for future research and clinical work. In case our results are validated, recognition of ADHD as a trait in the population that in the extreme may be associated with impairment would have important clinical ramifications. Most important, individuals without a diagnosis may fall high on a liability spectrum and may benefit from preventive tools. Second,

although impairment warrants treatment, recognition that the predisposition to ADHD is a continuous trait, like height and weight, may help destigmatize the condition because it may be seen as another dimension of human brain and behavioral diversity.

Disclosure: Dr. Moilanen is a member of the Eli Lilly Strattera Advisory Board, Finland. Dr. McGough receives grant research support from Eli Lilly, McNeil, New River Pharmaceuticals, Novartis, Shire, and Pfizer and is also a consultant to Eli Lilly, Novartis, and Shire and serves on the speakers' bureaus of Eli Lilly, McNeil, Novartis, and Shire. Dr. Swanson has received research support and honoraria from, served on the speakers' bureaus of, and served as a consultant or advisor to Alza, McNeil, Janssen, UCB, Cephalon, Eli Lilly, Novartis, and Shire U.S. The other authors have no financial relationships to disclose.

REFERENCES

- American Psychiatric Association (1994), *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV)*. Washington, DC: American Psychiatric Association
- Arminger G, Stein P, Wittenberg J (1999), Mixtures of conditional mean and covariance structure models. *Psychometrika* 64:475–494
- Brown RT, Freeman WS, Perrin JM et al. (2001), Prevalence and assessment of attention-deficit/hyperactivity disorder in primary care settings. *Pediatrics* 107:E43
- Dolan CV, van der Maas HLJ (1996), Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika* 63:227–253
- Faraone SV, Biederman J (2005), What is the prevalence of adult ADHD? Results of a population screen of 966 adults. *J Atten Disord* 9:384–391
- Hay DA, Bennett KS, Levy F, Sergeant J, Swanson J (2007), A twin study of attention-deficit/hyperactivity disorder dimensions rated by the Strengths and Weaknesses of ADHD-Symptoms and Normal-Behavior (SWAN) Scale. *Biol Psychiatry* 61:700–705
- Heinen T (1996), *Latent Class and Discrete Latent Trait Models: Similarities and Differences*. Thousand Oaks, CA: Sage Publications
- Hudziak JJ, Derks EM, Althoff RR, Rettew DC, Boomsma DI (2005), The genetic and environmental contributions to attention deficit hyperactivity disorder as measured by the Conners' Rating Scales-Revised. *Am J Psychiatry* 162:1614–1620
- Hudziak JJ, Heath AC, Madden PF et al. (1998), Latent class and factor analysis of *DSM-IV* ADHD: a twin study of female adolescents. *J Am Acad Child Adolesc Psychiatry* 37:848–857
- Hudziak JJ, Rudiger LP, Neale MC, Heath AC, Todd RD (2000), A twin study of inattentive, aggressive, and anxious/depressed behaviors. *J Am Acad Child Adolesc Psychiatry* 39:469–476
- Järvelin MR, Hartikainen-Sorri AL, Rantakallio P (1993), Labour induction policy in hospitals of different levels of specialisation. *Br J Obstet Gynaecol* 100:310–315
- Jedidi K, Jagpal HS, DeSarbo WS (1997), Finite mixture structural equation models for response based segmentation and unobserved heterogeneity. *Marketing Sci* 16:39–59
- Lubke GH, Muthén B (2005), Investigating population heterogeneity with factor mixture models. *Psychol Methods* 10:21–39
- Lubke GH, Neale MC (2006), Distinguishing between latent classes and continuous factors: resolution by maximum likelihood? *Multivariate Behav Res* 41:499–532
- Muthén B, Asparouhov T (2006a), Item response mixture modeling: application to tobacco dependence criteria. *Addict Behav* 31:1050–1066
- Muthén B (2006b), Should substance use disorders be considered as categorical or dimensional? *Addiction* 101:6–16
- Muthén B, Asparouhov T, Rebollo I (2006), Advances in behavioral genetics

- modeling using Mplus: applications of factor mixture modeling to twin data. *Twin Res Hum Genet* 9:313–324
- Muthén B, Shedden K (1999), Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics* 55:463–469
- Muthén LK, Muthén B (1998–2006), *Mplus User's Guide*, 4th ed. Los Angeles: Muthen & Muthen
- Neale MC, Kendler KS (1995), Models of comorbidity for multifactorial disorders. *Am J Hum Genet* 57:935–953
- Neuman RJ, Heath AC, Reich W et al. (2001), Latent class analysis of ADHD and comorbid symptoms in a population sample of adolescent female twins. *J Child Psychol Psychiatry* 42:933–942
- Neuman RJ, Todd RD, Heath AC et al. (1999), Evaluation of ADHD typology in three contrasting samples: a latent class approach. *J Am Acad Child Adolesc Psychiatry* 38:25–33
- Pickles A, Angold A (2003), Natural categories or fundamental dimensions: on carving nature at the joints and the rearticulation of psychopathology. *Dev Psychopathol* 15:529–551
- Rasmussen ER, Neuman RJ, Heath AC, Levy F, Hay DA, Todd RD (2002a), Replication of the latent class structure of attention-deficit/hyperactivity disorder (ADHD) subtypes in a sample of Australian twins. *J Child Psychol Psychiatry* 43:1018–1028
- Rasmussen ER, Neuman RJ, Heath AC, Levy F, Hay DA, Todd RD (2004), Familial clustering of latent class and DSM-IV defined attention-deficit/hyperactivity disorder (ADHD) subtypes. *J Child Psychol Psychiatry* 45:589–598
- Rasmussen ER, Todd RD, Neuman RJ, Heath AC, Reich W, Rohde LA (2002b), Comparison of male adolescent-report of attention-deficit/hyperactivity disorder (ADHD) symptoms across two cultures using latent class and principal components analysis. *J Child Psychol Psychiatry* 43:797–805
- Rohde LA, Barbosa G, Polanczyk G et al. (2001), Factor and latent class analysis of DSM-IV ADHD symptoms in a school sample of Brazilian adolescents. *J Am Acad Child Adolesc Psychiatry* 40:711–718
- Smalley SL, McGough JJ, Moilanen IK et al. (2007), Prevalence and psychiatric comorbidity of attention-deficit/hyperactivity disorder in an adolescent Finnish population. *J Am Acad Child Adolesc Psychiatry* 46:1575–1583
- Swanson JM (1995), *SNAP-IV Scale*. Irvine: University of California Child Development Center
- Swanson JM, Schuck S, Mann M et al. (2001a), Over-identification of extreme behavior in the evaluation and diagnosis of ADHD/HKD. Available at: <http://www.adhd.net>. Accessed June 6, 2006
- Swanson JM, Schuck S, Mann M et al. (2001b), Categorical and dimensional definitions and evaluations of symptoms of ADHD: the SNAP and SWAN ratings scales. Available at: <http://www.adhd.net>. Accessed June 6, 2006
- Thapar A, Harrington R, McGuffin P (2001), Examining the comorbidity of ADHD related behaviours and conduct problems using a twin study design. *Br J Psychiatry* 179:224–229
- Vermunt JK, Magidson J (2003), Latent class models for classification. *Comput Stat Data Anal* 41:531–537
- Volk HE, Neuman RJ, Todd RD (2005), A systematic evaluation of ADHD and comorbid psychopathology in a population-based twin sample. *J Am Acad Child Adolesc Psychiatry* 44:768–775
- Yung YF (1997), Finite mixtures in confirmatory factor analysis models. *Psychometrika* 62:297–330

Efficacy of Zolmitriptan Nasal Spray in Adolescent Migraine Donald W. Lewis, MD, FAAN, FAAP, Paul Winner, DO, FAAN, FAAP, Andrew D. Hershey, MD, PhD, Warren W. Wasiewski, MD, on behalf of the Adolescent Migraine Steering Committee

Objective: The goal was to evaluate the efficacy and tolerability of zolmitriptan nasal spray in the treatment of adolescent migraine. **Methods:** The “Double-Diamond” study used a novel, single-blind, “placebo challenge” in a multicenter, randomized, double-blind, placebo-controlled, 2-way, 2-attack, crossover design. A total of 248 U.S. adolescent patients (12–17 years of age) with an established diagnosis of migraine, with or without aura, were enrolled. A single-blind placebo challenge was used for each migraine attack. No additional medications were taken if a headache response to the initial placebo treatment was achieved at 15 minutes; if migraine intensity remained moderate or severe, then patients treated the attack with zolmitriptan (5 mg) nasal spray or placebo according to a randomized, crossover schedule (double-blind). The primary efficacy variable was headache response at 1 hour after treatment. A comprehensive range of secondary end points included sustained headache response at 2 hours. **Results:** A total of 171 patients (mean age: 14.2 years; 57.3% female) treated ≥ 1 attack with study medication (intention-to-treat population). The onset of significant pain relief was apparent 15 minutes after treatment with zolmitriptan nasal spray. At 1 hour after the dose, zolmitriptan nasal spray produced a higher headache response rate than did placebo (58.1% vs 43.3%). Zolmitriptan nasal spray was also significantly superior to placebo in improvement in pain intensity, pain-free rates, sustained resolution of headache, and resolution of associated migraine symptoms. Return to normal activities was also consistently faster with zolmitriptan nasal spray than with placebo, with less use of any escape medication. Treatment with zolmitriptan nasal spray was well tolerated. **Conclusions:** This novel, placebo-challenge study demonstrated that zolmitriptan nasal spray was well tolerated and provided fast and significantly effective relief of migraine symptoms in the acute treatment of adolescent migraine. **Pediatrics** 2007;120:390–396.