Joint developmental trajectories of internalizing and externalizing disorders between childhood and adolescence

MICHEL G. NIVARD, GITTA H. LUBKE, CONOR V. DOLAN, DAVID M. EVANS, BEATE ST. POURCAIN, MARCUS R. MUNAFÒ, AND CHRISTEL M. MIDDELDORP

Vrije Universiteit; University of Notre Dame; University of Queensland Diamantina Institute; University of Bristol; Max Planck Institute for Psycholinguistics; and VU University Medical Centre

Abstract

This study sought to identify joint trajectories of DSM-IV based internalizing (INT) and externalizing (EXT) problem scores across childhood and adolescence and to provide insight into the comorbidity by modeling the co-occurrence of INT and EXT trajectories. INT and EXT were measured repeatedly between age 7 and age 15 years in over 7,000 children and analyzed using growth mixture models. Five trajectories were identified for both INT and EXT, including very low, low, decreasing, and increasing trajectories. In addition, an adolescent onset trajectory was identified for INT and a stable high trajectory was identified for EXT. Multinomial regression showed that similar EXT and INT trajectories were associated. However, the adolescent onset INT trajectory was independent of high EXT trajectories, and persisting EXT was mainly associated with decreasing INT. Sex and early life environmental risk factors predicted EXT and, to a lesser extent, INT trajectories. The association between trajectories indicates the need to consider comorbidity when a child presents with INT or EXT disorders, particularly when symptoms start early. This is less necessary when INT symptoms start at adolescence. Future studies should investigate the etiology of co-occurring INT and EXT and the specific treatment needs of these severely affected children.

Longitudinal epidemiological studies have shown that the prevalence of joint internalizing (INT) disorders (anxiety and depressive disorders) increases from childhood to adulthood, whereas the prevalence of externalizing (EXT) disorders (attention-deficit/hyperactivity disorder [ADHD], conduct disorder [CD], and oppositional defiant disorder [ODD]) decreases (Costello, Copeland, & Angold, 2011). Research into the stability has shown that symptoms and disorders persist in varying degrees from childhood into adulthood (Hofstra, Van der Ende, & Verhulst, 2000, 2002). Moreover, comorbidity is common (Angold, Costello, & Erkanli, 1999), not only concurrently, but also successively, as has been established in longitudinal studies: INT disorders in childhood predict EXT disorders in adulthood and vice versa (Copeland, Shanahan, Costello, & Angold, 2009; Hofstra et al., 2000, 2002). It is well known that comorbidity is associated with poorer outcomes (Newman, Moffitt, Caspi, & Silva, 1998).

It is important to distinguish children who are at risk for developing chronic and/or comorbid symptoms and children whose symptoms are transient so that treatment programs can be developed that specifically target children at risk of long-term psychopathology or children whose symptoms are transient. Therefore, it is necessary to investigate the existence of subgroups (classes) with distinct developmental trajectories using growth mixture modeling (GMM). Few studies using GMM have analyzed data on INT and EXT problems in the period from childhood to adolescence in population-based cohorts (Dekker et al., 2007; Haltigan, Roisman, Susman, Barnett-Walker, & Monahan, 2011; Larsson, Dilshad, Lichtenstein, & Barker, 2011; Letcher, Smart, Sanson, & Toumbourou, 2009; Toumbourou, Williams, Letcher, Sanson, & Smart, 2011; Van Lier, Der Ende, Koot, & Verhulst, 2007). To the best of our knowledge, no study has investigated comorbidity in trajectories from childhood into adolescence. As pointed out by Angold et al. (1999), research in population-based cohorts is necessary to get unbiased estimates of comorbidity and its risk factors. Because the largest changes in prevalence rates are observed in the transition from childhood to adolescence, it is critical to investigate this period.

In the current study, we used GMM to model the development of DSM-IV based INT and EXT problem scores measured at four occasions between age 7 and 15 years in a birth cohort of over 7,000 children. INT comprised anxiety...
disorders and depression; and EXT comprised ADHD, ODD, and CD. The use of INT and EXT summary scores is consistent with the results of several factor analytic studies of these disorders, which revealed the presence of INT and EXT higher order factors explaining the covariance between individual disorders (Angold et al., 1999; Beauchaine & McNulty, 2013; Cosgrove et al., 2011). Following an initial separate analysis of INT and EXT trajectories, we focused on the co-occurrence of these trajectories in a combined model. We further added well-known risk factors for INT and EXT psychopathology to the model as predictors of class membership such as sex, birth weight, maternal smoking during pregnancy, and social class (Costello, Compton, Keeler, & Angold, 2003; Dolan et al., 2015; Groen-Blokhuizen, Middeldorp, van Beijsterveldt, & Boomsma, 2011; Hack et al., 2004; Linnet et al. 2003; Weissman, Warner, Wickramaratne, & Kandel, 1999). An added benefit of modeling early childhood risk factors as predictors of later trajectories in a longitudinal study is that the results are not affected by participant dropout associated with these childhood risk factors (Little & Rubin, 2014). The results provide insight into the trajectories of clinically relevant INT and EXT problems across childhood and adolescence as well as into the association between the INT and EXT trajectories in this period.

Based on the results of previous trajectory analyses of INT and EXT psychopathology measured during childhood and adolescence in population-based cohorts (Dekker et al., 2007; Haltigan et al., 2011; Larsson et al., 2011; Letcher et al., 2009; Toubourou et al., 2011; Van Lier et al., 2007), we expected for both INT and EXT a class of unaffected individuals. In addition, we expected at least a class with increasing symptoms for INT and a class with stable high and a class with decreasing symptoms for EXT. Although results of previous studies using GMM on INT symptoms are mixed regarding a class with persisting symptoms over time, we expected such a class given that other longitudinal studies suggest continuity over age (Hofstra et al., 2000, 2002). To the best of our knowledge, the current study is the first to investigate the co-occurrence of trajectories of INT and EXT problems from childhood to adolescence. Three previous studies investigated the concordance between INT and EXT trajectories in children up to age 12 (Brezo et al., 2008; Fantl & Henrich, 2010; Wiggins, Mitchell, Hyde & Monk, 2015). These studies suggested that children assigned to trajectories with high scores on INT problems were significantly more often assigned to trajectories with moderate or high scores for EXT problems, and vice versa. This signifies that the course of INT and EXT symptoms is also associated during childhood. We expect that association to continue into adolescence.

Methods

Subjects

The Avon Longitudinal Study of Parents and Children (ALSPAC, also known as Children of the 90s; http://www.bris.ac.uk/alspac/) is a long-term health research project (Boyd et al. 2012). More than 14,000 mothers from Avon County in the United Kingdom were enrolled during pregnancy in 1991 and 1992, and each returned at least one questionnaire. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally (Golding, Pembrey, & Jones, 2001).

The (psychological) health and development of these children has been followed in great detail. At ages 7, 10, 13, and 15 years, DSM-IV psychiatric disorders were assessed as part of the regular assessments. In total, 7,202 children were assessed at least once for psychiatric disorders and had data available on risk factors (see Table 1). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the local research ethics committees. Please note that the study website contains details of all the data available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary).

Instruments

The Development and Well-Being Assessment (DAWBA) is an instrument developed to diagnose DSM-IV psychiatric disorders (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). In addition to a dichotomous variable indicating whether or not a person satisfies the criteria for a diagnosis, the DAWBA instrument can be used to calculate an ordinal “DAWBA band” score. Each band indicates the probability of suffering from a psychiatric disorder as derived from the DAWBA psychiatric interview. Of the integer scores from 0 to 5 correspond to respective probabilities of <0.01%, 0.5%, 3%, 15%, 50%, and >70% of satisfying the diagnostic. We analyzed the combined EXT DAWBA band score, which includes ODD, CD, and ADHD, and the combined INT DAWBA band score, which includes major depression, generalized anxiety disorder, specific phobia, social phobia (at age 7, 10, 13, and 15), separation anxiety disorder (at age 7, 10, 13, and 15), and panic disorder and agoraphobia (at age 15). All disorders were assessed by the child’s mother except for self-reports of INT disorders at age 15.

The DAWBA band scores have shown a positive association with clinician-rated diagnosis (chance corrected 0.4–0.8, sensitivity 0.98–0.99) and a strong relation with indicators of mental health (Goodman, Heiervang, Collishaw, & Goodman, 2011). The INT DAWBA band scores (INT) and EXT DAWBA band scores (EXT) reflect the probability of satisfying the diagnostic criteria of any INT or any EXT disorder (Goodman et al., 2011). Because Category 0 did not occur in all assessments, 0 and 1 scores were collapsed into a single category (i.e., <0.5%). We used the DAWBA band scores because they provide more information than the dichotomous affected/unaffected variable.

We included maternal smoking during pregnancy (no/yes), maternal highest education (six categories), maternal and paternal social class (six categories), maternal age at de-
livery, birth weight, and sex as predictors of class membership. Maternal smoking was assessed at week 18; social class and educational attainment at week 32. Maternal age at delivery and birth weight were part of the pregnancy and child baseline data.

Statistical methods

As a baseline model, we fitted a latent growth curve model to the repeated measures of INT and EXT. This model described a single trajectory that can randomly vary over individuals and included three factors, an intercept, linear slope, and quadratic slope factor, where the quadratic slope factor allows for curvilinear development. Because we estimated the means (fixed effects) and the variances (random effects) of the intercept, linear slope, and quadratic slope factors, this is a random effects model, implying that each child was characterized by his or her own unique growth curve (Singer & Willett, 2003).

GMM extend the standard growth model with a latent class variable, each describing a subset of the entire population and featuring a distinct growth model within each latent class. Because class membership is unknown, subjects with similar trajectories are grouped into classes in a data-driven fashion. Fixing the variances of the intercept, linear slope, and quadratic slope factors to zero within each class results in a restrictive GMM, also known as latent growth curve models (LGCMs), in which only average within-class trajectories are estimated (i.e., means of intercept, linear slope, and quadratic slope), and all variability within classes is considered to be occasion specific (Nagin, 1999). In addition to LGCMs, we also fitted random intercept models in which the variance of the intercept was not fixed to zero, thus allowing for within-class individual differences in the intercepts (Jung & Wickrama, 2008; Muthén & Muthén, 2000). In fitting both LGCMs and the random intercept models, we considered models with increasing number of classes. Mixture models with random slope or random quadratic terms often failed to converge and are therefore not considered.

Based on the best fitting separate models, a combined model of INT and EXT trajectories between age 7 and 15 was tested, in which the EXT latent categorical class variable ($C_E$) was regressed on the INT latent categorical class variable ($C_I$; see Figure 1). This multinomial regression analysis provided an omnibus test of the null hypothesis that INT and EXT classes are unrelated. Note that the direction of this regression is arbitrary, and has no effect on the interpretation of the results. Reversing the direction of the regression to EXT on INT would result in exactly the same model fit and parameter estimates. INT and EXT class variables $C_I$ and $C_E$ were also regressed on maternal social class, paternal social class, maternal educational level, maternal age at delivery, birth weight, and sex as predictors of class membership. Maternal smoking was assessed at week 18; social class and educational attainment at week 32. Maternal age at delivery and birth weight were part of the pregnancy and child baseline data.

### Table 1. The number of individuals for each DAWBA band score for EXT and INT disorders, prevalence for EXT and INT, and total numbers, and the polychoric correlations between EXT and INT for each age

<table>
<thead>
<tr>
<th>DAWBA Band Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Prevalence</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT 7</td>
<td>1509</td>
<td>3948</td>
<td>612</td>
<td>160</td>
<td>96</td>
<td>0.058</td>
<td>6325</td>
</tr>
<tr>
<td>EXT 10</td>
<td>1668</td>
<td>3388</td>
<td>520</td>
<td>124</td>
<td>90</td>
<td>0.054</td>
<td>5790</td>
</tr>
<tr>
<td>EXT 13</td>
<td>1848</td>
<td>2682</td>
<td>608</td>
<td>115</td>
<td>79</td>
<td>0.055</td>
<td>5332</td>
</tr>
<tr>
<td>EXT 15</td>
<td>1714</td>
<td>1848</td>
<td>335</td>
<td>88</td>
<td>65</td>
<td>0.05</td>
<td>4050</td>
</tr>
<tr>
<td>INT 7</td>
<td>2749</td>
<td>3063</td>
<td>412</td>
<td>76</td>
<td>45</td>
<td>0.037</td>
<td>6345</td>
</tr>
<tr>
<td>INT 10</td>
<td>2092</td>
<td>3101</td>
<td>470</td>
<td>89</td>
<td>58</td>
<td>0.045</td>
<td>5810</td>
</tr>
<tr>
<td>INT 13</td>
<td>2181</td>
<td>2686</td>
<td>382</td>
<td>68</td>
<td>33</td>
<td>0.038</td>
<td>5350</td>
</tr>
<tr>
<td>INT 15</td>
<td>1506</td>
<td>1885</td>
<td>521</td>
<td>106</td>
<td>15</td>
<td>0.051</td>
<td>4033</td>
</tr>
</tbody>
</table>

Note: DAWBA, Development and Well-Being Assessment; EXT, externalizing; INT, internalizing. The correlations (all significant at $p < .05$ corrected for multiple testing) are below the diagonal, and the standard errors are above the diagonal.
ery, maternal smoking during pregnancy, birth weight of the child, and sex of the child to test whether these variables predict trajectories.

Models were fitted with Mplus 6.12 (Muthén & Muthén, 2007) using robust full information maximum likelihood. If initial settings did not result in replicated minima, the number of starts was increased from 500 to 2,000, and the number of final optimization from 50 to 200. If the best likelihood was not replicated with 2,000 starts, the model was considered to have failed. The choice of the best fitting model was based on the sample size adjusted Bayes information criterion. In cases of small differences in fit, the preciseness of individual assignment to a specific trajectory and the interpretability of the model were also considered. The certainty of class assignment of the individuals is captured by the entropy index presented in the Results section. A higher entropy implies a higher degree of certainty concerning the assignments.

Previous analyses have shown that missingness in the ALSPAC data is not random, but that it only marginally affects parameter estimates in statistical analyses (Wolke et al., 2009). As a form of attrition analysis, we regressed the number of missing DAWBA assessments per individual on the covariates. This attrition analysis showed that sex, smoking during pregnancy, maternal and paternal social class, maternal highest education, and maternal age at delivery significantly predicted missingness ($R = .274$, $F = 124.6$, $p < .0001$). Because these variables were included in the model, our model is robust for missingness conditional on these variables (i.e., data missing at random; Little & Rubin, 2014). We reran the models on listwise complete data to evaluate the models based on individuals with complete data, and the same conclusions regarding the class selection would be drawn.

**Results**

**Descriptives**

Table 1 provides an overview of the prevalences of the observed DAWBA band scores and the polychoric correlations between INT and EXT at age 7, 10, 13, and 15. Polychoric correlations quantify the association between ordinal variables (Ekstrom, 2011). The estimated prevalence of EXT disorders in our sample was around 5% between ages 7 and 15. The prevalence of INT disorders was around 4% in childhood, and grew to 5% at age 15. As expected, male average EXT scores were greater than female average EXT scores at all ages, and female average INT scores were greater than male average INT at ages 13 and 15. Correlations between EXT and INT were around .20. Longitudinal correlations for INT between age 7 and 15 were .15 to .48, whereas correlations for EXT between the ages of 7 and 15 were higher at .35 to .61.

**INT and EXT trajectories**

GMMs were fitted for INT and EXT separately. The single-class model with a random intercept, slope, and quadratic term showed a worse fit than models including a latent class variable, which indicates the existence of subgroups with different trajectories. Models with two to six classes were tested with (a) a fixed intercept, slope, and quadratic term; and (b) a random intercept, and a fixed slope and quadratic term. Table 2 provides the model fit statistics and entropy. We retained the quadratic term because models with the quadratic term generally outperformed models without a quadratic term (results available on request from the first author). INT data were best described by a model with five classes with...
Classes with the variances of the intercept and slopes fixed to zero in each class). The random models are for a random I (i.e., with the variance of the intercept estimated appropriately adjusted likelihood ratio between the model under consideration and the model with one class less. In each class) and fixed effects for S and Q. The reference model is with random ISQ. The LMR column provides the LMR, Lo–Mendell–Rubin; NR, best log-likelihood not replicated at 2000 starts and 200 final iterations. The fixed effects models are for ISQ slopes (i.e., with the variances of the intercept and slopes fixed to zero in each class). The random models are for a random I (i.e., with the variance of the intercept estimated in each class) and fixed effects for S and Q. The reference model is with random ISQ. The LMR column provides the p values for the LMR test, which tests the appropriately adjusted likelihood ratio between the model under consideration and the model with one class less.

For INT, there were two classes with low scores, called the most likely class membership) and the low-INT class (41.8%; Figure 2a). A third class contained individuals with decreasing INT scores (5.1%). The remaining two classes contained individuals with increasing scores. The increasing-INT class (17.8%) showed a steady rise in score from childhood on, while in the adolescent-risk INT class (12.6%) the scores are low until age 13 but sharply increase at age 15 years.

Four of the five EXT trajectories showed similar patterns as the INT trajectories (Figure 2b): that is, the very low EXT class (28%), the low-EXT class (54%), the decreasing-EXT class (7%), and the increasing-EXT class (8.3%). The final high-EXT class (2.4%) was different, because it contained individuals with persisting high scores from childhood to adolescence. Sex was a significant predictor of INT and EXT class membership, with girls being significantly more likely than boys to be a member of the decreasing (odds ratio $\text{OR} = 2.011$, $p < .001$), increasing ($\text{OR} = 7.800$, $p < .001$), or adolescent increasing INT classes ($\text{OR} = 3.128$, $p < .001$). In addition, girls were significantly less likely to be a member of the high ($\text{OR} = 0.074$, $p < .001$), increasing ($\text{OR} = 0.475$, $p < .001$), and decreasing ($\text{OR} = 0.178$, $p < .001$) EXT class. Maternal smoking during pregnancy was a significant risk factor for being a member of the decreasing INT class ($\text{OR} = 1.991$, $p < .001$) and for the high (OR = 2.237, $p < .001$), increasing (OR = 2.053, $p < .001$), and decreasing (OR = 2.765, $p < .001$) EXT classes. With respect to EXT, higher social class of the father was associated with a lower probability of belonging to the

### Table 2. Fit indices for the INT and EXT growth mixture models containing one to six classes

<table>
<thead>
<tr>
<th>Models INT</th>
<th>N</th>
<th>Classes</th>
<th>Entropy</th>
<th>AIC</th>
<th>Adj-BIC</th>
<th>LMR</th>
<th>Entropy</th>
<th>AIC</th>
<th>Adj-BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed ISQ</td>
<td>1</td>
<td>2</td>
<td>0.469</td>
<td>42326.785</td>
<td>42389.782</td>
<td>&lt;.001</td>
<td>0.323</td>
<td>41946.080</td>
<td>42016.489</td>
</tr>
<tr>
<td>Random I, Fixed SQ</td>
<td>2</td>
<td>0.432</td>
<td>41959.951</td>
<td>42063.711</td>
<td>&lt;.001</td>
<td>0.455</td>
<td>41674.774</td>
<td>41789.651</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Random ISQ</td>
<td>2</td>
<td>0.513</td>
<td>41650.452</td>
<td>41794.975</td>
<td>&lt;.001</td>
<td>0.392</td>
<td>41536.956</td>
<td>41696.303</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3</td>
<td>0.488</td>
<td>41536.394</td>
<td>41721.680</td>
<td>&lt;.001</td>
<td>NR</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.518</td>
<td>41505.755</td>
<td>41731.804</td>
<td>.021</td>
<td>NR</td>
<td>—</td>
<td>—</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: EXT, Externalizing; INT, internalizing; ISQ, intercept, linear, and quadratic; AIC, Akaike information criterion; BIC, Bayesian information criterion; LMR, Lo–Mendell–Rubin; NR, best log-likelihood not replicated at 2000 starts and 200 final iterations. The fixed effects models are for ISQ slopes (i.e., with the variances of the intercept and slopes fixed to zero in each class). The random models are for a random I (i.e., with the variance of the intercept estimated in each class) and fixed effects for S and Q. The reference model is with random ISQ. The LMR column provides the p values for the LMR test, which tests the appropriately adjusted likelihood ratio between the model under consideration and the model with one class less.

In the combined model, the association between INT and EXT was analyzed using the multinomial logistic regression of the five-class EXT trajectories on the five-class INT trajectories (Figure 1). The model including the multinomial regression parameters fitted the data better than a model that dropped these parameters (likelihood ratio = 477.894, df = 16, $p < .0001$). We first describe the INT and EXT trajectories and then discuss the association between the EXT and INT trajectory class variables.

For INT, there were two classes with low scores, called the very low INT class (22.7% of the sample based on a fixed intercept, slope, and quadratic term. For EXT, the best fitting model is the three-class random intercept model. However, this model has a very low entropy compared to the fixed intercept models. Among the fixed intercept models, the best fitting model is the six-class model, but the five-class model has substantially better entropy and only a slightly worse fit. Visual inspection of the trajectories showed that the six-class model adds a third unaffected class to the very low and low classes, which starts out low and progresses to very low EXT scores. Because this extra class is not very informative, the five-class fixed intercept model was preferred. The results of the analyses of the listwise complete data were similar and also resulted in the selection of a five-class model for both INT and EXT. Models fitted on listwise complete data had a higher entropy, reflecting that individuals with complete data available are easier to categorize.

**Combined INT/EXT model**

In the combined model, the association between INT and EXT was analyzed using the multinomial logistic regression of the five-class EXT trajectories on the five-class INT trajectories (Figure 1). The model including the multinomial regression parameters fitted the data better than a model that dropped these parameters (likelihood ratio = 477.894, df = 16, $p < .0001$). We first describe the INT and EXT trajectories and then discuss the association between the EXT and INT trajectory class variables.

For INT, there were two classes with low scores, called the very low INT class (22.7% of the sample based on most likely class membership) and the low-INT class (41.8%; Figure 2a). A third class contained individuals with decreasing INT scores (5.1%). The remaining two classes contained individuals with increasing scores. The increasing-INT class (17.8%) showed a steady rise in score from childhood on, while in the adolescent-risk INT class (12.6%) the scores are low until age 13 but sharply increase at age 15 years.

Four of the five EXT trajectories showed similar patterns as the INT trajectories (Figure 2b): that is, the very low EXT class (28%), the low-EXT class (54%), the decreasing-EXT class (7%), and the increasing-EXT class (8.3%). The final high-EXT class (2.4%) was different, because it contained individuals with persisting high scores from childhood to adolescence. Sex was a significant predictor of INT and EXT class membership, with girls being significantly more likely than boys to be a member of the decreasing (odds ratio $\text{OR} = 2.011$, $p < .001$), increasing ($\text{OR} = 7.800$, $p < .001$), or adolescent increasing INT classes ($\text{OR} = 3.128$, $p < .001$). In addition, girls were significantly less likely to be a member of the high ($\text{OR} = 0.074$, $p < .001$), increasing ($\text{OR} = 0.475$, $p < .001$), and decreasing ($\text{OR} = 0.178$, $p < .001$) EXT class. Maternal smoking during pregnancy was a significant risk factor for being a member of the decreasing INT class ($\text{OR} = 1.991$, $p < .001$) and for the high ($\text{OR} = 2.237$, $p < .001$), increasing ($\text{OR} = 2.053$, $p < .001$), and decreasing ($\text{OR} = 2.765$, $p < .001$) EXT classes. With respect to EXT, higher social class of the father was associated with a lower probability of belonging to the
Figure 3 displays the conditional probabilities of belonging to the EXT (INT) classes given membership of a given INT (EXT) class. These conditional probabilities showed that similar INT and EXT classes were associated. Focusing on the “affected” trajectories revealed that individuals in the decreasing INT class had a high probability of belonging to the decreasing EXT class (38%), and children in the increasing INT class had a substantial probability (22%) of being member of the increasing EXT class (Figure 3a). Likewise, 27% of the children in the decreasing EXT class belonged to the decreasing INT class, and children in the increasing EXT class had a substantial chance (46%) of belonging to the increasing INT class (Figure 3b). It further becomes apparent that the high EXT class was particularly associated with the decreasing INT class and less with the increasing INT class, whereas the adolescent onset INT class was independent from EXT trajectories.

**Discussion**

Based on analyses of DSM-IV based INT and EXT problem scores obtained in longitudinal studies from childhood into adolescence, we conclude that developmental trajectories for INT and EXT are largely similar, and that the INT and EXT trajectories are associated. Notable differences in trajectories are that for INT, there is a trajectory that is characterized by increased scores from adolescence onward, while for EXT there is a trajectory that is characterized by stable high scores. The adolescent-onset INT group showed no association with affected EXT classes. The high EXT group was most associated with the decreasing INT group, signifying that in some of the individuals that have both EXT and INT symptoms during childhood, the EXT symptoms will persist, while the INT symptoms attenuate.
Our findings for the INT and EXT trajectories are largely in line with a priori expectations. The most apparent discrepancies are the absence of a class characterized by stable high INT and the presence of an increasing EXT class. The absence of a stable high INT group may be due to the relatively low prevalence of these disorders, especially during childhood. GMMs require very large samples to reliably identify classes that consist of a small proportion of the sample. Our finding therefore do not rule out that there is a small group of children with persisting symptoms, as suggested by other studies (Hofstra et al., 2000, 2002). Previous results regarding an increasing trajectory for EXT symptoms were mixed. Van Lier et al. (2007) and Larsson et al. (2011) identified a class with increasing symptoms for conduct disorder, and for the inattentive subtype of ADHD, respectively. The increasing EXT class in our study probably comprises these groups.
The only two other studies (Brezo et al., 2008; Fanti & Henrich, 2010) that looked at the combination of INT and EXT trajectories up to age 12 also showed that increasing and decreasing INT and EXT trajectories are mutually dependent. It will be interesting to see whether future studies modeling the trajectories into adolescence will replicate our finding that INT disorders with onset in adolescence are independent of EXT disorders and that persisting high EXT is mainly associated with decreasing INT. Our results indicate that the previously reported longitudinal association between INT and EXT disorders starts in childhood. The recent finding in ALSPAC that adolescent depression is predicted by conduct problems in childhood (Stringaris, Lewis, & Maughan 2014) might be attributable to persistent childhood INT symptoms.

Because mixture modeling is an exploratory technique, the results presented here require replication (Lubke, 2012). A related issue is the relatively low entropy, or certainty of class assignments in the models, suggesting that class assignment based on the model is imprecise. We note that entropy was higher in the analyses carried out with listwise complete data: 0.76 for the five-class EXT model and 0.572 for the five-class INT model. This reflects the fact that class assignment is substantially more accurate in subjects with data available at each time point. However, while class assignment is more precise for individuals with complete data, including subjects with missing data has the advantage of rendering the results robust to missingness associated with a high score at an earlier measurement occasion, and with an included risk factor. Although we fit a model robust for missingness associated with the included covariates, we acknowledge dropout associated with other covariates, not included in our models, may have affected the results in an unknown way. We note that the associations between sex, prenatal risk factors, and trajectory membership probability were all in the expected directions. Boys were found to be more at risk for EXT problems and girls more for INT problems, and adverse prenatal risk factors were associated with EXT, and to a lesser extent, with INT problems (see, e.g., Ormel et al., 2014). In sum, the agreement between the estimated trajectories and expectations based on previous work increases the confidence in these developmental trajectories.

One of the other potential limitations, besides the attrition rate, could be the use of two broad INT and EXT problem scores. Studies investigating specific INT or EXT symptom domains (Barker, 2007; Broeren, Muris, Diamantopoulou, & Baker, 2013; Larsson et al., 2011; Van Lier et al., 2007) have detected differences in trajectories between the separate disorders. Given the low prevalence rates of the individual disorders, such analyses were not feasible here. Moreover, previous studies have also shown that the analyzed disorders load on common factors interpretable as our INT and EXT (see, e.g., Angold et al., 1999; Beauchaine & McNulty, 2013; Cosgrove et al., 2011). This indicates that studies focusing on measures of a general tendency to display INT or EXT disorders can also provide important information. Because sample size precluded separate analyses in males and females, gender was included as a covariate predicting class membership. We acknowledge that the slight variation in constituent disorders and raters over time is a limitation. INT disorders were measured by self-rating at age 15, and by maternal rating at ages 7 to 13. The design of the ALSPAC study is such that the set of INT disorders at age 7 to 13 includes separation anxiety, while the measure at age 15 includes agoraphobia and panic disorder. This is based on the expectation that agoraphobia in early childhood and separation anxiety in late adolescence are unlikely to be present. The change in disorders measured across the time span of this study can be viewed as an attempt to measure heterotypic continuity of the same underlying INT construct, a technique often applied when considering longitudinal development (Petersen, Bates, Dodge, Lansford, & Pettit, 2014). Finally, the correlations between INT and EXT disorders (around .20) were lower than previously reported by Cosgrove et al. (2011; .20 to .30). However, these differences are relatively small, and might be due to differences in the instrument used.

Other questionnaires, such as the Child Behavior Checklist (CBCL) and the Youth Self-Report (Achenbach & Rescota, 2001) are often used to measure INT and EXT psychopathology in large population samples. Growth curve analyses based on the CBCL/Youth Self-Report INT symptoms yielded evidence for heterogeneous development of INT problems, but this variation could not be easily disaggregated into different latent growth classes (i.e., a model including a random intercept and a slope provided the best fit; Lubke et al., 2015). This difference in outcome is likely to be because the CBCL is a continuous measure capturing individual differences in the general population, and as such it provides more information to estimate random growth parameters, whereas the measure used in this study is ordinal and developed for clinical practice.

The strengths of this study were the use of a large population-based sample with repeated measures from childhood into adolescence and a DSM-IV based psychiatric interview instrument. This enables the translation of these findings to clinical relevance. The results suggest that if an adolescent presents with INT symptoms and has no history of previous INT symptoms, a brief screening for EXT disorders will suffice. However, when confronted with childhood or adolescent EXT problems or with INT problems that were already apparent in childhood, comorbid symptoms should also be assessed at the start of the treatment. Risk and protective factors can also be taken into account. If the mother smoked during pregnancy, the chance to have a trajectory of increasing or persisting EXT symptoms is higher, whereas INT symptoms may decrease. A protective factor for unfavorable EXT trajectories is higher social class. Future studies should address the specific treatment needs of children with co-occurring INT and EXT disorders, especially because the co-occurrence is related to negative outcomes (Fanti & Henrich, 2010). An interesting question is whether successful treatment of an EXT disorder also leads to a remission of the INT disorder, or vice versa, or whether treatment of both disorders is necessary. There is some evi-
dence that INT symptoms respond to the treatment of EXT symptoms (Chase & Eyberg, 2008), and vice versa (Kendall, Brady, & Verduin, 2001). Moreover, it is important to identify the factors associated with the combination of the trajectories of decreasing INT and EXT symptoms versus the combination of persisting EXT symptoms and decreasing INT symptoms.

The current study does not address the etiology of comorbidity between INT and EXT disorders. Different hypotheses currently exist about the causes of comorbidity. It has been suggested that depressive symptoms in ADHD are due to demoralization (Brown, Borden, Clingerman, & Jenkins, 1988), but in line with our finding that a combination of trajectories of EXT symptoms in childhood and later increasing INT symptoms did not exist, Biederman, Mick, and Faraone (1998) concluded that this does not explain all comorbidity. The opposite (i.e., INT symptoms causing EXT symptoms) has also been hypothesized. Granic (2014), for example, proposes three mechanisms explaining how anxiety can cause aggression and suggests how future research could investigate whether these mechanisms play a role. Another explanation for comorbidity is that multiple disorders are caused by the same underlying mechanism, which agrees with the observed co-occurrence of similar trajectories. Cross-sectional twin studies have indicated that co-morbidity between INT and EXT disorders is partly explained by shared genetic risk factors (e.g., Cosgrove et al., 2011). It has previously been shown that ADHD trajectories are influenced by genetic factors (Larsson et al., 2011). This could also be the case for co-occurring trajectories, which would be interesting for gene-finding studies. Including a Genetic Variant × Course (i.e., decreasing or stable high) interaction term enables the identification of variants associated with a favorable or unfavorable outcome and reveal hints about biological differences in etiology between developmental courses.

The mechanisms underlying the association between the risk factors, such as maternal smoking during pregnancy, paternal social class and education, and the developmental trajectories are also a subject for further investigation. Many of the associated risk factors are influenced by genetic factors. The question whether the associations between exposures and adverse trajectories are explained by common genetic factors, common environmental exposures, or by direct causation can be addressed by genetically informative designs, such as studies of developmentally concordant and discordant monozygotic and dizygotic twins (McGue, Olsen, & Chtrisen, 2010; van Dongen, Slagboom, Draisma, Martin, & Boomsma, 2012) or children of twins designs (McAdams et al., 2014). The advent of high throughput genome wide (epi)genetic measurement will further allow for a more in-depth study of the origins of biological differences underlying developmental differences. Specifically, subtle developmental differences beyond lifetime diagnoses are interesting targets for studies relating epigenetic variation to variations in psychopathology.

To summarize, we showed that both INT and EXT disorders can have a favorable or unfavorable course in time from childhood into adolescence and that trajectories are associated with each other. Future research should focus on unraveling the etiology of the co-occurrence, and focus on the development of treatment designs for the most seriously affected children.

References


