Am J Med Genet B Neuropsychiatr Genet. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:
Am J Med Genet B Neuropsychiatr Genet. 2016 October ; 171(7): 948–957. doi:10.1002/ajmg.b.32375.

A powerful phenotype for gene-finding studies derived from trajectory analyses of symptoms of anxiety and depression between age 7 and 18

Gitta H Lubke[1,2,*], **Patrick J Miller**[1], **Brad Verhulst**[3], **Meike Bartels**[2,4,5], **Toos van Beijsterveldt**[2], **Gonneke Willemsen**[2,5], **Dorret I Boomsma**[2,4,5], and **Christel M Middeldorp**[2,4,6]

[1]Department of Psychology, University of Notre Dame, Notre Dame, Indiana [2]Department of Biological Psychology, VU University, Amsterdam, The Netherlands [3]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia [4]Neuroscience Campus Amsterdam, VU University Amsterdam, Amsterdam, The Netherlands [5]EMGO+ Institute for Health and Care Research, VU University Medical Centre, Amsterdam, The Netherlands [6]Department of Child and Adolescent Psychiatry, GGZ inGeest/VU University Medical Center, Amsterdam, The Netherlands

Abstract

To investigate the utility of longitudinal data in genetic analyses of symptoms of anxiety and depression, we assessed individual differences between age 7 and 18 using growth mixture models, and investigated the genetic and non-genetic factors contributing to the trajectories.

Mothers of 7,706 girl and 7,418 boy twins from the Netherlands Twin Register rated the anxious depression scale (SxAnxDep) of the Child Behavior Check List (CBCL) at age 7, 10 and 12 years. 2,706 girl and 1,856 boy twins completed the Youth Self Report (YSR) at age 14, 16 and 18.

While individual trajectories varied considerably, these differences were largely idiosyncratic and could not be grouped into separate latent classes with class-specific average growth curves. The intercept, which reflects the individuals' baseline level of SxAnxDep across time, explained 55-58% of the total phenotypic variance. The slope factor, which captures a common average trend over time, did not explain variance in the phenotype. This finding also underlines the high level of idiosyncrasy of trajectories that lack a common longitudinal structure.

The analyses of twin data showed that the random intercept factor of SxAnxDep during childhood and during adolescence is considerably more heritable than the observations at any single age, namely between 60% and 84%. One explanation is that different factors contribute to the level of symptoms of anxiety and depression at any given time point, including temporary events and emotions. When considering baseline stability, these temporary influences average out, with the result of a more reliable and more heritable phenotype.

[*]Correspondence to: Gitta Lubke & Christel Middeldorp, 110 Haggar Hall, Notre Dame, Indiana 46556, (574) 631-8789, Email: glubke@nd.edu.

Europe PMC Funders Author Manuscripts

Keywords

longitudinal; genetics; depression

Europe PMC Funders Author Manuscripts

Introduction

It has proven to be more difficult to identify genetic variants influencing major depressive disorder (MDD) than other psychiatric disorders. In a genome-wide association (GWA) study with a sample size comparable to, for example, a GWA analysis of bipolar disorder (Group 2011), no single nucleotide polymorphism (SNP) reached genome-wide significance (Wray et al. 2010). Estimates of the variance explained by all SNPs varied between 20 and 30% for MDD (Lee et al. 2013; Lubke et al. 2012) indicating that the SNPs analyzed in a GWA study do have an effect on MDD, that can be captured by increasing the sample size. One of several reasons that sample sizes for MDD need to be larger than for other psychiatric disorders is the often mentioned heterogeneity of the phenotype (Levinson et al. 2014; Wray et al. 2010). Levinson et al. (2014) sum up several variables that could explain heterogeneity, such as sex, age of onset, recurrence, symptom profile and longitudinal course. Decreasing the heterogeneity, in addition to an increase in sample size, could also lead to improved statistical power in a GWA study.

MDD is often comorbid with anxiety disorders. So far, these disorders have been less extensively investigated than MDD in GWA studies and no genome-wide significant hits have been observed (Erhardt et al. 2011; Otowa et al. 2012; Walter et al. 2013). Genetic epidemiological analyses in both adults and children showed that anxiety and depression share genetic risk factors (Kendler et al. 2011; Middeldorp et al. 2005; Rhee et al. 2015). This suggests that it could be useful to focus the search for genetic variants on a phenotype that comprises depression as well as anxiety. A GWA study on internalizing symptoms in children aged around three years old reported that around 20% of the variance was explained by SNPs (Benke et al. 2014). Moreover, genetic variants influencing later adult psychiatric disorders appeared to jointly have an effect on internalizing symptoms at age 3.

In the current paper, we focus on the longitudinal course of symptoms of anxiety and depression during childhood and adolescence, and investigate the role of genetic factors on the developmental course. The aim is to identify a more homogeneous phenotype which would provide increased statistical power in a genome-wide association study.

Epidemiological studies of the development of anxiety disorders and depression during childhood and depression, were reviewed by Merikangas et al (2009). Depression rates are low during childhood and increase during adolescence, especially in girls. For anxiety, the picture differs for the specific anxiety disorders, but overall there is an increase in prevalence from childhood to adulthood. Longitudinal studies show the heterogeneity in the developmental course of anxiety and depressive disorders during childhood and adolescence. Some of the children with anxiety and depression continue to have symptoms in adolescence, but some remit (Copeland et al. 2009). There is also heterotypic continuity, i.e., children with anxiety disorders are at risk for depression at later ages (Beesdo-Baum and Knappe 2012). In addition, as expected from the increased prevalence during adolescence,

there is a group with an onset of symptoms during adolescence without any preceding symptoms during childhood (reviewed in Costello et al (2011). These patterns are also seen when studying continuous measures of anxiety and depression. In the Young Netherlands Twin Register, which assesses anxious depression symptoms (SxAnxDep) in twins every two to three years, mean scores remained at the same level between age 7 and 12 and then showed an increase. The increase was larger in girls than in boys (Nivard et al. 2015). Further, the two year correlations for SxAnxDep were approximately 0.5 in childhood and 0.6 in adolescence indicating stability as well as change in this age period (Nivard et al. 2015).

The heterogeneity in the developmental course in anxiety and depression has been studied in more detail in several population based studies by analyzing whether different characteristic developmental trajectories can be identified using Growth Mixture Models (GMM) (Brendgen et al. 2005; Broeren et al. 2013; Cote et al. 2002; Crocetti et al. 2009; Dekker et al. 2007; Duchesne et al. 2008; Fanti and Henrich 2010; Feng et al. 2008; Legerstee et al. 2013; Letcher et al. 2012; Letcher et al. 2009; Marmorstein et al. 2010; Morin et al. 2011; Nivard et al. submitted; Rodriguez et al. 2005; Sterba et al. 2007; Toumbourou et al. 2011). However, only four studies performed in three different cohorts, covered the whole period from childhood and adolescence, until at least age 15 (Dekker et al. 2007; Letcher et al. 2009; Nivard et al. submitted; Toumbourou et al. 2011). The results differed between the three cohorts. All found trajectories with consistently low scoring individuals. However, Dekker et al (2007) identified additional gender specific trajectories. In girls, a stable high trajectory was identified, while in boys, decreasing trajectories were identified. Nivard et al (submitted) observed an increasing and decreasing trajectory, but no stable high trajectory. In the third cohort, in boys and in girls, high, increasing and decreasing trajectories were found in addition to the low trajectories (Letcher et al. 2009; Toumbourou et al. 2011). The studies performed in either childhood or adolescence, in general, found a low scoring and a high scoring class. Results also varied regarding the other classes, i.e., whether classes with increasing or decreasing scores over time were also observed (Brendgen et al. 2005; Broeren et al. 2013; Cote et al. 2002; Crocetti et al. 2009; Dekker et al. 2007; Duchesne et al. 2008; Fanti and Henrich 2010; Feng et al. 2008; Legerstee et al. 2013; Letcher et al. 2012; Letcher et al. 2009; Marmorstein et al. 2010; Morin et al. 2011; Rodriguez et al. 2005; Sterba et al. 2007; Toumbourou et al. 2011). Note that all but three studies (Crocetti et al. 2009; Morin et al. 2011; Nivard et al. submitted) fixed the variance of the intercept and slopes to zero in all classes. This lack of random effects within the latent growth curve model can lead to an over-extraction of classes because individual variability in intercepts and slopes is captured by additional classes (e.g. a high and low class)(Lubke and Neale 2006; Muthen and Muthen 2000).

Data obtained in twins can be used to investigate how genetic factors influence the longitudinal course of a trait. The outcomes of the twin analyses can thereby indicate the most suitable strategy to analyze longitudinal data in a GWA study. Earlier studies applying general multivariate or simplex models to longitudinal twin data have shown that stability in SxAnxDep is mostly explained by genetic factors, but no twin study so far has investigated longitudinal SxAnxDep data while attempting to take individual differences in the developmental course into account. In the current study, growth mixture models were fitted

to investigate, first, which longitudinal model provides the best description of individual differences in the SxAnxDep course, and, second, how the growth factors in the best fitting model are influenced by genetic and non-genetic factors. We modeled SxAnxDep between age 7 and age 18 in girls and in boys, and estimated the variances of intercepts and slopes in the classes. From age 7 to age 12, maternal ratings were available whereas from age 14 to age 18 data self-report ratings were available. Due to the change of rater from mother to self-reports between ages 12 and14, a piecewise growth model framework was used. In addition to the trajectory analyses, the twin data were used to estimate the influence of genetic and non-genetic factors on the latent factors.

Materials & Methods

Sample

The data for this study were collected by the Young and Adult Netherlands Twin Register between 1987 and 2015 (Boomsma et al. 2006; van Beijsterveldt et al. 2013). Children are enrolled in the YNTR by their parents at birth. For the ANTR, adolescent and adult twins were recruited through city-councils. The minimum age to participate in the ANTR was age 12. Every two to three years lifestyle, health and behavior are assessed by surveys in an ongoing data collection. Previous research has established that the NTR data can be considered as representative of the Dutch population (van Beijsterveldt et al. 2013).

Two groups of subjects were selected to, first, estimate the developmental trajectories and, second, to estimate the heritability of these trajectories. For the first set of analyses, data from twins with measures at two or more time points were included. These data either were collected between ages 7-12 years or between ages 14-18 years. This resulted in a sample size of n=15,124 (7,706 girls and 7,418 boys) between ages 7-12, and n=4,563 (2,706 girls, and 1,856 boys) between ages 14-18. The two subsamples included 1,970 female twins and 1,337 male twins with data in both age bins. The numbers of female and male monozygotic (MZ) and dizygotic (DZ) twins that were included in the trajectory analyses are presented in Table 1A.

For the heritability analyses all data were included. Data were available for 12,225 twin pairs between ages 7-12 (12,188 complete twin pairs), and for 8,241 twin pairs between ages 14-18 (6,716 complete twin pairs). The number of male and female monozygotic (MZ) and dizygotic (DZ) twin pairs is given in Table 1B.

Measures

Symptoms of anxiety and depression (SxAnxDep) were measured with the anxious depression scales of the mother rated Child Behavior Checklist (CBCL) (Achenbach and Rescorla 2001) and the Youth Self Report (YSR) (Verhulst et al. 1997) containing respectively 14 and 16 items. Example items are “cries a lot”, “fears’, “must be perfect”. The test-retest reliability of the SxAnxDep CBCL and YSR items are 0.82, and 0.74 respectively (Achenbach and Rescorla 2001). Achenbach & Rescorla (2001) also provide evidence for content, criterion and construct validity of the items. Moreover, the CBCL and YSR anxious

depression scales predict DSM-IV diagnoses of both anxiety disorders and depression (Bellina et al. 2013; van Lang et al. 2005).

In the current study we computed an average score at each age for each individual. Due to the skewness of the average score, the average score was categorized into 4 categories. This was done based on the sample quartiles for sxAnxDep at age 18 which ensured that there were observations in all cells at each age. Category endorsement rates of this aggregate score are presented in Table 2 separately for males and females at each age. Correlations between ages are given in Table 3, showing male correlations below and female correlations above the diagonal. As can be seen, correlations are generally higher between successive ages, and are higher over several ages for females than they are for males. Although there is an expected drop in correlations when raters change (mother vs. self-ratings), correlations between 12-year olds rated by mothers and 14, 16, and 18 self-ratings are similar to the correlation between 14 and 18 year old self ratings. This indicates that although raters change, the measured trait remains comparable.

Analysis Plan

The overarching goal of the analyses was to investigate the benefit of analyzing longitudinal data in genetic studies of anxiety and depression. The planned analyses consisted of two parts, a latent growth mixture analysis, and a twin-based heritability analysis.

The goal of the growth mixture analysis was to investigate whether we could identify different latent classes with characteristic growth trajectories representing the developmental course of SxAnxDep between age 7 and 18, potentially combining data from mother and self-ratings. The heritability analysis aimed at investigating how longitudinal data can be optimally used to identify genetic variants, and, in case the mixture models revealed multiple classes, investigate potential differences in heritability across classes.

1. Separate and Joint Piecewise Growth Mixture Modeling

In the first step, we fitted separate linear growth mixture models for males and females, and for mother and self-ratings. In the latent growth model, development is modeled by an intercept (I) and a linear slope (S), and if more than three time points are available, curvilinear growth can be modeled. Growth mixture models (GMM) extend the standard growth model with a latent class variable, featuring a distinct growth model within each latent class. Subjects with similar trajectories are grouped into classes in a data-driven fashion, since class membership is not known beforehand. Fixing the variances of the intercept (I) and linear slope (S) factors to zero within each class results in a restrictive GMM (also known as latent class growth models, LCGMs), in which only average within class trajectories are estimated (i.e., means of I and S), and all variability within classes is considered to be occasion specific (Nagin 1999). We fitted LCGM's as well as models allowing for within class individual differences in the intercepts and slopes, i.e., random intercepts and random slopes (Muthen and Muthen 2000).

Next, we applied piecewise growth mixture modeling to combine child and adolescent data. This type of growth mixture model can link two intervals that consist of several observed time points each. Importantly, piecewise growth modeling permits a change of slope, and, if

Europe PMC Funders Author Manuscripts

necessary, also a change in intercept between the two intervals. Therefore, piecewise growth modeling can capture an expected trajectory that starts increasing at the beginning of adolescence. Piecewise growth modeling is also suitable to handle the fact that the two intervals reflected a change in rater (mother vs. self-ratings). We compared piecewise models with increasing numbers of classes that only allowed for a change of slope to models that also allowed for a change in intercept. The first type of model would support that rater differences have a negligible impact on modeling continuous trajectories over the two intervals whereas models with a separate intercept for the second interval imply a discontinuity which can at least in part be due to rater differences.

All models were fitted with a large number of random starts. Models were considered as properly converged if three conditions were met: (1) at least 4 sets of starting values converged to the same maximum, (2) all parameter estimates were within their proper range (e.g., variance estimates >0), and (3) model estimation resulted in a positive definite information matrix.

In sum, the goals of the growth mixture analyses were (1) to identify the numbers of classes that best describe the longitudinal patterns of SxAnxDep, (2) to decide whether random intercepts and slopes were necessary to describe the structure in the data, and (3) to evaluate differences in developmental trajectories between the child and adolescent intervals.

2. Genetics analyses

The type of longitudinal model for the genetic analyses was chosen based on the results of the growth mixture analysis. We estimated additive genetic, and shared and non-shared environmental effects on the latent growth factors using standard twin modeling (Boomsma et al. 2002). Twin modeling utilizes the fact that monozygotic twins share the same genes whereas dizygotic twins are expected to share half of their segregating genes. Consequently, If MZ twin pairs are more similar for a trait than DZ twin pairs, this suggests that genetic factors influence this trait, for example when correlations in MZ pairs are 0.6 and in DZ pairs 0.3. If the correlation in DZ pairs is more than half the correlation in MZ twins (e.g. rMZ = 0.6 and rDZ=0.5), then there is additional familial resemblance which is not explained by genetic factors. Such factors are commonly referred to as shared or common environmental factors. The importance of individual-specific environmental factors is indicated by the differences within MZ twin pairs. Incorporating these expectations in a latent growth model permits estimating the percentages of variance explained by additive genetic effects, common environmental effects shared by children growing up in the same family, and non-shared environmental effects that contribute to the total variance of the latent growth factors.

Results

1. Growth Mixture Modeling for mother and self-ratings, and males and females

In all four groups, linear growth models with random intercepts for a single class emerged as the preferred model. Specifically, based on the BIC, models without random effects could be rejected. Likelihood ratio tests comparing single class models with random intercepts and

fixed slopes to models with random intercepts and random slopes were not significant for either male or female mother ratings, (mother/female: p=0.06, mother/male: p=1), but significant for self-ratings (self/female p=1.57e-5, self/male: p=0.036). However, the random slope effect explained essentially zero percent of the total variance in either model. Therefore, we rejected the need for random slopes. Extending the separate growth models to more than a single class did not result in proper convergence in either group. Model fitting results and more specific information regarding failure of proper convergence are presented in Tables 4 and 5 for models fitted to male and female data, respectively.

In the preferred single class linear growth models with random intercepts, for males, mother ratings during childhood showed a *higher* intercept mean compared to self-ratings during adolescence (mean mother rating = 0.692, SE=0.045 *vs.* mean self = 0.253, SE=0.100). For females, mother ratings had a *lower* average baseline compared to self-ratings (mean mother rating = 0.850, SE=0.44 *vs.* mean self = 1.287, SE=0.088). The difference in intercept between childhood and adolescence indicates that it might be necessary to account for a discontinuity when combining mother and self-ratings in a single model.

The linear growth model splits the total variance in the developmental course of symptoms into common factors (i.e., intercept and slope factors) and residual variance. The intercept factor represents the baseline, whereas the slope factor represents the common linear trend over time. Residual variance captures the more idiosyncratic differences in symptom course that cannot be described by the common intercept and slope factors.

In our analyses, the intercept variance contributed considerably to the total variance of the phenotype, both in childhood and adolescence, namely about 55% in males and about 58% in females. This means that a large part of the variability in symptom course is due to individual differences in baseline. The remaining variability was essentially idiosyncratic, with the slope factor variance being essentially zero in males and females. The slope mean was also zero in both time intervals for males, which means that there was no significant common average change. For females, the slope mean, while insignificant during childhood, was mildly positive during adolescence (mean=0.087, SE=0.022).

Taken together these results imply that a substantial part of the individual trajectories in SxAnxDep are in fact idiosyncratic, and cannot be easily disaggregated into different latent growth classes, or described in terms of a common linear development structure. Instead, the results point to age specific influences on the individual trajectories possibly including temporary effects of events and emotions.

2. Simultaneous analysis of mother and self-ratings

We compared single and 2-class models fitted to mother and self-ratings jointly. Although in the separate models we were unable to properly identify 2 classes, joining the data across all ages can help to identify multiple classes with characteristic developmental curves. More specifically, we compared piecewise growth models that only allow the slope to change between mother and self-ratings to models that allowed for a change in intercept as well. An intercept (or baseline) change is a discontinuity in average trajectories that can at least in part accommodate differences between mother and self-ratings. In case of a single class

model such a model is equivalent to fitting linear growth models separately to mother and self-ratings.

The results confirmed the findings of modeling child and adolescent data separately, namely that it was necessary to include a change in baseline when shifting from mother to self-ratings. This was true for observations from both males and females. When extending the single class version of this model to two classes, we found that the BIC was in fact lower for the 2-class model in the males and females. However, the difference in BIC with the single class model was small (see Tables 4 and 5). Bootstrapping a confidence interval for the BIC showed that the difference in BIC was not substantial. When evaluating the parameters of the 2-class models it was clear that for both males and females the two classes were basically dividing individuals into groups with a slightly higher baseline vs. a lower baseline (see Table 6). The entropy in these 2-class models was very low, which indicates that assigning individuals to either class was problematic (see Tables 4 and 5). We therefore concluded that the single class models were preferable. The parameter estimates of the intercept and slope factor means in the joint single class models that permitted a separate intercept for the self-ratings were as expected very similar to the separate single class models fitted initially (see Table 6).

In sum, the results of the growth mixture analyses indicate that combining mother and self-ratings of SxAnxDep does not add much information, that our data support a single class, and that the intercept or baseline factor explains a substantial amount of total variance whereas the slope factor does not. Therefore, in the following twin analyses, we fitted models for mother and self-ratings separately, and focused on the heritability of the intercept factor.

3. Genetic analyses

We fitted single class linear growth models to data from monozygotic (MZ) and dizygotic (DZ) twins. The variance of the intercept factor was decomposed into variance components of additive genetic variance, shared environmental variance, and non-shared/error variance. We did not decompose the slope variance since individual differences in the slope factor were small. The same model was fitted to mother and self-reports separately based on the longitudinal model results.

For all models we provide likelihood based confidence intervals for the estimates of the variance components (see Table 7). As can be seen, narrow-sense heritability is very similar for males and females, and is slightly higher in adolescent self-reports compared to childhood mother reports. The percentages of variance explained by additive genetic effects on the intercept of the maternal ratings during childhood were 71% and 63% for males and females, respectively. These percentages on the intercept of the self-ratings during adolescence were 83% and 84% respectively. Shared environment explained almost all of the remaining variance during childhood (23% and 31%, for males and females, respectively), implying that non-shared environmental effects were close to zero. As might be expected, the non-shared variance increased during adolescence and was estimated at 16% for males and 12% for females. This was at the cost of shared environmental effects that had a zero effect for males during adolescence and a close to zero effect for females.

Importantly, narrow sense heritability of baseline stability of anxious depression across age is considerably higher than at any given age. Age specific heritabilities were 43% (7 year olds), 48% (10 year olds), 46% (12 year olds), 54% (14 year olds), 51% (16-year olds), and 47% (18-year olds). We also compared the heritability of baseline stability in a growth model to the heritability of a simple average of the three ages during childhood and adolescence, respectively. The heritability of the average scores was also considerably lower, and closer to the heritabilities at each time point, namely 55% and 46%.

Discussion

GWA studies of anxiety and depression phenotypes have been largely unsuccessful so far. Possible reasons include phenotypic heterogeneity and the relatively moderate heritability (Levinson et al. 2014). These two factors are interrelated, however, and the results of our study show that the heritability of the phenotype can be substantially increased by reducing phenotypic heterogeneity.

The latent growth curve model separates individual variability into variance due to common factors (intercept and slope) and age specific residual variance such as temporary fluctuations and measurement error. The intercept factor captures individual variability that is common over time, and is therefore interpretable as baseline stability. Consequently, growth curve modeling can serve to extract a baseline measurement of SxAnxDep. Our study shows that this latent phenotype is substantially more heritable than the scores observed at any age, namely between 72% (childhood) and 83% (adolescence) for males and 64% (childhood) and 84% (adolescence) for females. A measure created by simply summing data from the different ages was also considerably less heritable. This gives further evidence that using a latent variable modeling approach significantly improves our ability to detect the role of additive genetic effects over simpler approaches. The high heritability of baseline stability is in line with the results of Nivard et al (2015) who found that stability in SxAnxDep was mostly influenced by genetic factors. Since baseline stability does not contain measurement error, it is more reliable, and likely a more informative phenotype for GWA studies.

In our study, there was considerable heterogeneity in individual trajectories over time both in the interval measured by mother ratings when children were 7, 10, and 12 year old, and in the interval measured by self-ratings when they were 14, 16, and 18 year old. However, these individual differences could not be clearly grouped into a number of latent classes. Previous studies have identified more classes (six at maximum) to describe individual differences in the developmental course of symptoms of anxiety and depression during childhood and adolescence (Brendgen et al. 2005; Broeren et al. 2013; Cote et al. 2002; Crocetti et al. 2009; Dekker et al. 2007; Duchesne et al. 2008 ; Fanti and Henrich 2010; Feng et al. 2008; Legerstee et al. 2013; Letcher et al. 2012; Letcher et al. 2009; Marmorstein et al. 2010; Morin et al. 2011; Nivard et al. submitted; Rodriguez et al. 2005; Sterba et al. 2007; Toumbourou et al. 2011). However, all, but three of these studies did not include random effects in the models (Crocetti et al. 2009; Morin et al. 2011; Nivard et al. submitted). This can explain the higher number of identified classes. In one of the studies that also modeled random effects, the results were very similar to ours (Crocetti et al. 2009).

They identified two classes, which only separated individuals with generally higher scores from the lower scoring majority. This type of differentiation is unlikely to improve power in GWA studies since the two classes basically reflect a categorization of the phenotype. The other two studies did identify 4 or 5 classes, but there were notable differences in methods that could explain these discrepancies. Morin et al (2011) analyzed seven time points and also included a quadratic term. Moreover, the time between the measurements was shorter (less than one year). Measurements in our data were separated by 2-3 years. Shorter measurement intervals and a larger number of observations across time might provide a better resolution to detect specific trajectory classes with characteristic developmental growth curves.

The way the phenotype is defined and measured can also have a substantial impact on finding latent classes in a phenotypic analysis. In Nivard et al (submitted) the phenotype was constructed by grouping individuals according to their chance of having a DSM-IV anxiety or depressive disorder. This type of measure is based on severity, and eliminates individual differences within each severity category. Such an operationalization of anxiety and depression favors the detection of latent classes compared to using a sum score of individual items because part of the individual differences are averaged out prior to fitting mixture models, and individuals are already grouped by severity. Furthermore, data were available at 4 ages which provides a better basis to detect average trajectories over time. In our study, a single class linear growth model with a random intercept was the best fitting model. Unfortunately we were unable to integrate mother and self-ratings smoothly in a single longitudinal trajectory model because of a discontinuity induced by rater differences, and instead modeled the two time intervals separately.

In sum, using growth curve modeling, we leveraged information about the phenotype at several ages, which cancels out age specific and measurement error variance, and therefore results in a more informative and reliable phenotype indicating baseline stability. We did not detect clear trajectories that group individuals into different classes, such as stable high in SxAnxdep. This signifies that it is difficult to predict developmental patterns during childhood and adolescence.

Our twin analyses clearly demonstrate the advantage of longitudinal data for genetic studies. Although it might be of interest to investigate age specific genetic effects on anxiety and depression, a promising first step would be to use the baseline stability as a more reliable and heritable phenotype in GWA studies. These results are consistent with our recent study using a different measure, the Hospital Anxiety and Depression Scale (HADS) (Laurin et al. 2015). Using latent variable methods we selected items from the HADS that permitted a narrow and less heterogeneous phenotype definition. The resulting higher phenotype reliability in turn improved statistical power in Genomic Complex Trait Analyses (GCTA) and GWA analyses (Laurin et al. 2015). The benefit of more narrowly defined phenotypes in gene-finding studies was also evident in a study of borderline personality (Lubke et al. 2014). For genome-wide association meta-analyses, we therefore recommend to establish more reliable phenotypes. One way to achieve this is to leverage longitudinal data and study the effects of genetic variants on baseline stability.

Limitations

A main limitation of our study was that a change in rater occurred between ages 12 and 14, and that there was an insufficient number of participants with mother and self-ratings at age 12 to directly model rater differences in a growth mixture model. Mother and self-ratings each consisted of three measurements, permitting only linear but not curvilinear growth trajectories. The combination of two linear intervals provided a basis to investigate slope differences between childhood and adolescence, thus theoretically permitting non-linear growth over the entire period. However, the rater differences in our data limited the options to smoothly combine childhood and adolescent trajectories and detect average longitudinal developmental patterns. Consequently, the growth modeling was done separately for three ages at childhood and at adolescent data. A quadratic term could not be included. As a result, the twin modeling was limited to estimating the genetic contributions to baseline stability of SxAnxDep in each interval separately. The analyses showed, however, that three time points are sufficient to extract a highly heritable phenotype that is likely to provide increased statistical power in GWAS. Although our study focused on SxAnxDep, we expect similar benefits of leveraging longitudinal data for other psychiatric symptom scales.

Acknowledgements

The first author was supported by DA-018673. The second author is funded through the NSF Graduate Research Fellowship Program under Grant No. 1313583. Further funding was provided by the Netherlands Scientific Organization (NWO) (912-100-20): “Genetic influences on stability and change in psychopathology from childhood to young adulthood”, and the European Research Council (ERC-230374): “Genetics of Mental Illness”. Data collection has been funded by multiple grants from NWO, NIMH and the Neuroscience Campus Amsterdam (NCA).

References

Achenbach, TM.; Rescorla, LA. Manual for the ASEBA School-Age Forms & Profiles. Burlington VT: University of Vermont, Research Center for Children, Youth, & Families; 2001.

Beesdo-Baum K, Knappe S. Developmental epidemiology of anxiety disorders. Child Adolesc Psychiatr Clin N Am. 2012; 21:457–478. [PubMed: 22800989]

Bellina M, Brambilla P, Garzitto M, Negri GA, Molteni M, Nobile M. The ability of CBCL DSM-oriented scales to predict DSM-IV diagnoses in a referred sample of children and adolescents. Eur Child Adolesc Psychiatry. 2013; 22:235–246. [PubMed: 23138539]

Benke KS, Nivard MG, Velders FP, Walters RK, Pappa I, Scheet PA, Xiao X, Ehli EA, Palmer LJ, Whitehouse AJ, Verhulst FC, et al. Lifecourse Epidemiology C. A genome-wide association meta-analysis of preschool internalizing problems. J Am Acad Child Adolesc Psychiatry. 2014; 53:667–676 e667. [PubMed: 24839885]

Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. Nat Rev Genet. 2002; 3:872–882. [PubMed: 12415317]

Boomsma DI, de Geus EJ, Vink JM, Stubbe JH, Distel MA, Hottenga JJ, Posthuma D, Van Beijsterveldt TC, Hudziak JJ, Bartels M, Willemsen G. Netherlands Twin Register: from twins to twin families. Twin Res Hum Genet. 2006; 9:849–857. [PubMed: 17254420]

Brendgen M, Wanner B, Morin AJS, Vitaro F. Relations with parents and with peers, temperament, and trajectories of depressed mood during early adolescence. J Abnorm Child Psychol. 2005; 33:579–594. [PubMed: 16195952]

Broeren S, Muris P, Diamantopoulou S, Baker JR. The course of childhood anxiety symptoms: developmental trajectories and child-related factors in normal children. J Abnorm Child Psychol. 2013; 41:81–95. [PubMed: 22836287]

Europe PMC Funders Author Manuscripts

Copeland WE, Shanahan L, Costello EJ, Angold A. Childhood and adolescent psychiatric disorders as predictors of young adult disorders. Arch Gen Psychiatry. 2009; 66:764–772. [PubMed: 19581568]

Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry. 2011; 52:1015–1025. [PubMed: 21815892]

Cote S, Tremblay RE, Nagin D, Zoccolillo M, Vitaro F. The development of impulsivity, fearfulness, and helpfulness during childhood: patterns of consistency and change in the trajectories of boys and girls. J Child Psychol Psychiatry. 2002; 43:609–618. [PubMed: 12120857]

Crocetti E, Klimstra T, Keijsers L, Hale WW, Meeus W. Anxiety Trajectories and Identity Development in Adolescence: A Five-wave Longitudinal Study. J Youth Adolescence. 2009; 38:839–849.

Dekker MC, Ferdinand RF, van Lang ND, Bongers IL, van der Ende J, Verhulst FC. Developmental trajectories of depressive symptoms from early childhood to late adolescence: gender differences and adult outcome. J Child Psychol Psychiatry. 2007; 48:657–666. [PubMed: 17593146]

Duchesne S, Vitaro F, Larose S, Tremblay RE. Trajectories of anxiety during elementary-school years and the prediction of high school noncompletion. J Youth Adolescence. 2008; 37:1134–1146.

Erhardt A, Czibere L, Roeske D, Lucae S, Unschuld PG, Ripke S, Specht M, Kohli MA, Kloiber S, Ising M, Heck A, et al. TMEM132D, a new candidate for anxiety phenotypes: evidence from human and mouse studies. Mol Psychiatry. 2011; 16:647–663. [PubMed: 20368705]

Fanti KA, Henrich CC. Trajectories of pure and co-occurring internalizing and externalizing problems from age 2 to age 12: findings from the National Institute of Child Health and Human Development Study of Early Child Care. Dev Psychol. 2010; 46:1159–1175. [PubMed: 20822230]

Feng X, Shaw DS, Silk JS. Developmental trajectories of anxiety symptoms among boys across early and middle childhood. J Abnorm Psychol. 2008; 117:32–47. [PubMed: 18266484]

Group PGCBDW. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near ODZ4. Nat Genet. 2011; 43:977–983. [PubMed: 21926972]

Kendler KS, Aggen SH, Knudsen GP, Roysamb E, Neale MC, Reichborn-Kjennerud T. The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. Am J Psychiatry. 2011; 168:29–39. [PubMed: 20952461]

Laurin CA, Hottenga JJ, Willemsen G, Boomsma DI, Lubke GH. Genetic Analyses Benefit From Using Less Heterogeneous Phenotypes: An Illustration With the Hospital Anxiety and Depression Scale (HADS). Genet Epidemiol. 2015; 39:317–24. [PubMed: 25832296]

Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, Perlis RH, Mowry BJ, Thapar A, Goddard ME, Witte JS, Absher D, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013; 45:984–94. [PubMed: 23933821]

Legerstee JS, Verhulst FC, Robbers SC, Ormel J, Oldehinkel AJ, van Oort FV. Gender-Specific Developmental Trajectories of Anxiety during Adolescence: Determinants and Outcomes. The TRAILS Study. J Can Acad Child Adolesc Psychiatry. 2013; 22:26–34. [PubMed: 23390430]

Letcher P, Sanson A, Smart D, Toumbourou JW. Precursors and Correlates of Anxiety Trajectories From Late Childhood to Late Adolescence. J Clin Child Adolesc Psychol. 2012; 41:417–432. [PubMed: 22551395]

Letcher P, Smart D, Sanson A, Toumbourou JW. Psychosocial Precursors and Correlates of Differing Internalizing Trajectories from 3 to 15 Years. Social Development. 2009; 18:618–646.

Levinson DF, Mostafavi S, Milaneschi Y, Rivera M, Ripke S, Wray NR, Sullivan PF. Genetic studies of major depressive disorder: why are there no genome-wide association study findings and what can we do about it? Biol Psychiatry. 2014; 76:510–512. [PubMed: 25201436]

Lubke G, Neale MC. Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? Multivariate Behavioral Research. 2006; 41:499–532. [PubMed: 26794916]

Lubke GH, Hottenga JJ, Walters R, Laurin C, de Geus EJ, Willemsen G, Smit JH, Middeldorp CM, Penninx BW, Vink JM, Boomsma DI. Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. BiolPsychiatry. 2012; 72:707–709.

Lubke GH, Laurin C, Amin N, Hottenga JJ, Willemsen G, van Grootheest G, Abdellaoui A, Karssen LC, Oostra BA, van Duijn CM, Penninx BW, et al. Genome-wide analyses of borderline personality features. Mol Psychiatry. 2014; 19:923–929. [PubMed: 23979607]

Marmorstein NR, White H, Chung T, Hipwell A, Stouthamer-Loeber M, Loeber R. Associations between first use of substances and change in internalizing symptoms among girls: differences by symptom trajectory and substance use type. J Clin Child Adolesc Psychol. 2010; 39:545–558. [PubMed: 20589565]

Merikangas KR, Nakamura EF, Kessler RC. Epidemiology of mental disorders in children and adolescents. Dialogues Clin Neurosci. 2009; 11:7–20. [PubMed: 19432384]

Middeldorp CM, Cath DC, Van Dyck R, Boomsma DI. The co-morbidity of anxiety and depression in the perspective of genetic epidemiology. A review of twin and family studies. PsycholMed. 2005; 35:611–624.

Morin AJS, Maiano C, Nagengast B, Marsh HW, Morizot J, Janosz M. General Growth Mixture Analysis of Adolescents' Developmental Trajectories of Anxiety: The Impact of Untested Invariance Assumptions on Substantive Interpretations. Structural Equation Modeling-a Multidisciplinary Journal. 2011; 18:613–648.

Muthen B, Muthen LK. Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. Alcohol Clin Exp Res. 2000; 24:882–891. [PubMed: 10888079]

Nagin DS. Analyzing developmental trajectories: A semiparametric, group-based approach. Psychol Methods. 1999; 4:139–157.

Nivard MG, Dolan CV, Kendler KS, Kan KJ, Willemsen G, van Beijsterveldt CE, Lindauer RJ, van Beek JH, Geels LM, Bartels M, Middeldorp CM, et al. Stability in symptoms of anxiety and depression as a function of genotype and environment: a longitudinal twin study from ages 3 to 63 years. Psychol Med. 2015; 45:1039–1049. [PubMed: 25187475]

Nivard MG, Lubke GH, Dolan CV, Evans DM, St Pourcain B, Munafo MR, Middeldorp CM. Joint developmental trajectories of internalizing and externalizing disorders between childhood and adolescence. submitted.

Otowa T, Kawamura Y, Nishida N, Sugaya N, Koike A, Yoshida E, Inoue K, Yasuda S, Nishimura Y, Liu X, Konishi Y, et al. Meta-analysis of genome-wide association studies for panic disorder in the Japanese population. Transl Psychiatry. 2012; 2:e186. [PubMed: 23149450]

Rhee SH, Lahey BB, Waldman ID. Comorbidity Among Dimensions of Childhood Psychopathology: Converging Evidence from Behavior Genetics. Child Dev Perspect. 2015; 9:26–31. [PubMed: 26019716]

Rodriguez D, Moss HB, Audrain-McGovern J. Developmental heterogeneity in adolescent depressive symptoms: Associations with smoking behavior. Psychosom Med. 2005; 67:200–210. [PubMed: 15784784]

Sterba SK, Prinstein MJ, Cox MJ. Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. Dev Psychopathol. 2007; 19:345–366. [PubMed: 17459174]

Toumbourou JW, Williams I, Letcher P, Sanson A, Smart D. Developmental trajectories of internalising behaviour in the prediction of adolescent depressive symptoms. Aust J Psychol. 2011; 63:214–223.

van Beijsterveldt CEM, Groen-Blokhuis M, Hottenga JJ, Franic S, Hudziak JJ, Lamb D, Huppertz C, de Zeeuw E, Nivard M, Schutte N, Swagerman S, et al. The Young Netherlands Twin Register (YNTR): Longitudinal Twin and Family Studies in Over 70,000 Children. Twin Res Hum Gen. 2013; 16:252–267.

van Lang ND, Ferdinand RF, Oldehinkel AJ, Ormel J, Verhulst FC. Concurrent validity of the DSM-IV scales Affective Problems and Anxiety Problems of the Youth Self-Report. Behav Res Ther. 2005; 43:1485–1494. [PubMed: 16159590]

Verhulst, FC.; Ende, Jv; Koot, HM. Handleiding voor de Youth Self Report. Rotterdam: Afdeling Kinder- en Jeugdpsychiatrie, Sophia Kinderziekenhuis / Academisch Ziekenhuis Rotterdam/ Erasmus Universiteit Rotterdam: 1997.

Walter S, Glymour MM, Koenen K, Liang L, Tchetgen Tchetgen EJ, Cornelis M, Chang SC, Rimm E, Kawachi I, Kubzansky LD. Performance of polygenic scores for predicting phobic anxiety. PLoS One. 2013; 8:e80326. [PubMed: 24278274]

Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, Ripke S, Macintyre DJ, McGhee KA, Maclean AW, Smit JH, et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol Psychiatry. 2010; 18:497–511. [PubMed: 22472876]

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1a –

Number of individual twins per age for the mixture trajectory analyses

	Age					
Zygosity	**7**	**10**	**12**	**14**	**16**	**18**
MZM	2258	2211	2067	366	484	450
DZM	2152	2098	1863	329	354	372
MZF	2586	2619	2373	663	778	865
DZF	2054	1989	1787	438	509	553
DZO	4282	4238	3698	776	884	861

Participants with measurements on at least two occasions between age 7-12 or between age 14-18 were selected for the mixture trajectory analyses. MZM(F) = Monozygotic Male (Female) twin; DZM(F) = Dizygotic Male (Female) twin; DZO = Dizygotic opposite sex twin.

Europe PMC Funders Author Manuscripts

Table 1b -

Number of twin pairs (% complete pairs) by age in the genetic analyses

	Age					
Zygosity	**7**	**10**	**12**	**14**	**16**	**18**
MZM	1510 (99)	1333 (99)	1147 (98)	519 (82)	647 (81)	468 (78)
DZM	1595 (99)	1281 (99)	1043 (99)	497 (76)	512 (77)	403 (71)
MZF	1721 (99)	1556 (99)	1292 (99)	796 (85)	880 (83)	838 (78)
DZF	1461 (99)	1226 (99)	1005 (99)	642 (83)	671 (75)	552 (72)
DZO	2992 (99)	2620 (99)	2075 (99)	1196 (71)	1289 (69)	957 (65)

MZM(F) = Monozygotic Male (Female) twin pair; DZM(F) = Dizygotic Male (Female) twin pair; DZO = Dizygotic opposite sex twin pair.

Europe PMC Funders Author Manuscripts

Table 2 -

Endorsement rates (%) for the categorized sum score for males/ females at each time point

	Age					
Category	**7**	**10**	**12**	**14**	**16**	**18**
0	33/ 30	31/ 30	37/ 34	43/ 26	39/ 22	37/ 21
1	36/ 37	34/ 34	33/ 35	26/ 23	27/ 23	27/ 21
2	24/ 25	24/ 25	21/ 22	21/ 30	22/ 31	23/ 29
3	08/ 09	11/ 11	09/ 10	10/ 22	12/ 25	12/ 29

Europe PMC Funders Author Manuscripts

Table 3 -

Polychoric Correlations between ages for males (lower) and females (upper).

Age	7	10	12	14	16	18
7		.70	.65	.40	.29	.23
10	.79		.74	.36	.29	.19
12	.52	.80		.43	.35	.20
14	.08	.27	.19		.70	.54
16	.17	.34	.27	.60		.71
18	.00	.15	.35	.26	.50	

Table 4 -

Latent Growth Model Fit in Males

Model	# Cl	# Params	LogLik	BIC	Entropy	Class Percentages
Mother Reports (Age 7-12)						
1. No Random Effects	1	4	-24172	48379		
	2	7	-22520	45103	.66	60, 40
	3	10	-22169	44427	.61	46, 36, 18
	4	13	-22121	44359	.63	45, 29, 22, 04
	5	Failed to replicate likelihood				
2. Random Intercept	**1**	**5**	**-22189**	**44423**		
	2	Failed to converge				
3. Random Intercept and Slope	1	6	-22189	44431		
	2	Failed to replicate likelihood				
Self-Reports (Age 14-18)						
4. No Random Effects	1	4	-4502	9033		
	2	7	-4317	8687	.60	67, 33
	3	10	-4287	8650	.50	43, 42, 15,
	4	13	-4283	8663	.42	35, 28, 21, 16
5. Random Intercept	**1**	**5**	**-4291**	**8619**		
	2	Failed to replicate likelihood				
6. Random Intercept and Slope	1	6	-4288	8622		
	2	Failed to replicate likelihood				
Piecewise Models (Age 7-18)						
7. Random intercept	**1**	**9**	**-29596**	**59273**		
	2	15	-29545	59224	.4	76, 24
*8. Rater Mean Difference = 0**	1	8	-29612	59297		

** $\Delta\chi 2=25.4$, df=1, p = 4.5E-8*

Latent class growth models for maternal and self-ratings separately, and jointly using a piecewise model. Models were estimated with an increasing number of classes (Cl). Shown are the number of estimated parameters in the model, the final log likelihood value, the Bayesian information criterion (BIC), Entropy, and class proportions. Models with multiple classes did not have significantly lower BIC than models with a single class (bold). For the piecewise models, we selected a single class model because entropy was low indicating poor class assignment. Secondly, the BIC of the two class model fell within 1 SD of bootstrapped BIC for the single class model (BIC SD = 1142). Model 8 tested whether the second intercept was necessary to model the discontinuity between raters at ages 12 and 14. This was done by testing whether the mean of the second intercept could be set equal to 0. This model fit significantly worse than the unconstrained model, indicating a mean difference between raters.

Table 5 -

Latent Growth Model Fit in Females

Model	# Cl	# Params	LogLik	BIC	Entropy	Class Percentages
Mother Reports (Age 7-12)						
1. No Random Effects	1	4	-25320	50675		
	2	7	-23645	47353	.65	58, 42
	3	10	-23289	46668	.61	46, 37, 17
	4	13	-23250	46616	.54	42, 27, 21, 10
	5	16	-23232	46607	.54	41, 22, 19, 10, 8
	6	Failed to converge				
2. Random Intercept	**1**	**5**	**-23302**	**46648**		
	2	Failed to replicate likelihood				
3. Random Intercept and Slope	1	6	-23301	46657		
	2	Failed to replicate likelihood				
Self-Reports (Age 14-18)						
4. No Random Effects	1	4	-7254	14540		
	2	7	-6894	13843	.59	51, 49
	3	10	-6823	13724	.54	48, 27, 25
	4	13	-6812	13727	.52	37, 27, 25, 11
5. Random Intercept	**1**	**5**	**-6824**	**13688**		
	2	15	-7600	15319	.36	59, 41
6. Random Intercept and Slope	1	6	-6819	13685		
	2	17	-7586	15307	.40	62, 38
Piecewise Models (Age 7-18)						
7. Random intercept	**1**	**9**	**-34056**	**68294**		
	2	15	-34000	68136	.31	.61, .39
8. Rater Mean Difference = 0*	1	8	-34190	68453		

** χ^2=216.4, df=1, p < 1E-16*

Latent class growth models for maternal and self-ratings separately, and jointly using a piecewise model. Models were estimated with an increasing number of classes (Cl). Shown are the number of estimated parameters in the model, the final log likelihood value, the Bayesian information criterion (BIC), Entropy, and class proportions. Models with multiple classes did not have significantly lower BIC than models with a single class (bold). For the self-reports, a likelihood ratio test showed that the random slope (6) did not improve fit, so model (5) was chosen. For the piecewise models, we selected a single class model because entropy was low indicating poor class assignment. Secondly, the BIC of the two class model fell within 1 SD of bootstrapped BIC for the single class model (BIC SD = 1264). Model 8 tested whether the second intercept was necessary to model the discontinuity between raters at ages 12 and 14. This was done by testing whether the mean of the second intercept could be set equal to 0. This model fit significantly worse than the unconstrained model, indicating a mean difference between raters.

Europe PMC Funders Author Manuscripts

Table 6:

Factor Mean Estimates in the Piecewise model

Model Name	Class (%)	Intercept (SE)	Slope (SE)	Intercept (SE)	Slope (SE)
		Age 12	Age 7-12	Age 12-14	Age 14-18
Male					
Random intercept					
1 Class	1	.64 (.04)	.00 (.01)	-.39 (.08)	.02 (.03)
2 Class	76	-.42 (.14)	-.09 (.02)	.32 (.18)	.17 (.06)
	24	3.2 (.30)	.18 (.04)	-2.1 (.29)	-.34 (.15)
Female					
Random intercept					
1 Class	1	.76 (.04)	-.01 (.01)	.99 (.07)	.10 (.02)
2 Class	61	-.37 (.13)	-.08 (.02)	1.6 (.15)	.34 (.10)
	39	2.7 (.37)	.09 (.03)	.03 (.27)	-.26 (.10)

Piecewise intercept and slope estimates (standard errors, SE) for males and females for the single and two class models. The first intercept was centered at age 12, and the second intercept is the mean difference between ages 12 and 14. Estimates are on the standard normal liability scale.

Europe PMC Funders Author Manuscripts

Table 7:

Latent Growth Decomposition (95% confidence intervals) of stability of symptoms of anxiety and depression in childhood and adolescence

Gender	Age	A	C	E
Male	7 - 12	71 (60-83)	23 (12-29)	6 (02-09)
Female	7 - 12	63 (53-75)	31 (30-41)	5 (02-06)
Male	14 - 18	83 (67-92)	0 (0-13)	16 (07-25)
Female	14 - 18	84 (63-94)	3 (0-22)	12 (06-18)

A: Additive genetic, C: common environment, E: non-shared environment + measurement error