JOINT TRAJECTORIES OF INTERNALIZING AND EXTERNALIZING PROBLEMS IN EARLY CHILDHOOD: TESTING THEORIES OF SYMPTOM COVARIATION

Sonya Kourany Sterba

A thesis submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology.

Chapel Hill
2005

Approved by
Advisor: Mitchell J. Prinstein
Reader: Daniel J. Bauer
Reader: Martha J. Cox
ABSTRACT

SONYA KOURANY STERBA: Joint Trajectories of Internalizing and Externalizing Problems in Early Childhood: Testing Theories of Symptom Covariation (Under the direction of Mitchell J. Prinstein)

Covariation between internalizing and externalizing problems is typically summarized using correlations. This assumes the same covariation pattern holds for all children, but there are empirical and theoretical rationales for qualitatively different patterns over time. In this study, growth mixture modeling of 8 waves of mother-reported internalizing and externalizing symptoms (ages 2-11; N=1,222) from the NICHD Study of Early Child Care were used to test four theories of internalizing/externalizing covariation. Growth mixture modeling probabilistically links developmental courses of internalizing and externalizing symptoms, while capturing heterogeneity in the expression of each. Results showed that stable covariation occurred at all levels of behavior (high, moderate, and low), for all functional forms of behavior (decreasing, persisting, or low-stable), and for the majority of children. Unique patterns of differentiation with development (among one fifth of boys) and rising internalizing (among one fifth of girls) evidenced in this study would have been obscured using correlational analyses.
TABLE OF CONTENTS

LIST OF TABLES ......................................................................................................................... ix
LIST OF FIGURES ...................................................................................................................... x

Chapter

I INTRODUCTION ..................................................................................................................... 1
  Development of Externalizing Problems ................................................................. 4
  Development of Internalizing Problems ............................................................... 9
  Externalizing and Internalizing Problem Covariation ........................................ 13
  Five Theoretical Predictions of Internalizing-Externalizing Covariation ........... 14
    Hypothesis #1: Differentiation with Development .............................................. 15
    Hypothesis #2: Internalizing Subtypes ............................................................... 16
    Hypothesis #3: Rising Internalizing ................................................................. 18
    Hypothesis #4: Covariation Stability ............................................................... 20
    Hypothesis #5: Gender differences ................................................................. 21
  Growth Mixture Modeling: A Conceptual Description ....................................... 22
  When is GMM an Appropriate Method of Analysis? ........................................... 26
  What Population Inferences Can and Cannot be Drawn from GMM? .................. 27
  Rationale for Choosing Nagin’s SPGMM over Muthén’s GMM ......................... 29
  Compatibility of Data and SPGMM Method of Analysis ..................................... 31
II METHODS

Participants
Overview of Data Collection
Measures
Demographic Characteristics
Child Behavior Checklist
Analyses
Diagnostics of Model Adequacy

III RESULTS

Correlational Analyses
Semiparametric Growth Mixture Analyses
Model fitting: Four Univariate Models
Number and Order of Trajectory Groups
Diagnostics of Model Adequacy
Univariate Internalizing and Externalizing Models
Internalizing Symptoms: Univariate Models
Low Symptom Groups
Persistors
Decreasers
Internalizing Symptoms: Correlative versus SPGMM Results
DISCUSSION

Efficacy of Group-Based Modeling versus Traditional Approaches

Group-Based Modeling Replicates and Extends Results

Groups as a Heuristic Device: Benefits

Patterns of Externalizing Symptoms from 2-11 Years

Key Features from Prior GMM Studies Replicated

Externalizing Patterns Differed from Prior GMM in Four Ways

Patterns of Internalizing Symptoms from 2-11 Years

Gender Differences in Longitudinal Patterns of Internalizing/Externalizing

Similar Developmental Pathways across Gender

Externalizing: Gender Differences in Intercepts and Slopes

Internalizing: Gender Divergence over Time

Gender Difference in Univariate Models: Theoretical Implications

Demonstrated Support for Five Main Theories of Covariation

Differentiation with Development

Internalizing Subtypes

Rising Internalizing

Covariation Stability

Moderate, not Extreme Behavior Modulable

Tracking of Symptom Covariation

Implications for Screening Efficacy, Detection

Limitations

Response Variable
<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. SPGMM Trajectory Groups for School-Age to Adolescent, Preschool to School-Age Samples</td>
<td>107</td>
</tr>
<tr>
<td>2. SPGMM Trajectory Groups for Late Adolescent Samples</td>
<td>109</td>
</tr>
<tr>
<td>3. Probabilities of Trajectory Membership Corresponding with Hypothesized Patterns</td>
<td>110</td>
</tr>
<tr>
<td>4. Jeffrey’s Evidence Scale for Bayes Factors</td>
<td>111</td>
</tr>
<tr>
<td>5. Invariant Internalizing, Externalizing Symptoms Across CBCL 2/3 and CBCL 4-18</td>
<td>112</td>
</tr>
<tr>
<td>6. Intercorrelations for the Form-Invariant Internalizing and Externalizing CBCL items</td>
<td>113</td>
</tr>
<tr>
<td>7. Boys’ Externalizing Symptom Lag+1 Correlations from 2-11 years (N= 598)</td>
<td>114</td>
</tr>
<tr>
<td>8. Boys’ Internalizing Symptom Lag +1 Correlations from 2-11 years (N= 598)</td>
<td>115</td>
</tr>
<tr>
<td>9. Girls’ Externalizing Symptom Lag +1 Correlations from 2-11 years (N= 559)</td>
<td>116</td>
</tr>
<tr>
<td>10. Girls’ Internalizing Symptom Lag+1 Correlations from 2-11 years (N= 559)</td>
<td>117</td>
</tr>
<tr>
<td>11. Concurrent Correlations Between Internalizing and Externalizing Symptoms</td>
<td>118</td>
</tr>
<tr>
<td>12. Boys’ Externalizing Univariate Model Fit Statistics</td>
<td>119</td>
</tr>
<tr>
<td>13. Boys’ Internalizing Univariate Model Fit Statistics</td>
<td>120</td>
</tr>
<tr>
<td>14. Girls’ Externalizing Univariate Model Fit Statistics</td>
<td>121</td>
</tr>
<tr>
<td>15. Girls’ Internalizing Univariate Model Fit Statistics</td>
<td>122</td>
</tr>
<tr>
<td>16. Diagnostics of Model Adequacy for Univariate Models</td>
<td>123</td>
</tr>
<tr>
<td>17. Degree of Functions Used to Model Trajectory Groups</td>
<td>124</td>
</tr>
<tr>
<td>18. Symptom Patterns Exhibited by Children in the Four Univariate Models</td>
<td>125</td>
</tr>
<tr>
<td>20. Girls’ Joint Membership versus Hypothetical Independent Model Probabilities</td>
<td>127</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Schematic of Hypothesized Externalizing and Internalizing Trajectories</td>
<td>129</td>
</tr>
<tr>
<td>2. Boys Externalizing Scores versus Age: 7 Group Model</td>
<td>130</td>
</tr>
<tr>
<td>3. Boys’ Internalizing Scores versus Age: 7 Group Model</td>
<td>131</td>
</tr>
<tr>
<td>4. Girls’ Externalizing Scores versus Age: 8 Group Model</td>
<td>132</td>
</tr>
<tr>
<td>5. Girls’ Internalizing Scores versus Age: 7 Group Model</td>
<td>133</td>
</tr>
<tr>
<td>6. Probability of Externalizing Conditional on Internalizing Membership: Boys</td>
<td>134</td>
</tr>
<tr>
<td>7. Probability of Internalizing Conditional on Externalizing Membership: Boys</td>
<td>135</td>
</tr>
<tr>
<td>8. Probability of Internalizing Conditional on Externalizing Membership: Girls</td>
<td>136</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Over a decade ago Kendall and Clarkin (1992) declared the study of comorbidity to be “the premier challenge facing mental health professionals in the 1990’s.” Comorbidity research, theorized Rutter (1997) and Jensen (2003), held promise both for identifying how disorders can serve as vulnerability factors for other disorders, and for illuminating similarities and differences in etiology within and across disorders. With this theoretical backing, comorbidity research was swiftly embraced in the wake of Kendall and Clarkin’s call to arms. Unfortunately, however, the resulting proliferation of studies mainly either documented snapshot rates of comorbidity from adolescent or adulthood cross-sectional data, or rates based on recall of the whole life course (see Angold, Costello, and Erkanli, 1999 for a review). Only a few longitudinal studies reported rates of childhood comorbid diagnoses, as well as dates of first “onset” of these comorbidities (e.g. Costello, Mustillo, & Erkanli et al., 2003; Cohen, Cohen, & Kasen, et al., 1993). The aims of comorbidity research (as stated by Jenson, Rutter, and others) have thus far been disconnected from the methods employed in this research, in two respects. The statistical techniques employed have not fit the goals of the research, and the late infancy and early childhood period has been neglected. The present study will address these two problems.

The first problem, more specifically, is that traditional statistical approaches have represented onset of comorbidity by highlighting when two problems first both met
diagnostic criteria (using survival analyses), and then summarizing the longitudinal stability of their co-occurrence (using a correlation or odds ratio). An age of comorbidity onset, however, could have very different meaning, for example, if the two problems had increased in severity together gradually, or abruptly; if increases in one had lagged after the other; or if the severity of one had peaked and was declining as the severity of the other was increasing and finally reached diagnostic threshold. Moreover, it could be that different subgroups of individuals show differing, yet equifinal, trajectory functions toward developing comorbidity (Cicchetti & Toth, 1998). A documented rate of comorbidity is but one point on the dynamic, unfolding trajectories of the problem behaviors, and the application of conventional variable-oriented statistical methods may obscure the very etiologic complexities the field seeks to uncover. Newly developed person-oriented methods for probabilistically linking the developmental courses of two behaviors using joint trajectories, and for capturing heterogeneity in the expression of each behavior (Nagin & Tremblay, 2001) will be used in the current study.

The second problem lies in the tendency of comorbidity researchers to only investigate covariation in suprathreshold pathology, and the tendency for infancy and early childhood researchers to restrict their investigations to the development and articulation of single problem behaviors (e.g. depression, Luby, 2000, and aggression, Campbell, 1995). This effectual division of labor has limited what etiologic insights could be gleaned from research on covarying conditions. An understanding of heterogeneity among patterns of covariation of even the broadest of symptom classes—internalizing problems (including anxiety, depression, social withdrawal, and somatic complaints) and externalizing problems (including aggressive, destructive, and delinquent behavior; Achenbach & Rescorla, 2000)—
from late infancy to middle childhood would help clarify how later comorbid states evolve. The current study aims to provide such an understanding.

Although heterogeneity among patterns of internalizing/externalizing covariation is largely unexplored, five main theories exist from previous literature that predict alternate patterns of when such covariation will begin and desist, whether all or subgroups of children will show covarying symptoms, and whether increases in one broadband factor will precede increases in the other. Using semiparametric growth mixture modeling (SPGMM; Nagin, 1999; Nagin & Tremblay, 2001) I will model the longitudinal course and patterns of covariation of children’s internalizing and externalizing problems using eight waves of data from ages 2 to 11 years. I will test the five theories to see whether the patterns predicted by any or all of them are reflected in the data. I will compare and contrast the information on covariation patterns obtained from SPGMM methods with that obtained from traditional correlative methods.

The first half of this review is devoted to the content area of this investigation, as follows. I begin with a review of available evidence about the development of externalizing problems, the development of internalizing problems, and the nature and validity of their covariation from 2-11 years. Next, I describe five main, competing theories of how internalizing and externalizing problems may covary in early childhood, and delineate the hypotheses predicted by each. The second half of this review is devoted to the statistical technique used in this investigation. I first give a conceptual description of growth mixture modeling, and then explicate when it is an appropriate method of analysis, what population inferences can be drawn from this technique, and why I chose Nagin’s (1999) version of GMM over an alternate version of GMM (Múthen, 2001) and over a parametric random
coefficient growth model. I end by discussing ways in which Nagin’s analytic method is compatible with my data.

**Development of Externalizing Problems**

The range of aggressive and delinquent behaviors known as externalizing problems (Achenbach & Rescorla, 2000) has been historically underexplored in the late infancy and preschool years because of the sentiment that undercontrolled behaviors at this age are normative, amorphous, and will be outgrown. We now know that, while age 2 represents the peak of aggressive behavior during the life course (Tremblay, 1998), strikingly, children who are not aggressive during infancy are unlikely to manifest clinical levels of aggressive behavior at any point later in the life course (Shaw, Gilliom, & Giovannelli, 2000).

Normative patterns do show gradual decreases in aggressive behavior from ages 2-5—when change is defined by shifts in mean rates of behavior. This normative decrease corresponds with developing language and cognitive abilities that permit the use of emotion regulation strategies other than physical aggression to settle disputes, for example. Nevertheless, stability of aggressive problems remains moderately high, when examined using correlative techniques. A recent review found that half of preschoolers with moderate to severe externalizing problems continue to have some externalizing problems at age 9, with 67% of these problems composed of attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), or conduct disorder (CD). Boys’ externalizing problems at age 9, however, were more severe than girls’ (Campbell, 1991).

This clearly speaks to the importance of taking an individual differences approach in order to differentiate among children whose normative levels of aggression declines with the acquisition of adaptive social skills, and children whose externalizing problems persist. There
has been progress in this area. Studies that have used Nagin’s (1999) semiparametric growth mixture modeling (SPGMM) method to model externalizing problems from ages 2-8 (e.g. Shaw, Gilliom, & Ingoldsby et al., 2003; NICHD ECCRN, 2004) and from ages 5 or 6 to late adolescence (e.g. Cote, Zoccolillo, & Tremblay, et al., 2001; Broidy, Nagin, & Tremblay, et al., 2003) have repeatedly shown a similar pattern of trajectories of externalizing problems (see Table 1), with some variations according to gender, the specific externalizing symptom measured, and age.

In the modal pattern over studies in Table 1, a small portion of children (typically 3-10% of the sample) evidence chronically high rates of the problem behavior throughout the observational period; another one fourth to one half of children show consistently low levels of externalizing problem behavior (either no problems or low and stable rates); finally, one fourth to one half of children show decreasing trajectories (either low rates declining to zero, moderate rates declining to low, or high rates declining to low/moderate). Some studies also find that 10-22% of children follow a moderate, stable trajectory. The meaning of this moderate, stable trajectory, however, is blurred by the fact that across studies investigators have measured externalizing symptoms using different manifest variables with different metrics. Some investigators have designated a stable level of problems in the middle of the scale as “low-stable,” and others as “moderate-stable.” In Table 1, I have sought to apply a more consistent classification. I have only labeled stable problems falling in the lowest fifth of the scale as “low-stable,” and labeled stable trajectories “moderate-stable” if they are above this, but below the chronic, stable trajectories near the scale ceilings. (For consistency, this labeling method will be applied throughout the presentation and discussion of the present study’s findings as well.)
In sum, Table 1 conveys profound similarities in the proportion of children following each externalizing trajectory type across samples collected in different continents, decades apart—even if the number of groups recovered by GMM varies somewhat (not shown). The modal pattern described above represents an integrated portrayal of accumulated findings from the development psychopathology field, including Patterson and colleagues’ (1993) subgroups of “early starter” aggressive children and Loeber and colleagues’ (1989) “desisting” versus “persisting” children. Additionally, Table 2, which details findings from GMM studies focusing on mid- to late-adolescence (e.g. Chung, Hawkins, & Gilchrist, et al., 2002; Lacourse, Cote, & Nagin, et al., 2002), provides some evidence of the “adolescence-limited” aggressive/delinquent subtype identified by Moffitt (1997). However, the existence of rising externalizing trajectories in adolescence is very controversial, as they appear in some studies but not others, and their shape and slope seem to be influenced by the length of the follow-up period.

Age/developmental differences in Table 1 are also notable. Both of the preschool/early school age samples (Shaw et al., 2003 and NICHD ECCRN, 2004) lack no-problem trajectories and have large portions of their samples in desistor trajectories. This fits with prior findings of a) normatively high levels of tantrums and disobedience in toddlers and b) mean developmental decreases in physical aggression by school entry, in favor of negotiation or more subtle, relational aggression in most children (e.g. Dodge, Coie, Pettit, & Price, 1990).

Some gender differences also emerge from Table 1. Lower proportions of girls follow high, chronic trajectories, and mean levels of physical aggression within the boys’ high chronic groups exceed levels in the girls.’ Also, girls exhibit fewer trajectories of
externalizing symptoms and a smaller range of best-fitting numbers of groups (as the number of trajectory groups best fitting the data was between 2 and 4 for girls, but between 3 and 7 for boys). (Fewer groups for girls could be due either to less heterogeneity in developmental patterns or weaker correlations among repeated measures over time.) While fewer girls may exhibit elevated and chronic aggression, when they do so, it predicts negative behavioral outcomes analogous to boys’ outcomes. Girls on moderate-high and high-chronic disruptive behavior trajectories from elementary school have significantly higher rates of conduct disorder diagnoses in adolescence than girls on either of the low trajectories (Cote et al., 2001).

There is evidence, moreover, that in the preschool years, girls with early-onset conduct problems have equivalent levels of total externalizing behaviors, noncompliance, verbal defiance, and instrumental aggression as do boys, but potentially still less physical aggression (Hay, Castle, & Davies, 2000; Broidy et al., 2003). The disparities between the behavior of preschool girls with early-onset conduct problems and their normative same-sex age-mates, however, is more extreme than for boys with early-onset conduct problems (Webster-Stratton, 1996). These data suggest a preliminary answer to the longstanding question of when gender differences in externalizing behavior first emerge. It may be that in toddlerhood and preschool, girls and boys exhibit similar levels of externalizing behaviors, but that fewer girls continue to exhibit homotypic high, chronic externalizing behavior (especially of the physically aggressive variety) into the elementary and high school years.

The initial overview of SPGMM studies of externalizing problems presented in Table 1 also exposes five gaping holes in this literature. First, there are very few studies that include toddlers and preschool children. Since those that are available evidence unique
externalizing patterns, further replication is needed, as it is therefore insufficient to draw conclusions based on extrapolating backwards from the many datasets starting in kindergarten. Second, most of the studies in Tables 1 and 2 have not included girls, possibly due to a stereotypic view that early externalizing problems do not represent a serious or prevalent enough problem to warrant investigation (Robins, 1986). Findings from the studies that did include girls have shown this not to be the case.

Third, as depicted in the “Sample Designation” column of Table 1, many of these studies have relied on the same samples, and thus replication of findings in different longitudinal samples is warranted. Fourth, as shown in the “Demographics” column, many of these studies have utilized high-risk (inner city, low-socioeconomic status) samples, so the external validity of the SPGMM externalizing trajectory patterns for unselected, community samples is still questionable, especially before school-age.

Fifth, since studies in Table 1 have each modeled the longitudinal course of different individual externalizing problems (e.g. physical aggression, hyperactivity, or opposition), they have consequently evidenced somewhat discrepant trajectory patterns (e.g. a higher proportion of boys follow chronic hyperactive trajectories than chronic aggressive or oppositional trajectories). At least one attempt has been made at integrating and synthesizing the results from these studies (see Broidy et al., 2003). While Broidy’s synthesis seems plausible, there are advantages to adopting the opposite strategy of first charting heterogeneity in the longitudinal course of the broad-band externalizing construct, and next modeling each component behavior (i.e. aggression, hyperactivity) individually. For if we are merely fitting single jigsaw pieces together without an image of the final puzzle to follow, we will not know if we have left any pieces out. No prior study has provided an
overarching understanding of how the entire externalizing construct develops over time. This could be helpful, especially in preschool samples, as a starting point from which to integrate findings of heterotypic and homotypic continuity among specific externalizing disorders over time.

Conversely, since the current study is measuring trajectories of the broad-band externalizing factor over time, we can assume that although some types of the component narrow-band behaviors will manifest in different phenotypic forms at different ages, we will still capture them, as they will still fall under the general externalizing/undercontrol construct. A broad-band factor approach is also effective in capturing *concurrent* phenotypic variation that may be particularly profound in the expression of early behavioral and emotion regulation difficulties (often appearing as concomitant impulsivity, physical aggression, and hyperactivity, Stormshak, Bierman, & Coie et al., 1998).

I aimed to fill the five aforementioned gaps by extending models of heterogeneity in externalizing problems to include toddlers/preschoolers, both genders, and all externalizing symptoms, in a community sample. I then link these findings to co-occurring patterns of internalizing problems.

*Development of Internalizing Problems*

Misconceptions that historically hampered research on early externalizing problems (e.g. the labeling of toddler/preschooler symptoms as temporary and inconsequential) parallel similar resistance within research on early internalizing problems. Some early notions—for example, that prepubertal children do not have a sufficiently well-developed sense of self or expectations of self to permit internalizing problems—have been long dispelled (Luby, 2000). However, there continues to be a bias toward attending to early internalizing
problems—consisting of anxiety, depression, somatic complaints, and social withdrawal (Achenbach & Rescorla, 2000)—only insofar as they are risk factors for later competencies, not as presently impairing conditions in their own right. This bias seems unwarranted as, for example, depression identified after school entry already can have a chronic and relapsing course (Luby, 2000).

Another fading misconception about early internalizing difficulties is that their symptom constellation shifts over time and is markedly different in young children than in adolescents and adults. Preschoolers and early school-aged children were formerly thought to exhibit “masked depression” in which internalizing difficulties were manifested indirectly as somatic complaints and non-affective symptoms rather than the typical phenotype of anxiety or anhedonia and sadness. In contrast, recent research has shown that depression in preschool-aged children is characterized predominately by the core, typical DSM-IV symptoms of sadness/irritability and neurovegetative signs (Luby, Heffelfinger, & Mrakotsky et al., 2003). These symptoms were more robust and specific markers of depression than was somatization, which was not more frequent among depressed preschoolers compared to those with other psychiatric disorders. In fact, the severity, rates of recovery, and recurrence of depression are similar in children and adolescents with depression (Birmaher, Williamson, & Dahl et al., 2004), and the only criteria that may not be met by preschoolers is the 2-week duration criteria (Luby et al., 2003).

Similar to depression, the phenotype of fears and anxiety from preschool to preadolescence shows consistency; although the frequency and severity of fears change with age, the highest-rated fears and worries remain relatively stable (Muris, Merckelbach, Gadet, & Moulaert, 2000). While there is evidence that school-aged children with anxiety disorders
report more somatic complaints than their non-disordered age-mates, this varies according to gender and complaint type (Egger, Costello, & Erkanli et al., 1999). Together, this evidence suggests that depression and anxiety problems are not so confounded with somatic or externalizing symptoms in early childhood as to be unrecognizable in their typically-manifesting forms.

To address a final misconception, internalizing disorders may be measured with at least moderate reliability and validity using parent-reported checklists as early as toddlerhood and preschool. The Child Behavior Checklist (CBCL) Version 1.5-5 internalizing disorder factor has a high 8-day test-retest reliability (r=.90) and substantial stability over a 12-month interval (r=.76) and from ages 2 through 9 (r=.46) (Achenbach & Rescorla, 2000). There is a significant relationship between preschoolers’ DSM-IIIR and CBCL clinical or borderline classifications of internalizing disorders, yet the overlap is only one third (Keenan, Shaw, & Walsh et al., 1997). The CBCL correctly classifies 63% of preschoolers’ DSM-III-R internalizing cases as true positives or true negatives. Although this classification accuracy is lower than that for preschoolers’ DSM III-R externalizing disorders, it is comparable to the CBCL classification accuracy for school-aged children’s DSM diagnoses based on either child or parent report (Jensen & Watanabe, 1999). Further, for school-aged children, CBCL clinical cut-off classifications are statistically equivalent to DISC classifications, compared against the discriminant function derived from a set of 22 external validator variables. These external validators include psychosocial, demographic, and developmental factors that have related to psychopathology in past studies (Jensen, Watanabe, & Richters et al., 1996).

There have been only a few explorations of the epidemiology and stability (using summary score estimates) of internalizing problems. Angold, Egger, Erkanli, and Keeler
(2004) reported a 3.5% incidence of internalizing disorders in a community sample of 2-5 year olds, (4.2% among girls; 2.7% among boys) using the first diagnostic assessment of preschool psychopathology (the Preschool Age Psychiatric Assessment; Egger & Angold, 2004). These rates were similar to those for externalizing disorders in the same age group (3.7%), indicating that internalizing problems in early childhood are already well developed, and merit research attention. One study found lower continuity of emotional disorder diagnoses (29%) from preschool to middle childhood, as compared to 40% and 48% of the conduct and attention deficit disorder groups, respectively (Beitchman, Wekerle, & Hood, 1987).

While there is considerable evidence of a gender difference in depressive symptoms in early to middle adolescence (Cicchetti & Toth, 1998; Costello, Mustillo, & Erkanli et al., 2003), and some evidence for slightly higher rates of anxiety problems in girls throughout childhood (Muris et al., 2000; Costello et al., 2003), the typology of gender differences in internalizing problems (in terms of variability and shape of symptom trajectories and proportions of boys and girls on chronic versus desisting versus low, stable trajectories) from toddlerhood through elementary school remains uncharted. Additionally, in contrast to the externalizing disorder field, there have been no descriptive studies of homotypic or heterotypic outcomes of symptom trajectories of specific internalizing problems (Costello, Pine, Hammen, et al., 2002). The present study addresses this need, and since I measured trajectories of the broad-band internalizing factor over time, even if some types of component narrow band symptoms phenotypically manifest differently at different ages or concurrently, I still captured them within the general internalizing construct.
I have shown that there is a need for characterizing patterns of internalizing symptoms for both genders over the toddler to school-age period. I chart these patterns and then link the findings to co-occurring patterns of externalizing problems, and the more-developed SPGMM literature base in that area.

*Externalizing and Internalizing Problem Covariation*

Early covariation of internalizing and externalizing problems (defined as the tendency of both symptoms to co-occur more often than expected by chance; Lilienfeld Waldman, & Israel, 1994) has been well documented in the literature (e.g. Garnefski & Diekstra, 1997; Achenbach & Rescorla, 2000). Summary estimates of overlap have evidenced moderate correlations, (ranging from .67, Gould, Bird, & Jaramillo, 1993, to .46, Weiss & Catron, 1994) indicating that these domains are related but distinct.

There are several reasons why the observed covariation between internalizing and externalizing problems appears to be an authentic psychological phenomenon rather than a methodological artifact. First, this covariation has been observed in community as well as clinical samples, and thus is not simply a result of Berkson’s or clinical referral biases (Weiss & Catron, 1994; Lilienfeld, 2003.) Second, although there are documented informant-specific individual-view components to parental ratings of CBCL internalizing and externalizing symptoms that could seemingly result in the appearance of internalizing/externalizing covariation (such as differing thresholds for endorsing behaviors; Rowe & Kandel, 1997), because such covariation has been observed across child-report, parent-report, respondent-based interviews, and interviewer-based interviews, such biases seem not to be driving the evidenced covariation (Angold, Costello, & Erkanli, 1999). Third, since the present study is concerned with what Angold terms *heterotypic covariation* (covariation between dimensions
of broad-band factor groupings rather than within narrow band factors of the broad-band groupings), any observed covariation is less likely to be attributable to single behaviors getting coded as multiple symptoms on both factors. Artifactual covariation is more likely when looking at homotypic covariation between, for example, ADHD and ODD externalizing problems where a single behavior could be doubly coded as “often leaving seat” for the ADHD criteria and “often actively defies rules” for the ODD criteria (Angold, Costello, & Erkanli, 1999).

Five Theoretical Predictions of Internalizing-Externalizing Covariation Patterns

The above review indicates that there is very likely non-artfactual covariation between internalizing and externalizing symptoms in early childhood. Because the previous literature has mainly summarized covariation using correlations, however, we have little knowledge of individual differences in covariation patterns and precisely how covariation patterns change over time. We do not know, for example, whether some children have pure (non-co-occurring) and stable symptoms across early childhood; whether others have co-occurring, stable symptoms; and whether still others have patterns of alternating co-occurring and then pure symptoms over time. While heterogeneity of internalizing and externalizing symptom covariation has not been specifically researched in early childhood, five theories—each backed by some empirical support—offer differing hypotheses of what patterns may exist. (The textual descriptions of each hypothesis are supplemented with hypothetical/heuristic graphical and tabular depictions for improved clarity—see Figure 1 and Table 3).

Hypothesis #1: Differentiation with Development. A number of theorists have proposed that broadband symptom covariation in early childhood may be a function of the
orthogenetic principle. For example, Angold and Costello (1992) suggest that early childhood is characterized by undifferentiated responding to stress (manifested, for example, as initially high internalizing and externalizing covariation) that over time coalesces into a specific disorder. Children with covarying internalizing and externalizing problems may be at a stage when the different cognitive and emotional developmental processes underlying these syndromes have not fully diversified into distinct trajectories (Lilienfeld, Waldman, & Israel, 1994; Nottleman & Jensen, 1995). For example, Gjone and Stevenson (1997) found that covariation between CBCL internalizing and externalizing behavior problems was most pronounced in the lowest age group studied (5-9 year-olds), as compared to 12-15 year-olds. Conceptualizing early internalizing and externalizing problems as non-specific symptomatic expressions of an underlying dysregulation would be strengthened if I observed a simultaneous onset of these symptoms (Biederman, Faraone, Mick, & Lelon, 1995).

This theory would lead us to hypothesize that some children with co-occurring internalizing and externalizing elevations in toddlerhood would tend to exhibit similar levels of both behaviors at early time points, but be more likely to be elevated on only one factor—as per the differentiation of pathology—by the middle of elementary school. This hypothesis will heretofore be referred to as the “Differentiation Hypothesis.” The differentiation hypothesis would correspond with a proportion of children with a high joint probability of being on a stably-elevated trajectory for one dimension (e.g. externalizing trajectory groups 1 or 3 in the heuristic Figure 1) and on an elevated-to-low desisting trajectory for the other dimension (e.g. internalizing trajectory groups 7 or 9, see Figure 1 and Table 3.)

Hypothesis #2: Internalizing Subtypes. Some theorists (e.g. Hammen and Ruldolph, 2003; Beauchaine, 2003) suggest that multiple subtypes of internalizing problem trajectories
may exist. Specifically, a co-occurring, stable internalizing and externalizing elevation may be a marker or prodrome for a different subtype of disorder than an early pure, stable internalizing elevation. The idea of a co-occurring internalizing/externalizing subtype was taken up by ICD-10 in their depressive conduct disorder diagnosis, and was supported by arguments that the psychosocial correlates and outcomes of comorbid depression and conduct disorder more resemble those of pure conduct disorder than pure depression (World Health Organization, 1993). Further, the idea of a pure, stable internalizing subtype lies at the intersection of several lines of research, discussed below. Both biological and behavioral evidence shows a stronger genetic basis, different long-term outcomes, and a possible less pernicious course for early, pure internalizing problems, as compared to early, co-occurring problems. This biological and behavioral level convergence follows the accepted practice of demonstrating differences between potential disorder subtypes at multiple levels before considering these subtypes to be valid (Angold, Costello, & Erkanli, 1999).

Beginning with the biological evidence, within an epidemiological sample of 526 identical and 389 fraternal 5-15 year-old twins, Gjone and Stevenson (1997) found that pure internalizing problems were discriminated by stronger genetic influence, and co-occurring internalizing and externalizing problems were discriminated by a stronger influence of shared environmental factors. Similarly, while children with early-onset MDD are at increased risk for developing later conduct and substance abuse disorders compared to controls, those early-onset children with major depression in first degree relatives are most likely to exhibit homotypically-continuous recurrent major depressive episodes (Weissman, Wolk, & Wickramaratne et al., 1999). Puig-Antich and colleagues (1989) also found less familial loading for comorbid depressive and conduct disorders than for pure depressive disorders.
Behavioral evidence of a possibly less pernicious internalizing-only trajectory comes from several sources. Verhulst and colleagues (1993) found that children who had non-comorbid early internalizing disorders were more likely to have decreasing problems from early to mid childhood than children with early externalizing disorders. Harrington and colleagues (1991) found that comorbid conduct disorders and depression in children predict significant adult pathology, but show less continuity with adult depression diagnoses than do early pure depressive problems. Zoccolillo (1992) argues that this may be because the more severe the externalizing problem (such as conduct disorder), the higher the likelihood of co-occurrence with emotional disorders, which can have an onset at the same time as the conduct problem. Early childhood depressive problems only predict adulthood depressive diagnoses in the context of low comorbidity, family history of depression, and stability of depressive problems (Harrington et al., 1991).

From this evidence, we can conclude that if children with early and stable co-occurring internalizing and externalizing problems develop later problems, they tend to be in the externalizing spectrum, while children with solely early internalizing problems tend to develop only later internalizing problems, or no later problems. I respond to Hammen and Rudolph’s (2003) plea for researchers “to consider subtypes of depressed children with and without co-occurring symptoms when researchers are developing and testing models of depression” (p. 262).

As implicated by the above research, I hypothesize that I will find evidence consistent with a mixture of distinct subgroups of internalizing trajectories: one subgroup will have early, stable, and co-occurring elevations in internalizing and externalizing symptoms, and another subgroup will have only stable, moderately-elevated internalizing symptoms. Finding
separate trajectories of internalizing problems with and without comorbidity is an initial step towards identifying subgroups of internalizing problems that future research may show to have different risk factors, etiologies, and treatments. For convenience, this hypothesis will summarily be referred to as the “Internalizing Subtypes” hypothesis. The first internalizing subtype (stable, co-occurring elevations) would correspond with a proportion of children having a high joint probability of being on stably-elevated internalizing and externalizing trajectories (e.g. externalizing groups 1 or 3 and internalizing groups 6 or 8 in Figure 1). 2) The second internalizing subtype (pure, stable internalizing elevation) would correspond with a proportion of children with a high joint probability of being in a moderately elevated internalizing trajectory (internalizing group 8 in Figure 1) and a low, stable externalizing trajectory (externalizing group 5 in Figure 1; also see Table 3.)

Hypothesis #3: Rising Internalizing. A third hypothesis is based on the argument that toddler and preschool externalizing problems are stage-specific manifestations of a broad variety of complex later disorders that include both externalizing and internalizing symptoms (Bird, Gould, & Staghezza, 1993; Lavigne, Arend, & Rosenbaum et al., 1998b). These theorists postulate that among high-chronic externalizing children there is increasing co-occurrence of internalizing and externalizing problems over time. This idea was originally inspired by Cicchetti and Schneider-Rosen’s (1984) theory that externalizing and internalizing problems increasingly co-occur over time as initially disruptive children cognitively develop the capacity to become more articulate about their feelings.

Some evidence has supported an increased risk for later combinations of internalizing and externalizing problems among children displaying early externalizing problems (Campbell, 1995). For example, Lavigne and colleagues (1998a) found that the vast majority
(80%) of 2-5 year-olds with disruptive disorders did not show comorbidity initially, but by the early school years, 70% of those with continuing disruptive disorders had comorbid diagnoses. Specifically, only 10% of children with initial externalizing disorders had comorbid internalizing (emotional) disorders, but by the early/middle school years, this percentage had risen to 30-50%. In contrast to Hammen and Rudolph’s (2003) idea of a subgroup with early, pure internalizing problems, Lavigne found that for children with early internalizing (emotional) disorders, comorbidity was common at all time points. At least 60% of children with emotional disorders had comorbid diagnosis at every time point.

Biederman and colleagues (1995) also documented a rise in internalizing disorders following early elevations in externalizing disorders. Externalizing disorders such as ADHD, Conduct Disorder, and ODD preceded the onset of major depression disorder by at least a year (with average onset ages of 3.5 to 6.5, as compared to 7.5 for depression). Diagnoses of comorbid depression and externalizing disorders were retained even after removing symptoms shared by both conditions. This suggests that the later-developing internalizing/externalizing disorder comorbidity was not an artifact of overlap of diagnostic criteria.

From the available data, however, it is unclear when internalizing problems begin to increase among children with early, stable externalizing problems. Biederman and colleagues (1995) only discuss when these internalizing problems reach diagnostic thresholds. Given this uncertainty, this line of research and theoretical reasoning leads us, in general, to hypothesize an increase in internalizing symptoms in early/middle childhood, especially among children who manifested externalizing problems in late infancy and toddlerhood. If found, this phenomena could represent either heterotypic continuity of an underlying disorder
that manifests with different symptoms at different developmental stages due to brain
development and hormonal changes, or an addition of symptoms, possibly due to early
externalizing problems serving as risk factors for later internalizing problems (Biederman et
al., 1995). This hypothesis will summarily be referred to as the “Rising Internalizing
Hypothesis.” The rising internalizing hypothesis would correspond with a proportion of
children with a high joint probability of being on a stable, elevated externalizing trajectory
(externalizing groups 1 or 3 in Figure 1) and an increasing internalizing trajectory
(internalizing group 11; see Figure 1 and Table 3.)

_Hypothesis #4: Covariation Stability._ The fourth theory draws on literature reviewed
earlier showing moderately stable covariation between internalizing and externalizing
problems at multiple time points in early and middle childhood. For example, Lavigne and
colleagues (1998a) demonstrated diagnostic stability of .54 for disruptive disorders comorbid
with another disorder among preschoolers followed for 2-5 years. _From these results, I
hypothesize that there will be a proportion of children who will exhibit similar patterns of
behavior on both symptom dimensions—this could be evidenced either as exhibiting both
desisting internalizing and externalizing, both stably-elevated internalizing and
externalizing, or low-stable internalizing and externalizing problems._ This hypothesis will
heretofore be referred to as the “Covariation Stability Hypothesis.” The covariation stability
hypothesis would correspond with a proportion of children with a high joint probability of
being on the same type of trajectory on both symptom dimensions (e.g. trajectory groups 1
and 6; or groups 3 and 8; or, alternately, groups 5 and 10. See Figure 1 and Table 3.)

_Hypothesis #5: Gender Differences._ In older children, some gender differences in
patterns of internalizing/externalizing problem comorbidity, albeit inconsistent, have
surfaced. Some studies have shown more comorbidity among depression and anxiety disorders for adolescent girls and more comorbidity among depression and conduct disorders for adolescent boys (Kessler, Avenevoli & Merikangas, 2001). Other studies have shown equal rates of depression and anxiety comorbidity for both genders across adolescence, but more comorbidity among depression and externalizing disorders for adolescent boys and more comorbidity among depression and eating disorders for adolescent girls (Kovacs, Obrosky, & Sherrill, 2003). Yet, several investigators have found no major differences in the maternal-reported internalizing/externalizing comorbidity patterns between preschool or early school age boys and girls (Angold et al., 2004; Keiley, Lofthouse, & Bates et al., 2003). Since these studies found no gender differences in preschool and early school age internalizing/externalizing comorbidity patterns, I hypothesize that the covariation patterns explicated in hypotheses 1-4 will be similar across gender, when separate joint trajectory analyses are conducted for boys and girls. Nonetheless, I hypothesize that there will be gender differences in the distribution and number of trajectory groups within the externalizing dimension. Following Broidy and colleagues (2003; see Table 1) I hypothesize that there will be fewer girls than boys in the respective high, chronic externalizing groups and fewer number of trajectories for girls’ externalizing behaviors. These hypotheses will be summarily referred to as the “Gender Differences Hypotheses.”

The primary aim of this study is to use growth mixture modeling (GMM) to identify whether my available data is consistent with one of these five hypothesized patterns of internalizing/externalizing covariation for most or all children, or whether, in fact, my data are consistent with groups of children evidencing heterogeneous, distinct patterns of internalizing/externalizing covariation, such that multiple competing hypotheses are
simultaneously supported. In addition, adopting a multi-pronged approach, I will compare these GMM trajectory patterns (a person-oriented analysis) to traditional correlative measures of covariation (a variable-oriented analysis) to see if GMM affords us unique or redundant information with traditional statistical approaches.

**Growth Mixture Modeling: A Conceptual Description**

Since growth mixture modeling has only recently been adopted as an analytic tool in developmental psychopathology, I find it important to provide the following for this proposal: a) what GMM does conceptually; b) what can and cannot be inferred from the GMM procedure c) how the specific tradition of GMM used in this paper differs from other GMM approaches; d) how the tradition of GMM used here differs from conventional growth modeling; and e) why Nagin’s (1999) GMM technique fits with my questions and data.

Growth mixture modeling is an improvement upon methods which group individuals ex ante based on subjective classification rules (Nagin, 2004). For my specific research questions, GMM is also an improvement on multiple-group growth modeling, in which the grouping variable must be observed in order to act as an exogenous predictor of trajectory parameters (Bauer & Curran, 2003a).

The semiparametric growth mixture modeling (SPGMM) method used in this study estimates unobserved trajectory classes within a population by modeling an outcome variable as a polynomial function of age or time. The form of the outcome variable determines the specifics of the statistical model used to identify and estimate trajectory groups. My outcome variable is CBCL psychometric scale data, and such data typically show sizable clustering at the scale minimum and smaller clustering at the maximum. I assume that the data are
generated by an underlying censored normal process, as follows (Jones, Nagin, and Roeder, 2001).

Specifically, I assume that data for each group $j$ are generated by an underlying latent variable, $y_{it}^{*j}$. The latent variable is linked to its observed, censored counterpart such that if the latent variable is less than the minimum possible score on the measurement scale, the observed variable will be set equal to the minimum possible score. If the latent variable is between the minimum and maximum possible scores on the measurement scale, it will be exactly equal to the observed variable. Finally, if the latent variable is greater than the maximum possible score on the measurement scale, the observed variable will be set as equal to the scale maximum. The latent variable $y_{it}^{*j}$ is interpreted as the potential for exhibiting internalizing (or externalizing behavior) for an individual $i$'s age at time $t$ given membership in group $j$ (Nagin, 1999). Thus, the relationship assumed between $y_{it}^{*j}$ and age is shown in equation (1) below.

$$y_{it}^{*j} = \beta_{0j} + \beta_{1j} \text{Age}_{it} + \beta_{2j} \text{Age}_{it}^2 + \beta_{3j} \text{Age}_{it}^3 + \epsilon_{ir}$$

(1)

$\text{Age}_{it}$, $\text{Age}_{it}^2$, and $\text{Age}_{it}^3$ are the subject $i$'s age at time $t$, the square of subject $i$'s age at time $t^2$, and the cubic of subject $i$'s age at time $t^3$, respectively. The $\epsilon_{ir}$ error term is assumed to be normally distributed with a homoskedastic variance of $\sigma^2$ across groups and a mean of 0 (Nagin, 1999). The shapes of the trajectories for each of the $j$ groups are structured by the coefficients $\beta_{0j}, \beta_{1j}, \beta_{2j}$ and $\beta_{3j}$. The parameters determining the level and shape of the trajectory vary by group (denoted by the superscript $j$), but every member of the same group has the same slope and intercept. The fact that parameters defining the shape of trajectories vary across groups permits identification of population heterogeneity over time. If, for example, $\beta_{1j}, \beta_{2j}$ and $\beta_{3j}$ were all $= 0$, this implies stable behavior over age; if $\beta_{1j}, \beta_{2j}$ and $\beta_{3j}$
were all > 0, this would imply steadily increasing behavior over age; and if \( \beta_{j1} > 0, \beta_{j2} < 0, \) and \( \beta_{j3} < 0, \) this would imply a single peak in behavior that declines thereafter.

As group membership is unobserved, the Bayesian Information Criteria (BIC) guides the selection of a model with the optimum number of groups, as shown in equation (2).

\[
\text{BIC} = -2\log(L) + \log(n) * k
\]  \tag{2}

In equation (2), \( L \) represents the log of the model’s maximum likelihood, \( n \) represents the sample size, and \( k \) represents the number of parameters in the model. The BIC rewards models that best reproduce the observed data, while being parsimonious—as the improvement in the log likelihood for each additional trajectory group must exceed the penalty for more parameters (similar to the adjusted R-square in regression). (This process is detailed further in the methods section below, as well as in (D’Unger, Land, McCall, & Nagin, 1998; Nagin, 1999).

Because the BIC is in the metric of the log likelihood, I use a standard criteria to provide the probability that the BIC-based “\( j \)” group model is, in fact, the correct model, as compared to models with different numbers of groups. This standard criteria is an approximation of the Bayes factor (Wasserman, 1997). It is the natural log of the BIC of group “\( i \)” minus the BIC of group “\( j \)” and it measures the odds of two possible models being the correct model. Strong evidence for model “\( j \)” corresponds with a Bayes factor approximation of < 1/10, and strong evidence for model “\( i \)” corresponds with a Bayes factor of >10 (see Table 4.)

The parameters for the fixed number of groups are estimated using maximum likelihood. Specifically, the log of the unconditional probability of observing individual \( i \)’s longitudinal sequence of behavioral measurements is maximized for the entire sample. This
serves to estimate the coefficients of the parameters and the probability of membership in
each group. Because individuals’ membership in trajectory groups is not observed,
individuals cannot be sorted ex ante into these groups. Instead, groups are formed through
assigning each individual to the group for which they have the highest posterior probability
of group membership. The posterior probability of group membership is simply the
conditional probability of group membership given the data and estimated model parameters.
This is done by multiplying the probability of observing the behavior--conditional on
membership in group \( j \) for person \( i \)--times the estimated proportion of the population in group
\( j \), and dividing this numerator by the sum of this function for each of the \( j \) groups (Nagin &
Tremblay, 2001).

The semiparametric mixed censored normal method outputs, for each of the two
behaviors (internalizing and externalizing), a number of groups of individuals who exhibit
similar patterns of behavior over time. The two univariate models are linked to form a joint
model by specifying the likelihood function to also output the probability of internalizing
behavior given externalizing behavior, the probability of externalizing behavior given
internalizing behavior, and the joint probability of belonging to each pair of internalizing-
externalizing trajectory groups (Nagin & Tremblay, 2001). For the specific patterns of
trajectories and conditional probabilities expected for each of the current study’s five
hypotheses, refer to Table 3 and the earlier section delineating these hypotheses.

*When is GMM an Appropriate Method of Analysis?*

GMM is an appropriate technique when we can make the assumption, from
theoretical and empirical work that *either* 1) a population of interest is composed of
subgroups with qualitatively different developmental pathways *or* 2) that a population has a
complex, intractable, continuous distribution that is difficult to model without breaking it into a number of simpler component distributions (Bauer & Curran, 2003a). Nagin and Tremblay (2001) explain that in the latter case, when a continuous population distribution is of unknown form, this distribution can be usefully approximated using a discrete distribution of a number of “points of support” (like a histogram overlaid on a continuous curve). An example of a research area deemed appropriate for GMM is outlined below:

“Consider a study of changes in depression. It makes no sense to assume that everyone is increasing (or decreasing) with respect to depression. In a normative sample, many persons will never be high in depression, whereas others will always be high; some persons will be recovering from serious depression, but others will become increasingly depressed, and perhaps another group will oscillate between high and low levels of depression. Such “depression curves” can certainly be represented by a polynomial of sufficiently high degree …however …may not capture the qualitatively different types of trajectories found in the population” (Raudenbusch, 2001; p. 512).

If, on the other hand, we are modeling a variable that is thought to vary according to a common function (such as vocabulary growth), with the growth parameters only differing in magnitude, a conventional growth curve would be a preferable technique (Raudenbush, 2001). Conventional growth curve modeling would, at the first level, provide us with a mean trajectory parameter for the whole population, and, at the second level, allow us to estimate quantitative variations around this parameter (Bauer & Curran 2003b).

Fundamentally, these two techniques give us different ways of viewing the same reality. As researchers, deciding which to use may be guided by a) consistency with past empirical evidence and theory—whether there is evidence that the behavior under study follows a tractable/continuous distribution (like the multivariate normal distribution), manifests in different intensities in clusters of individuals, or follows an unknown/complex/intractable distribution and b) scientific goals—which angle of the reality we want to capture (Bauer & Curran, 2003b; Múthen, 2003).
In choosing to use GMM for this study, I was motivated both by past theory and evidence, and by my scientific goals. As discussed in earlier sections, theory and evidence on internalizing and externalizing covariation imply a mixture of trajectories growing according to different functions, and GMM is a methodological complement that allows me to sort out competing hypotheses about patterns of strength and form of this covariation. Past research that measured covariation by correlating symptoms at each time point and then looking at the continuous (parametric) function of the correlations across time points may have obscured potential differences in covariation patterns. My scientific goals included exploring insights that could be gained by not making assumptions of continuous, unitary internalizing and externalizing trajectories, in hopes that an alternate depiction of the data that could spawn new hypotheses.

What Population Inferences Can and Cannot be Drawn from GMM?

A common misconception of growth mixture modeling is that the existence of groups can be inferred because they can be estimated. I should not base my choice of analytic method on the desire to prove the existence of groups. As pointed out by multiple theorists, it is tautological to say that a model can discover what it assumes (Rindskopf, 2003; Múthen, 2003). Put another way, it is a reversal of the hypothetical-deductive process of science to claim that finding latent classes affirms the existence of true subgroups because these subgroups were hypothesized from theory (Bauer & Curran, 2003a, p. 358). Moreover, conventional growth modeling cannot be used to disprove a mixture hypothesis, and growth mixture modeling cannot be used to disprove a hypothesized single, continuous distribution. Model misfit is not reliably detected when fitting a single, conventional growth curve model to data from a mixed group population (Bauer & Curran, 2001), nor when fitting a growth
mixture model to data from a nonnormal, single continuous distribution (Bauer & Curran, 2003a). In fact, common data checking procedures (such as plotting) do not even reveal mixtures of two normal distributions separated by either one or two standard deviations. Only when the means of the two distributions are around 5 standard deviations apart (extremely unlikely in social science studies) and the variance within classes is low can a mixture be visually detected (Rindskopf, 2003).

Proving whether latent classes represent true subgroups or a single, complex continuous distribution, assert Bauer and Curran (2003a), “ultimately requires a programmatic series of studies aimed at testing the validity of the assumption of heterogeneity” (p. 359). In this study I make several small steps towards this goal. I test for consistency of trajectory covariation with specific, theoretical expectations, and inconsistency with patterns that would be expected if the population was homogenous. For example, finding all parallel trajectories that map evenly onto a normal distribution would be consistent with an expectation of homogeneity, but finding trajectories with diverse functional forms would be consistent with an expectation of heterogeneity (Raudenbush, 2001). I will revisit the issue of which is the more reasonable inference about the structure of the population, given my data, in the discussion section.

While I have heretofore discussed what GMM is conceptually, how it compares with conventional growth curve modeling, and what cannot be inferred from GMM results, I now give my rationale for selecting Nagin’s SPGMM over a common alternative, Muthén’s GMM, and end by explaining why SPGMM is an appropriate analytic method given my data.
Muthén (2001) and Nagin (1999) have both developed growth mixture modeling programs that have many similar advantages: both can accommodate missing data, time-imbalanced repeated measures, and non-normally distributed data. These features are relevant to the current project, as I have both missing data for some children and a psychometrically-scaled outcome variable (CBCL scores) that is likely to be censored at both tails rather than strictly normally distributed. The primary difference between Muthén’s Mplus and Nagin’s SAS PROC TRAJ versions of GMM is not at the software level. While both Muthén’s and Nagin’s software allows for random effects, the primary difference between these two approaches is the importance they place on including random effects in their models and their interpretation of trajectory groups. As discussed below, Nagin’s model conceptualization is more closely allied with the theory and research that underlies the present study, and thus I use his PROC TRAJ program in my analyses.

Nagin and Muthén differ in their conceptualization of groups in GMM. Nagin conceptualizes trajectory groups as a statistical device, or heuristic, for approximating an unknown or highly irregular single distribution and summarizing complexity (Nagin, 1999; 2004). While the latent trajectory classes may correspond with distinct population subgroups, this is not a requirement. According to Nagin, conceptualizing the population as a mixture of groups defined by distinct developmental trajectories has utility insofar as it highlights differences in probable covariation over time, and is not meant to imply that certain children actually reside in actual trajectories, which they could switch into and out of. Since groups are a heuristic device for Nagin, he does not believe that individuals actually “belong” to certain functional groups, following their group trajectory in “lockstep.” He also does not
believe that the number of trajectory groups is “immutable” and should be “reified” (D. Nagin, personal communication, May 26, 2004).

It is therefore not damaging to find that when length of follow-up changes from 7-43 years to 7-63 years in a simulation, there is small shifting of when trajectories peak, and some shifting of membership among groups still exhibiting the behavior of interest (Eggleston, Laub, & Sampson, 2004). This instability of group membership at the individual level, claims Nagin, is merely a byproduct of using homogenous groups as an approximation device (Nagin, 2004). Homogeneity within groups is assumed for simplification of Nagin’s group-heuristic. This means that he assumes that the group mean trajectories (fixed effects) encompass all individual variability in growth and any individual deviations from the group mean trajectories are considered random error (Bauer & Curran, 2003a).

In contrast, Muthén endorses a description of trajectory groups as heterogeneous classes that can be described by the same probability function. According to Muthén’s (2003) conceptualization, groups represent an actual mixture of distributions, and since individuals actually reside within groups, including random effects within trajectory groups is pivotal for accounting for individual variation. While Nagin’s program can accommodate random effects, to include them in a model runs counter to the idea of groups as a heuristic device. Since my goal is to explore how a non-traditional, heterogeneous depiction of covariation patterns can inform theory on the development of internalizing and externalizing problems, rather than to predict the specific characteristics of children who follow certain trajectories, I chose Nagin’s software program and conceptual model. In support of this decision, Carrig and Bauer (2001) found that while in a simulation Muthén’s GMM correctly classified individuals into subgroups at moderate to high rates, random effects models were not reliably
superior to fixed-effects models in recovering trajectory subgroups. I agree with Eggleston, Laub, and Sampson’s (2004) summary that Nagin’s groups are at least a way to explore discontinuous patterns in longitudinal data and at most, if replicated with multiple techniques, can give insights into social ontology.

Compatibility of Data and SPGMM Method of Analysis

There is currently no way to do a formal power analysis for Nagin’s SPGMM method (D. Nagin, personal communication, May 26, 2004). Yet, the data used for this investigation match and meet requirements of the SPGGM technique in four ways: the sample size, number of time points; the developmental period spanned, and the repeated measure of behavior. I briefly discuss each in turn.

Sample size. The number of latent classes that can be identified by SPGMM increases with sample size until a certain point (just as an exploratory factor analysis justifies more factors with a larger sample; Cudeck & Henly, 2003). This was shown in Eggleston, Laub, and Sampson’s (2004) sensitivity analysis of SPGMM. They selected random sub-samples of differing sizes (from 25 to 500) from the same population, and estimated the optimal groups by applying the same objective BIC statistic. The number of groups extracted plateaued at about sample size 200. D’Unger and colleagues (1998) also found that the optimal number of groups was invariant using samples from 500-2000, and Nagin recommends at least 300-400 cases. My sample of 1,222 far exceeds this minimum, and thus I can be confident that I will be able to discern distinct patterns in the data. Even after dividing the 1,222 in half to do separate analyses for each gender, my N is well above the necessary minimum.

Number of Assessment/Assessment Periods. Estimating linear trajectories with SPGM requires at least three assessments points; these do not have to be evenly spaced, but
they should span a developmentally relevant period (Nagin, 1999). Sampson, Laub, &
Eggleston (2004) warn that using short lengths of follow up in childhood and adolescence
and extrapolating prematurely in labeling a high-rate chronic group can make results
vulnerable to misappropriation by “those predisposed to believe in the idea of a …life course
persistor, superpredator group, especially policymakers who seek to intervene” (p. 38-39).
This technique bears a somewhat greater burden than conventional growth curve modeling in
this regard. The current study utilizes more than the minimum number of time points (7 time
points). Further, my time points span an important period of cognitive and emotional growth
over the transition to formal schooling that has been underinvestigated in prior studies (see
Table 1), and I use extra caution in interpreting behavioral trajectories.

Response Variable. My use of dimensional broad-scores from the CBCL rather than
DSM-IV diagnoses is most compatible with a SPGMM analysis given that I want to model
heterogeneity in trajectories over the entire range of the behaviors, not only over the most
severe portion. CBCL broad-band scores capture much wider ranges of severe, moderate, and
low problem behaviors than the categorical DSM approach, and thus retain more information
(Verhulst et al., 1993). When children with subthreshold problems are lumped into an
“absent” category despite having psychosocial impairments, they can be as disturbed as
children meeting criteria but lacking psychosocial impairments (Jensen & Watanabe, 1999;

Because the current study represents an initial foray into modeling co-occurring
internalizing and externalizing symptoms with joint SPGMM, my main focus is on testing
competing theories of covariation, and secondarily, the utility and pros and cons of SPGMM
for this purpose. Given this focus, I decided to use only maternal reports of the CBCL and
leave an investigation of the sensitivity of covariation patterns to conjunctive, disjunctive, and compensatory approaches to combining child behavioral data from multiple informants to a future study. When children reach middle childhood and adolescence, child and maternal reports of internalizing symptoms and co-parental and teacher reports of externalizing symptoms represent the state of the art (Rowe & Kandel, 1997). But in late infancy and early childhood, although child self-report measures of internalizing problems are in development and have shown promise (e.g. Berkeley Puppet Interview and MacArthur Story-Stem Battery), using solely maternal report of symptoms is standard practice (see, for example, Egger & Angold, 2004; Keenan et al., 1997; Lavigne et al, 2001; Jean & Guskin, 2001). Further, maternal report of internalizing and externalizing problems is thought to be more reliable for this age group than for older children (Egger & Angold, 2004). Finally, although the use of maternal report may lead to over-estimates of symptom stability because influences of informant-specific biases, this method did allow us to maximize the amount of useable data across the seven year observational window.
CHAPTER II
METHODS

Participants

Data from children in the NICHD Study of Early Child Care and Youth Development were examined in this study. Recruitment and selection procedures are outlined here, but are described in detail in several publications (NICHD ECCRN, 1997; 1998; 1999) as well as at the study’s website (http://secc.rti.org/). Families were recruited in 1991 through hospital visits shortly after the birth of a child at 10 sites in the U.S. (Little Rock, AR; Irvine, CA; Lawrence, KS; Boston, MA; Philadelphia, PA; Pittsburgh, PA; Charlottesville, VA; Morganton, NC; Seattle, WA; and Madison, WI). All women giving birth during selected 24-hour sampling blocks were approached with a screen for eligibility and asked if they were willing to be contacted again. The eligibility screen included: maternal age of at least 18, maternal ability to speak English, infant not one of a multiple birth and not released for adoption, family residing within one hour of a research site; family not planning to move in the next year; neighborhood not deemed too dangerous for police to visit). Additionally, a list of infant medical exclusions (including chromosomal, metabolic, or congenital abnormalities causing developmental handicaps) and mother medical exclusions (including physical handicaps or intensive care treatment that precluded interviewing) are available on the study’s website (http://secc.rti.org/). Of the 8,986 total mothers who gave birth during the sampling intervals, 5,416 (60%) both met the eligibility criteria and gave permission to be called two weeks later. Of the 5,416, 3,015 (56%) were conditionally randomly sampled for
the phone call; conditioning ensured that there would be adequate representation—at least 10%—of each single mothers, ethnic minority mothers, and mothers without a high school degree.

The two-week post-birth call excluded families whose infant had been hospitalized for more than 7 days, who were planning to move in the next 3 years, or who were not able to be reached with at least three phone contacts, leaving a remaining sample of 1,526 who met all criteria and agreed to an interview. Of these, 1,364 became the study participants by completing a home interview at 1 month post-birth. The resulting sample was diverse; 24% were ethnic minority children; 11% of mothers had not completed high school, and 14% were single mothers. Although the 1,364 participating families were not randomly chosen, they were similar in maternal education, ethnic minority composition, and rate of single parenthood to the eligible hospital sample of 5,146. The final sample may underrepresent some risk categories, such as adolescent motherhood and diagnosed infant disabilities.

Following recommendations of Nagin (1999), only the 1,222 children whose mothers completed Child Behavior Checklists for at least 2 of 8 potential data collection points) were included in analyses. The 10.4% of the sample (N= 142) that was dropped from final analyses was compared to the 89.6% of the sample retained in analyses on a variety of demographic, parenting, child, and familial risk variables to determine whether the analysis sample is biased upward on levels of resources and risk factors.

No differences were found in the proportion of boys versus girls based on inclusion or exclusion in analyses (t= .11, ns). Among the included 1,222 children, 51.6% were boys and 48.4% were girls, and among the excluded 142 children, 52.1% were boys and 47.9% were girls. Included children, however, were less likely to be from a minority group (t= -3.18,
p<.01); 29.6% of excluded families were from a minority racial/ethnic group, but 18.4% of included families were from a minority group.

Mothers from included families were more likely to be living with a husband or partner (86.7% of included and 74.6% of excluded lived with a husband or partner; M=.87 for included and M=.75 for excluded; t=3.89, p<.001.) However, mothers included versus excluded from analyses did not differ on marital conflict at 1 month post birth (M= 3.15 for included and M= 3.10 for excluded; t=.46, ns).

Mothers of participants included in the analyses were older (M= 25.96 years for excluded and M=28.36 years for included, t=4.85, p<.001), and more educated (M= 13.13 years of education for excluded and M=14.36 for included, t=5.55, p<.001). However, mothers included in analyses did not differ from excluded mothers on maternal stimulation of child’s cognitive development at 15 months post birth, (M= 2.61 for included mothers and M=2.43 for excluded mothers, t= .765, ns).

Additionally, at 6 months post-birth, mothers of participants included versus excluded from analyses did not have significantly different levels of maternal sensitivity (M= 9.23 for included and M=9.00 for excluded, t=1.14, ns), or levels of neuroticism (M=29.68 for included mothers, M=31.18 for excluded mothers, t= -1.85, ns) or levels of depressive symptoms (t=-.68, ns; M=8.93 for included mothers and M=9.58 for excluded mothers).

Families of participants included in analyses had lower income to needs ratios (M= 1.88 for excluded, M=2.86 for included, t=4.00, p<.001). Yet, mothers of participants included in analyses did not differ on total life stress at 6 months post-birth (t= -1.90, ns); M=1.01 for included mothers and M=1.24 for excluded mothers. Further, mothers of participants included in analyses did not differ in the number of hours worked at all jobs at
15 months post-birth (t=.40, ns); M=22.05 hours for included mothers and M=20.92 for excluded mothers.

Overview of Data Collection

During the enrollment home visit at 1 month post-birth, and at subsequent visits at 6 months, 15 months, and at approximately 2, 3, 4.5, 6, 7, 9, 10, and 11 years, demographic information as well as multiple measures (questionnaire, testing, trained observers, and interviewers) of child, family, and peer functioning were collected. Data were collected in multiple settings including the home, school, day care, and after-school care contexts. All instruments used are described and listed by informant and by time point at http://secc.rti.org/.

Measures

Demographic Characteristics. At the 1 month post-birth home visit, mothers reported their own age, work status, educational attainment, family income, infant’s ethnicity and gender, and presence of a partner or husband in the home. Income-to-needs ratios were calculated as the ratio of family income to the poverty threshold for the household size and number of children below 18, as per the U.S. Census Bureau tables.

Among the 1222 participants used in the current analyses, 51.6% were boys and 48.4% were girls. Participants’ ethnic backgrounds included 1.6 % Asian or Pacific Islander; 11.8% African American; 81.6% Caucasian, and 5.1% of another racial/ethnic category. The majority of mothers (78.7%) were married at 1 month post-birth. Further, 86.3% of mothers were living with their baby’s father at 1 month post-birth, and 86.7% lived either with a husband or partner. The average age of mothers for the present sample was 28.36. Mothers’ average amount of completed education was 14.36 years, but 9% of mothers did not
complete high school. The average income-to-needs ratio in the analysis sample was 2.86, and 42.5% of families had very low incomes, as indicated by income-to-needs ratios of less than 2.0 at 1 month post-birth. Mothers worked an average of 22.05 hours per week at 1 month post birth.

*Child Behavior Checklist.* The 99-item Child Behavior Checklist/2-3 (Achenbach, 1992) was used to assess problem behaviors for the 2 and 3 year time points, and the 113-item Child Behavior Checklist/4-18 (Achenbach, 1991) was used to assess these behaviors thereafter (at the 4.5, 6, 7, 9, 10, and 11 year time points.) Both versions are empirically-derived and are among the most widely used measures in child clinical psychology. Both show acceptable test-retest reliability; construct, content, and criterion-related validity; and inter-parent agreement (see extensive documentation in the CBCL manuals for details; Achenbach, 1991; 1992).

In prior studies, both versions of the CBCL have been factor analyzed using principal components analyses into narrow-band syndromes. The 6 syndromes for the CBCL/2-3 are: anxious/depressed, withdrawn, sleep problems, somatic problems, aggressive behavior, and destructive behavior. The 8 syndromes for the CBCL/4-18 are anxious/depressed, withdrawn, somatic complaints, aggressive behavior, delinquent behavior, social problems, thought problems, and attention problems. For both versions, withdrawn, somatic problems, and anxious/depressed syndromes were combined to form an internalizing broad-band score, and destructive and aggressive (2-3) or delinquent and aggressive (4-18) were combined to form an externalizing broad-band score. I opted to use broad-band scores rather than narrow-band scores because these broad-band scores have greater reliability than specific narrow-band syndromes in early childhood as a function of their greater number of items.
The CBCL/2-3 contains 59 items in common with the CBCL/4-18, and the remaining 40 items were specifically designed for the younger age group. The CBCL/2-3 has no sex-specific scales, in contrast to the CBCL/4-18. Of the 59 items that appear on both the CBCL 2/3 and CBCL 4-18, only those 28 items that fall into the internalizing or externalizing domains will be used in the present analyses. A list of these included items is provided in Table 5. If one of the 14 internalizing items was missing, the participant received a missing score for that time point. If one of the 14 externalizing items was missing, the participant received a missing score for that time point. Only if more than 6/8 time points had missing scores was the entire case dropped from the analyses altogether (following Nagin, 1999). The form-specific items will not be used because such a change form content could masquerade as developmental change, and serve as a confound. In growth modeling, we must assume that the meaning and metric of the dependent variable are constant across time points, so we can interpret changes in absolute level of scores as real increases or decreases. Inter-correlations for the 14 of these 28 items that load on the internalizing factor, and separate inter-correlations for the other 14 items that load on the externalizing factor are shown in Table 6. These Pearson’s r intercorrelations ranged from moderately high to high, and were slightly lower for internalizing than externalizing symptoms (.78 to .82 for externalizing symptoms and .70 to .77 for internalizing symptoms.) These intercorrelations indicate the reduced-item internalizing and externalizing factors each still represent cohesive constructs.

For a similar reason, I use the CBCL raw scores, rather than the standardized T-scores calculated by the Cross Informant Program software that are based on a nationally-representative sample. T-scores are problematic for a GMM analysis because the precise
developmental change that I want to examine empirically is effectively removed by making the means and variances constant across age. Because the conversion formula from raw scores to T-scores differs across ages, it would be difficult to interpret the meaning of any age-related changes in T-scores. On both versions of the CBCL, mothers rated how well each item described the target child on a 3-point scale from 0-2 (with 0 = not true, as far as you know; 1 = somewhat or sometimes true; 2 = very true or often true).

Analyses

First, the longitudinal pattern of each behavior and of their covariation was examined using correlational statistics. These summary statistic results were compared to results from the SPGM joint trajectory model to illuminate what unique and overlapping information is conveyed the former variable-oriented and the latter person-oriented technique.

As outlined in the introduction, for the SPGMM analyses, I used the SAS-based PROC TRAJ software (detailed in Jones et al., 2001) to fit optimal univariate models for internalizing and externalizing behaviors. The optimal number and shapes of trajectories from both univariate models were used to estimate the joint model. I first estimated a joint model for boys and girls combined and then divided the sample by gender and estimated separate joint models for boys and girls.

I matched the form of the response variable (here psychometric scale data) to its appropriate distribution (here the censored normal) from which the mixture model was built. The censored normal model is also appropriate for continuous data that is not censored, but is approximately normally distributed (Jones et al., 2001).

Two model-fitting criteria were used to guide model selection. First, I used the objective of maximizing the Bayesian Information Criteria (BIC) to guide my search for the
optimal number of groups to include in each univariate model. Because my potential alternative models are not nested, the BIC is recommended over the likelihood ratio test for determination of the optimal model (D’Unger et al., 1998). Second, I used an approximation of the Bayes factor (see Wasserman, 1997) to measure the odds of two possible models being the correct model (as discussed earlier).

Diagnostics of Model Adequacy

The adequacy each univariate model was then assessed. I calculated the posterior probability of each individual’s membership in each group, in order to quantify the precision of group assignments. Following Nagin (1999), mean assignment probabilities for each group exceeding .70 are considered good, and over .90 is high; (1.0 would represent complete certainty of assignment). These posterior probabilities were converted into odds of correct classification statistics for comparison of the adequacy of models with different numbers of groups (formula provided in results section).
CHAPTER III

RESULTS

Correlational Analyses

Correlational analyses evidenced considerable symptom stability from ages 2-11 years, for both internalizing and externalizing symptoms, and for both boys and girls. Symptom stability is indicated by moderately-high lag+1 correlations (i.e. the correlation of a behavior at time X with the same behavior at time X+1). These lag+1 correlations ranged from .66 to .76 for boys’ externalizing symptoms, .56 to .73 for boys’ internalizing symptoms, .66 to .78 for girls’ externalizing symptoms, and .60 to .73 for girls’ internalizing symptoms (see Tables 7, 8, 9, and 10). Unfortunately, correlational analyses cannot tell us whether these moderately-high, stable lag+1 correlation coefficients are driven solely by the large number of children in this community sample who always exhibit low rates of each symptom, or whether symptom stability occurs at high symptom levels as well.

To test whether the first four hypothesized patterns of symptom covariation were upheld in correlational analyses, I examined the longitudinal patterns of concurrent correlations between internalizing and externalizing symptoms. Correlational analyses could only provide information to address parts of the first and the fourth hypotheses. In support of hypothesis #1, differentiation with development, contemporaneous correlations between internalizing and externalizing symptoms attenuated from ages 2 to 9 for girls (see Table 11). Such attenuation was not observed for boys (see Table 11). Before concluding that differentiation with development does not occur for boys, we must bear in mind that would only expect to
see correlation attenuation if most or many of the boys showed differentiation with
development. If only a subset show differentiation, contemporaneous internalizing-
externalizing correlations over the sample would not evidence attenuation over time.
SPGMM analyses will help us sort out whether symptom differentiation with development
does in fact occur for some groups of boys.

Correlational analyses do not permit testing hypotheses # 2 and 3 (whether subtypes
of children show pure-internalizing elevations or co-occurring-stable elevations, or rising
internalizing symptoms), so SPGMM must be employed to detect whether these patterns
occur. (We will see later that the steepest rise in girls’ internalizing symptom trajectories in
SPGMM does correspond with an attenuation in contemporaneous correlations between
girls’ internalizing and externalizing symptoms from 7-9 years. If we were not able to refer
to the SPGMM output, however, we would not be able to infer whether this correlation
attenuation represents a rise or a fall in one of the behaviors, nor which behavior rose/fell,
like we can using SPGMM.)

Correlational analyses did show support for hypothesis #4, covariation stability, as
the contemporaneous correlations between internalizing and externalizing scores were
consistently moderately-strong for boys (ranging from .66 to .72), and for girls (ranging from
.65 to .76), as shown in Table 11.

Finally, with respect to the gender differences hypothesis (hypothesis #5),
correlational analyses do not allow us to compare the proportion of girls versus boys whose
behavior pattern fits with each of the four hypotheses. Correlational analyses only make the
point that both genders exhibit moderately-strong covariation stability, as per hypothesis #1.
Correlations also do not allow us to compare the proportion of girls versus boys who exhibit each level and functional form of internalizing and externalizing behavior.

Thus, correlational analyses were mainly effective in identifying stability of each symptom and stability in their covariation. We now turn to the results of the complementary SPGMM analyses to both gain a more nuanced understanding of these findings (for example, whether there is stability at all levels of the symptom; what changes in what proportion of children underlie periods of correlation attenuation), and to test hypotheses 1, 2, 3, and 5 that correlational analyses were unable to fully address.

**Semi-Parametric Growth Mixture Analyses**

*Model Fitting: Four Univariate Models.* All analyses were conducted separately for each gender, as initial findings showed that the number, order, and composition of trajectory groups in the boys’ best-fitting models differed from the girls’. Joint trajectory analyses with a combined-gender sample obscured these differences.

**Number and Order of Trajectory Groups** Tables 12, 13, 14, and 15 depict model fit criteria, here BIC scores and Bayes factors, for competing univariate models of boys’ and girls’ externalizing and internalizing behaviors with different numbers of groups. As detailed earlier, the best fitting models had the maximum Bayesian Information Criteria (BIC) (i.e. the least negative value). BIC scores can also be converted into a probability that model \( j \) is the best model, as follows in equation (3).

\[
p_j = e^{BIC_j - BIC_{\text{max}}} / \sum_j e^{BIC_j - BIC_{\text{max}}}
\]  

(3)

In equation (3), \( p_j \) = the probability that model \( j \) is the best model, \( BIC_j \) = the BIC of model \( j \), and \( BIC_{\text{max}} \) = the BIC of the model with the maximum BIC score. The best fitting model
should also show stronger odds of being the ideal model than do competing models, as indicated by its Bayes’ factor. Following Kass and Wasserman (1995), I used $e^{\text{BIC}_i - \text{BIC}_j}$ as a means of comparing the odds of each of two competing models being the ideal model. This formula is a good approximation of the Bayes factor. Wasserman (1997) and Nagin (1999) use the Jeffrey’s Scale shown in Table 4 as a rubric for interpreting Bayes factors.

In fitting these models, I adhered closely to results of the model fit statistics in selecting, in each case, the model that maximized the BIC. In contrast, Nagin (1999) has argued that if the “distinctive patterns in the data emerge” one can stop model estimation earlier, with fewer groups, at a solution that does not necessarily correspond with the best BIC or Bayes factor. I based my rationale on reasoning from the exploratory factor analysis literature that argues that underfactoring is a more serious error than overfactoring. Underfactoring collapses (combines) factors and the resulting rotated solutions have a hard-to-interpret loading structure. On the other hand, overfactoring usually only results in rotated solutions with one or more factors having none or one high loading, indicating that they do not represent major common factors (Wood, Tataryn, & Gorsuch, 1996). In SPGMM, similarly, underestimating the number of groups risks obscuring unique patterns found in small groups of children. Overestimating the number of groups in SPGMM can just produce groups that “split” a trajectory into two similar trajectories. If such unproductive splitting of one trajectory into two likeminded trajectories occurs, it is easy to discuss results by referring to all low, stable groups together, for example. When this splitting occurred, I summarily discussed likeminded groups together. More often, until the maximum BIC is reached, adding groups separated former groups into new, distinctive ones with different slopes and intercepts. I also examined the sensitivity of each univariate solution to changes in starting
values. I tested a random sample of starting values for each of the four univariate models with maximum BIC scores to ensure that these solutions were not merely local maxima.

In describing the shape of the following trajectories, I try to remedy a problem in the literature. Past studies have labeled trajectories as low, moderate, or high relative to the range of that specific problem behavior, leaving open the possibility that for different genders a high level trajectory could correspond with a very different amplitude of behavior. For both genders, I will refer to a behavior as low-level if a score of $\leq 5$ was given on the 14 items, moderate-level if a score $> 5$ but $\leq 9$ was given, and high-level if a score $> 9$ was given. Hence, trajectory groups with an intercept $\leq 5$ initially exhibit low-level behavior; trajectory groups with an intercept $> 5$ but $\leq 9$ initially exhibit moderate-level behavior; and trajectory groups with an intercept $> 9$ initially exhibit high-level behavior. Thus, “low, “moderate,” and “high” here refer to severity with respect to the sample of girls and boys, not with respect to clinical diagnostic cut-offs.

The highest BIC score corresponded with a 7-group model for boys’ externalizing behavior, a 7-group model for boys’ internalizing behavior, an 8-group model for girls’ externalizing behavior, and a 7-group model for girls’ internalizing behavior (see Tables 12, 13, 14, and 15). Each of these four best fitting models had Bayes factors that indicated “strong evidence,” on the Jeffrey’s scale, for the chosen number of groups over the next best alternatives (e.g. 6 or 8 group models), as shown in Tables 12, 13, 14, and 15. Each of these four best fitting models also had a $>.85$ probability of being the correct model (range .86-.99), as shown in Tables 12, 13, 14, and 15. The order (e.g. zero-order, linear, or quadratic) of each trajectory group within the four best-fitting univariate models is listed in Table 17.
Diagnostics of Model Adequacy. Each individual has a probability of being in each trajectory group. We would hope that there are clear cut distinctions, with the members of a given group having a high posterior probability of membership in that group, and near zero probability of being classified into other groups. The posterior probabilities of group membership tell us both how similar group members are to one another based on their levels of the dependent variable, and how different they are from members of other groups. Nagin (1999) specifies that adequate model performance corresponds with > .70 average posterior probability for members of that group being assigned to that group. Odds of Correct Classification, according to Nagin (1999), should be > 5:1 for each group for model adequacy. OCC are calculated as follows, in equation (4).

\[
\text{OCC}_j = \frac{\text{AvePP}_j}{1-\text{AvePP}_j} \quad \pi_j / 1-\pi_j
\]

(4)

The OCC calculation penalizes groups with a high proportion of children in the group, but gives an advantage to groups with high posterior probabilities of group assignment. This prevents fitting a model where the OCC are unjustly elevated because most of the children are in a certain group.

As can be seen in Table 16, the average posterior probability of group membership for each trajectory group within all four univariate models exceeds the acceptable minimum of .70. Average posterior probabilities of group membership range, for the seven groups in the boys’ externalizing model, from .80-.98; for the seven groups in the boys’ internalizing model, from .77-90; for the eight groups in the girls’ externalizing model, from .78 to .95; and for the seven groups in the girls’ internalizing model, from .74 to .94. Furthermore, the Odds of Correct Classification well exceed the acceptable minimum of 5:1 for each group within each of the four univariate models, as also shown in Table 16. Interestingly, the odds
of correct classification for the high, stable groups in each of the four univariate models are the largest, ranging from 424:1 for boys’ internalizing behavior to 1730:1 for boys’ externalizing behavior. This indicates that I can be most sure of accurate classification of those children on persistent problem trajectories, which are especially interesting to us in these analyses. Given that each group within each of the four univariate models fared well on the diagnostics of model performance, I can be confident that my model does an acceptable job of classifying children to groups and specifying the proportion of children in each group.

*Univariate Internalizing and Externalizing Models.* At the most general level, three patterns of mother-reported symptoms were found in the univariate models: stable, decreasing, and increasing. Moving to the next level of specificity, each of these symptom patterns occurred in several different forms. For example, symptom stability could take the form of groups of children following low-stable symptom trajectories, moderate-stable symptom trajectories, or high-chronic symptom trajectories. Declining symptoms could take the form of groups of children following moderate-desisting, high-desisting, or high-declining symptom trajectories. The term “desisting” is used to describe symptoms that drop, over time, to low or zero levels—i.e. that move into a low-problem range. The term “declining” is used to describe symptoms that drop over time, but never reach the low-problem range—i.e. “high-declining” symptoms might only drop from high to moderate levels from 2-11 years. Increasing symptoms mainly took a single form, as moderate-rising symptoms.

The univariate models sometimes included more than one trajectory exhibiting a similar form (e.g. if trajectories #2 and #5 both showed moderate-desisting symptoms). The
most detailed depiction of the four univariate models, which graphically portrays the form of every trajectory and the percentages of children classified to every trajectory, is found in Figures 2, 3, 4, and 5.

The high, persisting group in Figures 2, 3, 4, and 5 (trajectories 7, 7, 8 and 7 respectively) are all referred to as “high-chronic groups” even though three of the four slopes are significantly different than zero (p<.05). While the two high-chronic internalizing groups do increase slightly and the two high-chronic externalizing groups do decrease slightly, they all remain the high-problem range across all timepoints. The stability of their behavior is thus most prominent, and, as such, the chronic groups were listed among the stable groups above.

Since we can think about children who follow similar forms of symptom trajectories as functionally similar (Wood, Tataryn, & Gorsuch, 1996), for clarity of presentation I henceforth will rarely distinguish trajectories by their number or distinguish among trajectories of the same form within a given model. Rather, I will combine trajectories of the same form under the same label within a model (e.g. boys’ “moderate-desisting” externalizing trajectories rather than “trajectories #2 and #5.”) Table 18 compares the proportion of children exhibiting each symptom pattern across each of the four univariate models.

Internalizing Symptoms: Univariate Models

Low Symptom Groups. Table 18 shows that over one third of all boys and girls exhibited low mother-reported internalizing problems at all timepoints. These children either followed low-stable trajectories, or trajectories dipping from low to zero over time. It is notable that no boys’ or girls’ groups exhibited zero internalizing problems as toddlers. A
somewhat larger proportion of girls than boys followed low internalizing trajectories (45% for girls versus 36% for boys).

Persistors. Nearly another third of children (36% of boys and 19% of girls), maintained stable or increasing mother-reported symptoms in the moderate or high range from 2 to 11 years (see Table 18). Such symptom patterns took the form of high-chronic trajectories, moderate-stable trajectories, or moderate-rising trajectories. There were two prominent gender differences among the symptom patterns of these children with persisting problems. First, girls had a unique rising internalizing trajectory (15% of girls) that increased from moderate to moderate/high symptoms. Second, boys had two moderate-stable groups (34% of boys), and the girls had none. Additionally, the girls’ high-chronic internalizing trajectory contained a larger proportion of the sample than did the boys’ (4% versus 2% for boys), yet was lower in amplitude (had a smaller intercept). The girls’ high-chronic trajectory hovered around a score of 10 and the boys’ trajectory hovered around a score of 12. Although both genders’ high-chronic groups showed mainly stable symptoms throughout the duration of the study, both started showing slight increases at age 8.

Decreasers. A final third of children showed decreasing mother-reported internalizing symptoms. Girls had three decreasing trajectories (36.1% of girls) compared to the boys’ two decreasing trajectories (27.7% of boys), as shown in Figures 3 and 5. Girls did have a greater number of decreasing symptom pathways and boys a greater number of moderate-stable pathways. Yet, more boys shifted from the high or moderate symptom range into the low-symptom range between ages 2 and 11. This can be seen from the fact that fewer boys had low internalizing symptoms as toddlers and a similar percentage of boys and girls had low internalizing symptoms as pre-adolescents. This apparent contradiction is resolved.
by noting that one of the girls’ decreasing trajectories only dropped from high to moderate levels, without shifting those girls into the problem-free range (i.e. without desisting).

**Internalizing Symptoms: Correlative versus SPGMM Results**

Overall, the SPGMM analysis evidenced considerable stability of mother-reported internalizing symptoms from ages 2-11 for both genders. This is indicated by the 48.8% (N=288) of girls following stable trajectories (either low- or high-stable), and the 72.3% (N=456) of boys following stable trajectories (either low-, moderate-, or high-stable). This corroborates the moderately-high lag+1 correlation results for girls’ and boys’ internalizing symptoms described earlier. Additionally, we can map the correlation attenuation observed in girls from 2-6, and 7-9 years and in boys from 3-6 years, onto Figures 3 and 5. Doing so, we see that the first correlation attenuation for girls (from 2-6 years) corresponds with the steepest decrement in internalizing symptoms for three groups; the second attenuation for girls (from 7-9 years) corresponds with an increase in internalizing for one group; and the correlation attenuation for boys (from 3-6 years) corresponds with the period of steepest decrement in internalizing symptoms for the two declining trajectory groups. Correlational analyses were unable to reveal such information on whether attenuation was due to some children showing decreases or increases in the behavior, and, if so, how rapidly. From the SPGMM analyses, we can also see that it is *not* merely the heavily-populated low-symptom groups that are driving the stable lag+1 correlations. *Rather, this stability is a function of multiple groups of children who exhibit stable mother-reported symptoms at low, moderate, and high levels.*

**Externalizing Symptoms: Univariate Models**
**Low Symptom Groups.** Similar proportions of girls and boys exhibited consistently low mother-reported rates of externalizing symptoms from 2-11 years; 43% of girls followed one of three low externalizing trajectories, and 44% of boys followed one of two low externalizing trajectories (see Figures 2 and 4 and Table 18). These low externalizing trajectories took the form of either low-stable symptoms or low rates that subsequently dipped to zero. Paralleling the findings from the internalizing models, there were no children exhibiting zero externalizing problems as toddlers.

**Persistors.** Only 11% of boys and 19% of girls exhibited persistent mother-reported problems--stable rates of externalizing problems in the moderate to high range. These persistent problem pathways either took the form of moderate-stable or high-chronic trajectories. No boys or girls followed trajectories showing increasing levels of externalizing symptoms. A slightly greater proportion of girls exhibited moderate-stable externalizing symptoms (15% of girls versus 9% of boys), and a slightly greater proportion of girls exhibited high-chronic symptoms (4% of girls versus 2% of boys). The amplitude of the girls’ moderate-stable externalizing group (which hovered around a score of 6) and the girls’ high-chronic group (which hovered around a score of 12), however, were less than their counterpart boys’ moderate-stable group (which hovered around a score of 9) and boys’ high-chronic group (which hovered around a score of 13). Additionally, while remaining consistently within the high symptom range (i.e. scores > 9) from ages 2-11, both the boys’ and the girls’ high-chronic groups appeared to start slightly decreasing around age 8.

**Decreasers.** The percentage of girls on decreasing trajectories (39%) was similar to the percentage of boys on decreasing trajectories (45%), as shown in Table 18. Additionally, boys and girls showed two similar types of decreasing trajectories: high-desisting, and
moderate-desisting (see Figures 2 and 4). As in the internalizing model, the third type, a high-declining trajectory, was only found among girls. Interestingly, as with internalizing symptoms, there was considerably more shifting over time from the moderate/high problem range to the low problem range for boys. Only 36% of boys had low externalizing symptoms at age 2, but 89% did so at age 11; for girls, 43% had low externalizing symptoms at age 2, but 64% did so at age 11. Strikingly, fully 36% of girls maintained moderate to high externalizing symptoms across all timepoints.

Externalizing Symptoms: Correlations versus SPGMM Results

Overall, SPGMM results indicated substantial mother-reported symptom stability from 2-11 years, mirroring the strong lag+1 correlations for externalizing symptoms. Fully 61.1% (N= 362) of girls and 55.4% (N= 350) of boys followed stable trajectories (either low-, moderate-, or high-stable). From the SPGMM analyses, we can see that the stability is a function of multiple groups of children who exhibit stable symptoms at low, moderate, and high levels, not just due to the sizable group of children consistently exhibiting low externalizing symptoms. Furthermore, SPGMM results now clarify that the low lag+1 correlations between 2-4.5 years for girls and 3-6 years for boys correspond with the period of steepest decrement in two externalizing trajectories (see Figures 2 and 4).

Conditional and Joint Probabilities of Group Membership

Hypothesized theories of symptom covariation were investigated in two ways: through examining patterns of conditional probabilities outputted by PROC TRAJ, and through comparing joint probabilities outputted by PROC TRAJ with hypothetical “independent-model” probabilities. Each is described here, before turning to the results.
Conditional Probabilities. As described in the methods section, we can describe symptom co-occurrence through examining the conditional probabilities of membership in \( j \) trajectory on one symptom dimension given membership in \( k \) trajectory on the other symptom dimension (i.e. \( \pi_{jk} \)) and of membership in \( k \) trajectory on one symptom dimension given membership in \( j \) trajectory on the other symptom dimension (i.e. \( \pi_{kj} \)). These conditional probabilities are presented in Figures 6, 7, 8, and 9. Looking at Figure 6, for example, the lines represent probabilities; the left column of boxes represents internalizing trajectory patterns; and the right column of boxes represents externalizing trajectory patterns. All the lines that originate from each box in the left column of boxes cumulatively total 1.00. It is intuitive that the conditional probabilities sum to 1.00 because each member of a given internalizing trajectory group must follow some sort of externalizing trajectory as well; we are just parsing out the likelihoods of which this will be. We read Figure 6 left to right, as follows, for the first box: "Given membership in a low, stable internalizing trajectory, boys have a 97% chance of also following a low, stable externalizing trajectory and a 3% chance of following a moderate-desisting externalizing trajectory." In Figures 6, 7, 8, and 9 conditional probabilities over .25 are shown in bold to highlight the more likely patterns of co-occurrence.

Because there is a wealth of information portrayed in each figure, and because the examination of these conditional probabilities must be hypothesis driven to avoid capitalization on chance variations in the dataset, I will not dissect the meaning of each separate conditional probability. Rather I will highlight the patterns that correspond with each of my first four hypotheses. In fact, we will see that these four hypotheses together
explain the behavior of 94% of the sample of boys and 88% of the sample of girls (without
doubly counting children whose pattern of behavior supports two hypotheses).

Joint Probabilities. Secondarily, I will examine the joint probabilities of group
membership (which indicate what proportion of the sample synonymously follows $j$
internalizing trajectory and $k$ externalizing trajectory; i.e. $\pi_{jk}$). Joint probabilities tell us the magnitude of importance we should attribute to any given conditional probability pattern, on the basis of whether the pattern applies to a small subset or large subset of children.

Joint probabilities of group membership for each pair of internalizing and
externalizing trajectories will be compared to probabilities calculated from a hypothetical independent model (following the logic of Configural Frequency Analysis; Bergman, 1998). This independent model was constructed by calculating the proportion of the sample that would follow each pair of internalizing and externalizing trajectories (given the number of children in each trajectory) if both symptoms were unrelated, and comparing these to the observed joint probabilities. The rule of thumb adopted is that when the observed joint probabilities are twice as large as the hypothetical independent probabilities, this constitutes evidence that the observed probabilities are substantively different from the hypothetical. Cells meeting this criterion are shown in bold in Tables 19 and 20.

Boys’ Conditional Probabilities

Hypothesis #1. There was some support for differentiation occurring among boys’ mother-reported problem behaviors. A subset of boys who had both externalizing and internalizing elevations in toddlerhood maintained internalizing elevations over time while their externalizing problems desisted. This differentiation pattern was more likely if the boys’ internalizing levels were moderate-stable, not high-chronic. (If their internalizing levels
remained high over time, their externalizing levels tended to remain stable as well.)

Specifically, boys with moderate-stable internalizing symptoms had a one in two chance of having moderate-desisting externalizing symptoms, in Figure 6. The reverse probability was nearly equivalent (boys with moderate-desisting externalizing symptoms had a .45 chance of also being moderate-stable internalizers, in Figure 7). In comparison, boys with high-chronic internalizing symptoms had a small (\(\pi_{jk} = .08\)) chance of also exhibiting high-desisting externalizing symptoms; (the reverse conditional probability, \(\pi_{kj}\), was nearly equivalent.)

The joint probabilities in Table 19 show that this differentiation with development pattern of stable internalizing elevations co-occurring with desisting externalizing elevations applies to slightly more than 19% of the sample. This joint probability is nearly double the hypothetical independent probability, suggesting that it may be clinically meaningful. Since this pattern applied only to a subset of boys, it was not detected in the omnibus correlation tests that summarized change over the whole sample. Correlational analyses therefore misled us in obfuscating this result.

Note also that no similar pattern of differentiation existed for boys with initial co-occurring elevations whose externalizing symptoms persisted but internalizing symptoms desisted.

**Hypothesis #2.** There was substantial support for the existence of Internalizing Subtype #1: boys with co-occurring, stable mother-reported internalizing and externalizing symptom elevations. The joint probabilities in Table 19 indicate that Internalizing Subtype #1 describes the behavior of a substantial subset of the sample (12% of boys). This is quadruple the hypothetical independent probability, in which only 3% of boys would have shown high- or moderate-stable symptoms on both dimensions. Although no formal criteria
exist for comparing these independent and observed probabilities, from these results, it seems reasonable to conclude that this finding is clinically meaningful.

Yet the conditional probabilities show that this finding is complex. To summarize, an observation of elevated externalizing symptoms in boys from ages 2-11 is synonymous with elevated internalizing symptoms from 2-11, but an observation of elevated internalizing symptoms from 2-11 does not imply an elevation in externalizing symptoms from 2-11.

Specifically, Figure 7 shows that if a boy has moderate or high mother-reported externalizing symptoms, he has a 100% chance of also having moderate-stable or high-chronic mother-reported internalizing symptoms. The reverse conditional probability is not equal and differs depending on whether we are referring to moderately-stable or high-chronic internalizers. If in a moderate-stable internalizing group, boys only have a one in three chance of being in a moderate-stable or high-chronic externalizing group. However, if in a high-chronic internalizing group, boys have a .92 probability of being in a moderate-stable or high-chronic externalizing group. So only if a boy has high and chronic internalizing symptoms is it likely that he also has elevated externalizing symptoms.

At the outset, there seemed also to be support for the existence of Subtype #2, boys with a pure, stable internalizing-only elevation. Figure 6 shows that, if on a moderate-stable internalizing trajectory, boys have a 14% chance of also having low externalizing symptoms; (the reverse conditional probability was nearly equal). As hypothesized, this pure-stable internalizing subtype had symptom levels that were moderate, not high, in magnitude. Interestingly, there was no converse pure, stable externalizing-only subtype, nor did I hypothesize that one would exist. In general, this indicates that an observation of low
externalizing symptoms in boys does not imply low internalizing symptoms, but an
observation of low internalizing symptoms implies low externalizing symptoms.

The Joint probabilities in Table 19 show that 6% of the sample of boys appeared to
form Subtype #2, a pure-internalizing elevation subtype with moderate-stable internalizing
and low-stable externalizing symptoms. Yet this finding is qualified because the joint
probability is less than half the hypothetical independent probability of 15%, and so occurs
substantially less than would be expected by chance.

Hypothesis #3. Hypothesis #3--that a subset of boys will show stable mother-reported
externalizing elevations and rising internalizing elevations--is only supported to the extent
that the boys’ high-chronic internalizing group (trajectory #7 in Figure 3) is slightly
increasing. While this group’s internalizing symptoms are mainly high and stable, the slope
of their trajectory ($\beta_{0}^{j} = .17$) is significant at the $p<.05$ level. This is sufficient for testing
hypothesis #3. The conditional probabilities show that these boys with high-chronic/(slightly
rising) internalizing symptoms have a one in two chance of showing high-chronic
externalizing symptoms; (the reverse probability is nearly equivalent). These boys also have
a one in three chance of having moderate-stable externalizing symptoms. (Yet the reverse
conditional probability shows that boys on a moderate-stable externalizing trajectory only
have 9% chance of showing high-chronic/(slightly rising) internalizing, as most show
moderate-stable internalizing instead).

The joint probabilities show that hypothesis #3 describes the behavior of a very small
portion of the sample (only 2% or N=13), but that this proportion is twice as large as would
be expected by chance (<1%), so it may be clinically meaningful for a few boys (see Table
19).
Hypothesis #4. Covariation stability—evidenced by children exhibiting similar patterns of mother-reported behavior on both symptom dimensions, was found for each trajectory group, and was somewhat more pronounced low and high levels of the behavior. Given membership in a low or high externalizing trajectory, the modal pattern would be membership in the same trajectory group for internalizing symptoms. Synonymously, given membership in a low or high internalizing trajectory, the modal pattern would be membership in the same trajectory group for externalizing symptoms. Low externalizers have a high conditional probability ($\pi_{jk}^{jk}=.82$) of also being low internalizers; high-desisting externalizers have a high ($\pi_{jk}^{jk}=.95$) probability of being high-desisting internalizers, and high-chronic externalizers have a .56 probability of being high-chronic internalizers. For each, the reverse conditional probabilities are similar.

In contrast, covariation stability is inconsistently observed in the moderate symptom range. For symptoms in the moderate range, there are differences in the likelihood of having a similar behavioral pattern on the second dimension conditional on the form of the behavior on the first dimension. Having moderate-stable externalizing symptoms implies also having moderate-stable internalizing symptoms, but the reverse conditional probability is much less frequent (.91 versus .31). Finally, having moderate-desisting internalizing symptoms implies also having moderate-desisting externalizing symptoms, but the reverse conditional probability is much less frequent (.85 versus .46).

While there is variation in the extent to which boys’ internalizing and externalizing behaviors follow similar courses depending on the level and function of the behavior, these data show that when we see a high correlation of internalizing and externalizing symptoms over time reported in the literature, this is not solely driven by the large number of children
who always exhibit low behaviors on each dimension. Rather, at each level of one behavior, there are children who show the same level of the other behavior, and this cumulative effect yields the observed covariation stability.

The joint probabilities show that the covariation stability hypothesis #4 explains the behavior of more boys than any other hypothesis (see Table 19). Fully 67% of boys exhibited the same level and functional form of mother-reported behavior on both symptom dimensions (34% were low on both; 9% were moderate-stable on both; 17% were moderate-desisting on both; 6% were high-desisting on both, and 1% were high-chronic on both). These probabilities are between two and six times greater than would be expected by chance, and so may be clinically meaningful.

Girls’ Conditional Probabilities

Hypothesis #1. There was minimal support for differentiation occurring among girls, and most of the existing symptom differentiation for girls took the form of stable or increasing mother-reported internalizing symptoms co-occurring with decreasing externalizing symptoms.

Girls with moderate-increasing internalizing symptoms had a small ($\pi = .17$) chance of showing high-decreasing externalizing symptoms. Reverse probabilities were equivalent. Further, if a girl had high-chronic symptoms on either dimension, she had between a one in three and a one in four chance of having declining symptoms on the other dimension. The reverse conditional probabilities were small: if a girl followed a high-declining trajectory on either symptom dimension, she was unlikely to have high-chronic symptoms on the other dimension.
Joint probabilities indicated that total of 5% of the sample are best described as having differentiation with development: 3% of girls had moderate-rising internalizing but high-decreasing externalizing; 1% had high-decreasing internalizing but high-chronic externalizing; and 1% had high-chronic internalizing but high-decreasing externalizing (see Table 20). The sum of these joint probabilities is not substantially different than the sum of their independent probabilities, however, so this finding seems not to be clinically meaningful. Yet, this finding corroborates with the internalizing/externalizing correlation attenuation found from ages 2-9.

Hypothesis #2. There was modest support for the existence of Subtype #1, girls with co-occurring, stable mother-reported internalizing and externalizing elevations. If a girl has high-chronic symptoms on one dimension, she has a one in two chance of also having high-chronic symptoms on the other dimension. This pattern of co-occurring high-chronic symptoms is overwhelmingly more likely than co-occurring, stable elevations that are moderate on one dimension and high on the other. No pattern of co-occurring moderate-stable symptoms was found because a moderate-stable internalizing group did not exist—girls with moderate levels of internalizing either had increasing or decreasing levels. Joint probabilities show that this subtype #1—evidencing high-chronic or moderate-stable symptoms on both dimensions—explains the behavior of 3% of the sample of girls. Albeit tiny, this represents more than twice the percentage that would be expected by chance in an independent model, and thus may be clinically meaningful.

There was minimal support for the existence of Subtype #2—girls with a pure, internalizing-only elevation. If on a moderate-rising internalizing trajectory, girls have a 3% chance of having low or no externalizing symptoms. As hypothesized, this pure-internalizing
subtype had symptom levels that were moderate, not high, in magnitude. Joint probabilities in Table 20 showed that only 1% of the sample appeared to form a pure-internalizing elevation subtype with moderate-rising internalizing and low or no externalizing symptoms, and this is much less frequent than would be expected by chance in an independent symptoms model. Thus, this appears not to be a clinically meaningful finding.

**Hypothesis #3.** There is strong evidence for a group of girls with mother-reported stable, elevated externalizing symptoms, and rising internalizing symptoms. Most girls with rising internalizing symptoms followed trajectory #2 in Figure 5, increasing from moderate to moderate-high levels from 7-11 years. These girls were most likely to have co-occurring moderate-stable externalizing symptoms ($\pi = .75$), and least likely to have high-chronic externalizing symptoms; ($\pi = .05$); (the reverse conditional probabilities were similar.)

The high-chronic internalizing group (trajectory #7 in Figure 5) also evidenced a slight rise in internalizing symptoms starting at age 8, as the slope of trajectory #7 ($\beta_j = .21$) was significantly greater than zero (p<.05). One out of two high chronic/(slightly rising) internalizers are likely to also have high-chronic externalizing symptoms.

The joint probabilities in Table 20 showed that stable, elevated externalizing and rising internalizing symptoms described the behavior of 15% of girls in the sample, and was three times that would be expected by chance in an independent model. Thus, it may be a clinically meaningful finding.

**Hypothesis #4.** Girls frequently exhibited similar mother-reported behavior on both symptom dimensions, and this covariation stability was somewhat more pronounced at low symptom levels. Unlike for boys, each of the girls’ internalizing trajectories did not have an externalizing trajectory counterpart with the same functional form, so covariation stability
could only be looked at for those groups with counterparts on both dimensions. Low externalizing girls were almost certain to have low internalizing ($\pi = .99$); high-to-moderate externalizers had a one in two chance of being high-to-moderate internalizers, and high-chronic externalizers had a one in two chance of being high-chronic internalizers. The reverse probabilities were similar. Moderate-desisting externalizers are overwhelmingly likely to be moderate-desisting internalizers, but moderate-desisting internalizers have an equal probability of being moderate-desisting or high-desisting externalizers.

Again, while there is some variation in the extent to which girls’ internalizing and externalizing behaviors follow similar courses, this indicates that when we see a high correlation of internalizing and externalizing symptoms reported over time for girls in the literature, this is not solely driven by the large number of girls who always exhibit low behaviors on each dimension.

The joint probabilities highlighted that the covariation stability hypothesis #4, explained the behavior of more girls than any of the other hypotheses (see Table 20). Fully 70% of girls exhibited the same level and functional form of mother-reported behavior on both symptom dimensions (43% were low on both; 15% were moderate-desisting on both; 10% were high-declining on both; and 2% were high-chronic on both). These probabilities are more than two times greater than would be expected by chance in the independent model, and so may be clinically meaningful. This corroborates with the consistently moderately-strong correlations between girls’ internalizing and externalizing scores across timepoints.

Hypothesis #5: Gender Differences.

Differentiation with Development. Differentiation occurred more often among boys, and mainly involved boys in the moderate range of symptom severity. The typical pattern of
differentiation was for the internalizing symptom to become the primary/dominant/persistent symptom, at increasing or stable-elevated rates, and for the co-occurring externalizing symptom to decline over time, respectively. Interestingly, no differentiation pattern consisting of persisting externalizing and desisting internalizing symptoms emerged for either gender.

*Internalizing Subtypes.* Subtype #1--children with co-occurring, high- or moderate-stable mother-reported elevations on both symptom dimensions—occurred 4 times as often for boys than for girls, yet for both genders co-occurrence rates exceeded chance levels. For both genders, the conditional probability of being in an elevated internalizing group given membership in an elevated externalizing group was greater than the conditional probability of being in an elevated externalizing group given membership in an elevated internalizing group. While 6% of boys and 1% of girls appeared to evidence Subtype #2—a moderate, pure-internalizing elevation with low-stable externalizing—both appeared less often than expected by chance and were not deemed clinically relevant findings.

*Rising Internalizing.* Rising internalizing symptoms among children with already elevated externalizing occurred five times more often for girls than boys. For both genders, rates of rising internalizing were substantially greater than expected by chance. Boys with chronic/slightly rising internalizing symptoms were most likely to have co-occurring high, chronic externalizing symptoms. But girls with rising internalizing symptoms were most likely to have co-occurring moderate-stable externalizing symptoms. Moreover, girls with rising internalizing symptoms had a higher conditional probability of also exhibiting declining externalizing symptoms than boys.
**Covariation Stability.** Covariation stability occurred frequently for both genders, and was somewhat more pronounced at low levels of the behaviors. Fully two thirds of girls and boys exhibited the same level and functional form of mother-reported behavior on both symptom dimensions, which far exceeded the expected rates in an independent model. Of these, more girls than boys were low on both symptom dimensions, but similar numbers of children were moderate-declining on both, and high-chronic on both.

Comparing the conditional probabilities of group membership across gender showed that, if *either* boys or girls were exhibiting low levels of one behavior, they were very likely to exhibit low levels of the other behavior, but if exhibiting high-chronic levels of one behavior, they have about a 50% chance of exhibiting high levels of the other behavior. Also, for both genders, decreasing internalizing symptoms implied decreasing externalizing symptoms, but decreasing externalizing symptoms did not imply decreasing internalizing symptoms. Covariation stability was inconsistently observed at moderate symptom levels mainly for boys: boys’ moderate stable externalizing implied high or moderate stable internalizing, but not vice versa.
CHAPTER IV

DISCUSSION

The objectives of this study were to apply a relatively new methodology, growth mixture modeling, to test and compare the efficacy of five theories of internalizing-externalizing symptom covariation, and in so doing, also a) compare GMM to traditional correlative results to validate this method, and b) chart and describe heterogeneity in each behavioral domain (internalizing and externalizing) before combining them in joint analyses. Given these multiple goals, that relate, yet stand alone, I have subdivided this discussion into seven sections. The ordering of the sections is purposeful, as the former sections in some ways build upon the latter. For example, before discussing specific GMM results, as a prerequisite, I justify GMM’s advantages over correlative methods. Each section focuses on one of the following topics: 1) demonstrated efficacy of a group-based modeling approach over traditional analytical approaches; 2) patterns of externalizing symptoms from 2-11 years; 3) patterns of internalizing symptoms from 2-11 years; 4) gender differences in longitudinal patterns of each internalizing, and externalizing behavior; 5) demonstrated support for the five main hypotheses of symptom covariation; 6) limitations; and 7) future research directions.

Efficacy of Group-Based Modeling versus Traditional Approaches.

I began this thesis by posing a host of questions about individual differences in patterns of symptom covariation that were empirically unverifiable using traditional,
variable-oriented statistical methods. (For example, what proportion of children have co-occuring, elevated symptoms over time versus patterns of alternating co-occurring and then pure symptoms? Do co-occurring symptoms increase in severity together gradually or do increases in one symptom lag increases in the other?) A more nuanced depiction of covariation patterns for boys and girls over time, I contended, was crucial for the advancement of comorbidity research. I argued that variable-oriented statistical methods may obscure the complexities in covariation patterns that the developmental psychopathology field seeks to uncover, and offered GMM as a promising person-oriented statistical complement to chart heterogeneity in covariation patterns.

*Group-based Modeling Replicates and Extends Results.* As expected, GMM recovered all of the information conveyed in correlational analyses, and also provided some unique information obscured by correlational analyses. Correlational analyses mainly showed, for boys and girls, moderately-strong a) stability of internalizing behavior across time, b) stability of externalizing behavior across time, and c) stability of associations between internalizing and externalizing symptoms across time. This replicated correlational results found in past studies (such as Gould, Bird, & Jaramillo, 1993, and Weiss & Catron, 1994). It was impossible to tell, however, from correlations whether this symptom stability/covariation stability was driven by a large group of children who consistently exhibited low behavior in both domains (or whether symptom stability occurred at all levels of the behavior). It was also impossible to tell whether periods of correlation attenuation were due to all (or some) children showing decreasing (or increasing) behavior. This pitfall has, in prior work, been referred to as the “correlation trap”—correlations calculated for a
sample do not allow researchers to draw conclusions about individuals (von Eye & Bergman, 2003).

GMM analyses successfully replicated these findings of symptom stability and covariation stability, providing some evidence of convergent validity for this new approach. GMM also resolved the aforementioned ambiguities by indicating a) the proportion of boys and girls who began with high, moderate, or low levels of the behavior, b) the proportion who exhibited change (decreasing or increasing trajectories), and c) what these rates of change were. Thus, it seems that we can only gain by replacing correlational analyses with GMM analyses in research on covarying symptoms. However, did the benefits of using GMM outweigh the costs?

Groups as a Heuristic Device: Benefits. Using GMM, and consequently adopting its group-based model as a heuristic device, risks reifying for policymakers the idea of children inhabiting fixed developmental paths--a notion that we do not support in reality, and a notion that is not engendered by variable-oriented correlative approaches (Nagin, 1999). I contend that the heuristic value of probabilistic homogenous trajectory groups is justifiable on two grounds. 1) It is indispensable for communicating the complexity of covariation patterns evidenced in this dataset. 2) It follows an established precedent in scientific thought and theory. I discuss each in turn.

1) Even after grouping similarly-behaving individuals into probabilistic trajectories, and grouping these trajectories by functional form (i.e. discussing all declining groups together), patterns of covarying symptoms were complex, within and across genders. If I had looked at mean behavior for the entire sample, I would have risked overlooking intricacies happening to some children, and risked making faulty generalizations for clinical
practice. If I had paid attention to even smaller clusters of individuals, or single individuals, findings may have become too sample-specific, and less transferable into real-world settings. Referring to children together who had a >.85 probability of exhibiting a certain behavioral pathway, was a compromise simplification in this study. It allowed a fair degree of nuanced, yet not overly-specific clinical inference.

2) Since this practice of disaggregation of a population into groups is so hotly contested in psychology (see Bergman, 1998), I highlight that my use of GMM groups as an approximation device follows an accepted precedent in other areas of science. For example, a pedagogical fixture of most chemistry curricula is Niels Bohr’s 1913 model of the atom. It explains that neutrons and protons reside in the nucleus of the atom and electrons travel around the nucleus, in orbitals, within concentric levels/shells. This model, students later learn, is a vast simplification, but one that provides the framework necessary for conceptualizing the complexities that follow. In fact, electrons no more sit on discrete circular orbitals than children travel on the discrete developmental trajectories featured in Figures 2, 3, 4, and 5. The orbital is a region of space for which the probability of finding an electron exceeds some arbitrary value (usually .95). Similarly, a trajectory is a region of space for which the probability of finding one of its group member’s symptom scores exceeds .85. Further, these orbitals/regions of space have different shapes, depending on their energy levels, just as the shape of each trajectory is allowed to vary in our GMM model. Chemists retain the heuristic Bohr model because its framework facilitates the understanding of the more detailed model. Similarly, the GMM group-based approximation makes the exploration and testing of multiple hypotheses of covariation more comprehensible—hypotheses that would never been identified using correlative methods.
Patterns of Externalizing Symptoms from 2-11 years

**Key features of Prior GMM Studies Replicated.** The longitudinal patterns of externalizing symptoms evidenced in the current study replicate key features of patterns produced in previous studies using the same analytic methodology (e.g. Brame et al., 2001, Broidy et al., 2003; Nagin & Tremblay, 1999). First, my results replicated the primary types of trajectories (persisting, declining, or low-stable) found in past studies. In other words, I similarly found no evidence of a group with “late onset” of high level externalizing for either gender, contradicting the commonly held contention that externalizing problems increase in frequency over time (Cairns & Cairns, 1994; Tremblay, Boulerice et al., 1996). While it remains to be seen whether a late-onset pattern emerges among this sample in late adolescence, from toddlerhood through preadolescence there is no evidence of the beginnings of such a trend. The present results confirm Shaw and colleagues’ (2000) finding that children who do not have elevated externalizing problems in toddlerhood are unlikely to exhibit them at any later point.

Second, like other GMM studies, I found considerable change, in the absolute sense, among children who initially exhibited high or moderate symptoms. Most of these boys had transitioned to low levels of externalizing behavior by preadolescence. For both genders, the relative sizes of the high-decreasing groups compared to the high-chronic externalizing groups further reify “Robins’ Maxim” (1978)—that "adult antisocial behavior virtually requires childhood antisocial behavior [yet] most antisocial children do not become antisocial adults.” (p. 611). Only a fraction of the initially elevated externalizers continued to evidence high symptom levels over time. This prominent decreasing trend among many early externalizers, together with the decreasing slope of even the high-*chronic* groups may, as
others have argued, be a function of general socialization processes and cognitive maturation (e.g. Lacourse et al., 2002). In most children, negotiation strategies, for example, replace physical aggression as a primary means of conflict resolution by school age (Dodge, Coie, Pettit, & Price, 1990).

Similarities between my findings and findings from past GMM studies speak to the robustness and generalizability of this mother-reported constellation of externalizing trajectories across levels of contextual risk (community versus high-risk samples), and across specific symptom domains (all externalizing symptoms, versus only physical aggression.)

*Externalizing Patterns Differed from Prior GMM in Four Ways.* Three of these four differences appear to be attributable to age/developmental factors, since we followed children from toddlerhood to preadolescence, whereas most studies of externalizing heterogeneity followed children from kindergarten to late adolescence.

No Zero-Symptom Beginners. GMM studies with older children typically find that the most populous externalizing trajectory is a no-problem group. As hypothesized, my most populous trajectories were low-problem groups. As others have suggested (e.g. Shaw et al., 2003 and NICHD ECCRN, 2004, Bongers et al., 2004), this may be because it is normative for all parents to observe some externalizing symptoms in toddlerhood, in the form of tantrums and disobedience. My results provide further confirmation that early, normative externalizing elevations occur for girls as well.

The Early Starter Pathway starts Earlier. This study confirms what others (e.g. Tremblay, 1998; Nagin & Tremblay, 2001) have theorized, but have lacked the data to explore—that the “early starter pathway” starts as early as age 2. All the person-oriented studies charting externalizing trajectories from ages 6 to mid-adolescence find that the high-
chronic and moderate-stable groups *already* have elevated symptoms by age 6. These authors have extrapolated from knowledge that the highest mean rates of aggression appear at age 2 to say that, “it is most likely that boys in the high level and chronic physically aggressive trajectories were already highly physically aggressive by age 2” (Nagin & Tremblay, 1999, p. 1193). This study confirms that these high-chronic groups are, in fact, already elevated by age 2, and for both boys and girls.

No Consistent Rank Order. Broidy and colleagues (2003) synthesized findings from six studies that tracked externalizing symptoms from children ages 6 to 15 and concluded that, “patterns of physical aggression appear to be relatively stable with some evidence of gradual increases or decreases over time but consistent rank stability across sites and sex” (p. 235). We did not evidence said rank stability. Some high-decreasing trajectories for both genders had such steep slopes that they crossed paths with other group’s trajectories. This lack of rank stability was replicated in two other person-oriented studies of externalizing problems that included preschoolers (Shaw et al., 2003, and NICHD ECCRN, 2004). If we refer back to Figures 2 and 4, we can see that the steepest drops in externalizing across groups occurred *before* the age of 6 (i.e. between 2 and 6). Further, if we *only* look at the patterns after the age of 6 (i.e. from 6-1), we *do* see predominately rank stability among the trajectories. This clarifies how Broidy and colleagues could have been misled in drawing conclusions about the rank stability of externalizing problems by not including very young children in their analyses. My finding, moreover, fits with prior evidence that the largest mean normative decreases in externalizing across the lifespan are from ages 2-5, as language abilities and emotion regulation strategies mature (Shaw, Gilliom, & Giovannelli, 2000).
Number of Groups. Although my models optimized BIC and Bayes factor model fit indices and far exceeded requirements for model adequacy, the number of groups in my optimal models (7) was at the top of the range found in other GMM studies (which is typically between 3 and 7 groups, see Broidy et al., 2003). There are four possible reasons why I may have fit more than the mean number of groups.

First, some studies (e.g. Shaw et al., 2003) had samples that were one fourth the size of my sample, and thus they may not have been able to support a model with as many groups. Second, most prior GMM studies in the externalizing field have only modeled single symptoms (usually physical aggression) and physical aggression exhibits less variability than does broad-band externalizing, and so would support fewer groups. Third, these prior studies that modeled a single physical aggression symptom over time may have evidenced weaker correlations among repeated measures over time due to less overall stability of single problem behaviors compared to broad-band externalizing (Achenbach, 1991; 1992). This also could have led to those researchers fitting optimal models with fewer numbers of groups than supported by my models. Fourth, it also may be that some researchers stopped adding groups before the maximum BIC was reached because adding more groups did not reveal new, distinctive patterns (e.g. Nagin & Tremblay, 2001). However the final groups I added each evidenced a distinctive pattern. Moreover, as emphasized by Nagin (1999) it is the shape, slopes, intercepts, and distribution of the groups that is important, not the number of groups, as “groups” are just a heuristic. Importantly, my models correspond to others in the literature on the former criteria.

*Patterns of Internalizing Symptoms from 2-11 Years*
While a number of prior GMM investigations have charted the stability and heterogeneity of externalizing problems over time, person-oriented methods have seldom been used to chart heterogeneity in internalizing symptoms. Hence, there is little prior evidence of either continuity or heterogeneity in the expression of early internalizing problems. The present study showed that longitudinal patterns of early, mother-reported internalizing symptoms are substantially different than those evidenced for externalizing symptoms in two ways.

First, the proportion of children exhibiting each main type of symptom trajectory (persisting vs. decreasing vs. low-stable) were different for internalizing and externalizing domains from ages 2-11. For internalizing symptoms, roughly one third of boys and girls showed (moderate or high) persisting trajectories, another third showed decreasing, and another third showed low trajectories. For externalizing symptoms, only one fifth of girls and boys showed (moderate or high) persisting trajectories, two fifths showed decreasing, and two-fifths showed low symptom trajectories. This suggests that, for both boys and girls, some elevation that persists throughout childhood is normative for internalizing behavior, but only decreasing or low symptoms are normative for externalizing behavior. This finding adds to growing evidence (e.g. Beitchman, Wekerle, & Hood, 1987) of the continuity of internalizing problems across childhood. This finding also indicates that internalizing problems do appear consistently in their typically-manifesting forms from as early as 2 years, in line with Luby and colleagues (2003), and in opposition to the idea that internalizing problems are completely “masked” as somatization and externalizing in young children.

Second, another difference between the longitudinal course of mother-reported internalizing and externalizing symptoms is that girls’ and boys’ high chronic externalizing
groups were slightly decreasing but girls’ and boys’ high chronic internalizing groups were slightly increasing. In other words, the internalizing symptoms of the most disturbed children in the sample got slightly worse over time and slightly better in externalizing (but, as we will discuss later, overwhelmingly remain comorbid for both).

These results suggest a multi-faceted answer to the question, posed by Zahn-Waxler, Klimes-Dougan & Slattery, (2000) of why prior evidence of the continuity of early internalizing symptoms has been lacking, since it is clearly demonstrated here, even more so than for externalizing problems. In research contexts, unique patterns of internalizing symptom continuity and change in small groups of children may have been obscured by variable-oriented methods. In clinical contexts, the high-chronic internalizing groups may have been targeted and treated for their more obvious often co-occurring externalizing problems, and the moderate-persisting internalizers who lacked externalizing problems may have gone undetected and untreated.

*Gender Differences in Longitudinal Patterns of Internalizing/Externalizing.*

*Similar Developmental Pathways Across Gender.* In line with Bongers and colleagues (2004) and Cote and colleagues (2002), we found striking similarities across gender in the developmental pathways followed for mother-reported internalizing symptoms, and for mother-reported externalizing symptoms. For each symptom domain, similar percentages of girls and boys followed each type (i.e. functional form) of trajectory (persisting vs. decreasing vs. low-stable), and there were similar numbers of trajectories evidenced for boys and girls. Hence, counter to the gender differences hypothesis #5, and in contrast with Broidy and colleagues (2003), I did not find that girls exhibited fewer numbers of trajectories of externalizing symptoms, nor were there fewer girls than boys in high, chronic groups. Broidy
and colleagues used samples of children aged 6-15, so it could be that older girls show less heterogeneity in externalizing behavior, compared to younger girls, whose symptoms are undergoing more differentiation.

Externalizing: Gender Differences in Intercepts and Slopes. Given such similarities in developmental pathways of externalizing symptoms, why then do boys’ early externalizing problems, for example, seem more serious to parents and teachers, such that they are referred for treatment more often? My results showed that boys’ persistent (elevated, stable) externalizing groups had higher intercepts than their counterpart groups among girls (following Broidy and colleagues, 2003).

One inference from this finding is that boys with elevated and persisting externalizing symptoms have more severe symptoms than their girl counterparts. Webster-Stratton (1996) and Zoccolillo (1993) disagree. They have suggested that this difference may be a product of a different phenotypic presentation of early-emerging conduct problems in boys versus girls rather than simply a more severe symptom profile for conduct-disordered young boys. Webster-Stratton found that preschool boys with conduct problems are more physically negative, and destructive/aggressive, while preschool girls with conduct problems exhibit more noncompliance and verbal bullying, which are less visible and less disruptive in the classroom. Mothers then rate boys higher on externalizing symptoms because of their greater physically negative behavior (Webster-Stratton, 1996). Zoccolillo, in fact, has argued for gender-specific criteria for conduct disorder that emphasize the detection of nonaggressive conduct problems in girls while permitting girls to have a lower threshold of physically aggressive behavior.
If Webster-Stratton is correct and a) mothers’ generally give higher CBCL ratings for physically aggressive behavior than for oppositional behavior and b) high CBCL ratings for preschool boys are driven more by physically aggressive symptoms and high CBCL ratings for preschool girls are driven more by oppositional symptoms, then we should see a) more precipitous declines among externalizing scores of initially-elevated boys (as physical aggression desists or morphs into oppositionality from ages 2-5), and b) more stability among girls’ externalizing ratings. This is exactly what I found. Boys in high-decreasing groups had higher intercepts to start with but steeper negative slopes on average than high-decreasing girls. Further, girls’ symptoms overall showed more stability and boys’ symptoms more change.

These findings together indicate that gender differences in externalizing behavior start as early as preschool, and occur for some subgroups of children (decreasing groups) more than others, and that further research is necessary to clarify how gender differences in specific symptom profiles affect symptom trajectories for boys and girls. It was helpful to map trajectories of the entire externalizing symptom domain, instead of just charting aggressive symptoms, or just charting oppositional symptoms, for it has generated hypotheses for how these symptoms might operate differently in boys and girls. Hence, future research should focus on both symptoms and both genders to fully appreciate gender differences in the longitudinal course of externalizing behavior. This could lead to a better understanding of socialization practices that lead girls and boys to act out in different ways. This would represent an important shift from the present systematic exclusion of girls from studies on early externalizing problems, and the present overly-narrow focus on single symptoms (mainly physical aggression).
Internalizing: Gender Divergence over Time. Among children with stable and elevated mother-reported internalizing symptoms, girls were more likely to get worse over time and boys were more likely to stay the same. One fifth of girls with moderate levels of internalizing symptoms at age 2 showed rising internalizing starting at age 8, and there was no similar pattern among boys. While gender differences in depression at adolescence (Cicchetti & Toth, 1998; Costello, Mustillo, & Erkanli et al., 2003), and gender differences in anxiety in middle childhood (Muris et al., 2000; Costello et al., 2003) are well established, my results clarify more specifically the nature of gender differences in mother-reported internalizing problems in early childhood, as follows.

First, there is a discernable inflection point where girls’ internalizing symptom patterns diverge from boys. This point of inflection is earlier than expected from past studies. Second, the girls whose symptoms rise at age 8, diverging from their boy counterparts, themselves had a history of some symptoms from age 2. The early, rising symptoms among one fifth of girls may have gone undetected in variable oriented analyses until a later age when more girls showed increasing symptoms—enough to raise the group mean. Currently, much research focuses on “why” there are gender differences in internalizing symptoms in adolescence. Perhaps more future research should focus on the risk and protective factors specific to girls who show high and chronic internalizing over time versus those specific to girls with rising internalizing symptoms. Further, it is important to investigate how the girls’ moderate-rising subgroup differs from their counterpart boys’ moderate subgroup, whose symptoms remained stable.

Gender Differences in Univariate Models: Theoretical Implications. Results from univariate models of boys’ and girls’ internalizing and externalizing symptoms have
implications for theories of gender differences in early psychopathology. In their 2003 review paper, Crick and Zahn-Waxler outline the three major theories of early gender differences in the development of psychopathology, as follows. Only the third theory is consistent with the results from the present study.

The first theory, called the “benign childhood hypotheses” suggests that girls are buffered from adjustment problems across early and middle childhood, whereas boys’ problems accumulate over this timeframe. This difference is attributed to girls’ relatively greater social strengths (e.g. more empathetic responding) and cognitive strengths (e.g. more developed language skills) in early childhood (Sommers, 2000). The second theory dictates that both boys and girls experience adjustment problems over early/middle childhood, yet girls’ problems are restricted to mainly internalizing problems (e.g. Keenan & Shaw, 1997). The third theory suggests that boys and girls experience significant problems in the internalizing and externalizing domains throughout childhood that are equally prevalent, and girls’ problems “just have been overlooked due to a failure to define and assess adjustment problems that are most salient to females” (Crick & Zahn-Waxler, 2003, p. 726; Rudolph, 2002).

My results are consistent with the third theoretical model that allows that both boys and girls can follow childhood-onset trajectories of pathology, and allows that different symptom constellations may be more likely for one gender versus another. Further, in-contrast with the adolescent-delimited, or adolescent-onset theories, my results bore no hints of girls developing internalizing symptoms in pre-adolescence from zero-levels in childhood (i.e. out of the blue), nor boys developing late-onset externalizing problems from zero-levels in childhood. Girls’ internalizing symptoms did rise in middle childhood, but only among
girls who had exhibited moderately elevated levels of internalizing since toddlerhood.

Hopefully these results will help to further dissipate generalizations that “during childhood girls exhibit far fewer externalizing problems (and less overall psychopathology) than boys, with no clear sex differences evident in depressive and anxiety disorders” (Zahn-Waxler, Klimes-Dougan & Slattery, 2000, p. 457).

**Demonstrated Support for Five Main theories of Covariation.**

One of the main objectives of this investigation was to empirically test five theories of early symptom covariation defined earlier: 1) differentiation with development, 2) internalizing subtypes, 3) rising internalizing, 4) covariation stability, and 5) gender differences. In short, boys’ joint trajectory analyses evidenced strong support for two of the four hypothesized internalizing-externalizing covariation patterns—*differentiation with development*, occurring for 19% of boys and *covariation stability*, occurring for 70% of boys—and weak support for the *internalizing subtypes* and *rising internalizing* hypotheses. Girls’ joint trajectory analyses also evidenced strong support for two of the four hypothesized covariation patterns. These were *rising internalizing*, occurring for 15% of girls, and *covariation stability*, occurring for 67% of girls. Weak support was obtained for the *internalizing subtypes* and *differentiation with development* hypotheses for girls. The *gender differences hypothesis* was thus partially upheld, as covariation patterns were similar across gender for two thirds of children, but different for the remainder.

Discussion will focus sequentially on each of the covariation patterns for which I obtained support, and will delineate nuances within these findings, gender differences within these findings, and theoretical rationale for these findings. Secondarily, discussion will focus on covariation patterns I did *not* obtain support for and instances where no gender differences
were found. I adopt the latter “anti-types” or “white spots” as worthy topics for discussion following Bergman and Magnusson (1997). They argue that, under circumstances of sufficient sample-size and power to test one’s hypotheses (as fulfilled in the present study) pertinent information may be obtainable from the “non-occurring developmental patterns” as these are “pathways of pattern development that for some reason are closed” (Bergman & Magnusson, 1997, p. 313).

Differentiation with Development. Evidence of heterotypic continuity was found for approximately one fifth of boys and one fifth of girls. For girls, heterotypic continuity took the form of rising internalizing symptoms among those with already-elevated externalizing symptoms; this rising internalizing pattern will be discussed subsequently in another section. For boys, heterotypic continuity took the form of differentiation from initial elevations in both externalizing and internalizing domains, to single elevations in the internalizing domain by preadolescence. In other words, for one fifth of boys the internalizing symptom became the dominant/primary symptom over time, as their externalizing behavior desisted. This is consistent with each Angold, Lilienfeld, and Jensen’s notion that a) early childhood is characterized by undifferentiated responding to stress which can manifest as dysregulation across internalizing and externalizing domains, and b) this dysregulation can eventually articulate into a single problem or disorder following cognitive and emotional maturation (Angold & Costello, 1992; Lilienfeld, Waldman, & Israel, 1994; Nottleman & Jensen, 1995).

Other researchers have replicated this finding among preschool boys and have attributed it to a cumulative process in which disruptive young boys generate social failures that lead to sadness and social anxiety, even after they have outgrown early physically-aggressive behavior (Gilliom & Shaw, 2004). Whether internalizing problems were always
present (but perhaps overlooked until externalizing problems declined) as suggested by the present results, or whether internalizing problems actually rise as externalizing problems fall, across childhood (as suggested by Tremblay and colleagues, 1996; Gilliom & Shaw, 2004) may depend on the sensitivity of detection methods for early internalizing symptoms. My results indicate that one fifth of boys had elevated internalizing problems all along (possibly indicative of a vulnerability in that domain), so it is not merely the ‘social scar’ of early rejection that precipitated later internalizing elevations. Replication of this finding is needed, however.

Neither Angold, nor Lilienfeld, nor Jensen, however, articulated whether said differentiation would occur 1) for most or a subgroup of children, or 2) for both or one gender. They also did not postulate whether 3) the directionality of the differentiation would be from co-occurring symptoms into internalizing-only, or co-occurring symptoms into externalizing-only. Hence, it is a novel finding that this differentiation was observed 1) only for a subgroup of children, who had moderate symptom severity 2) only for boys, and 3) only occurred in one direction (i.e. there was no differentiation from early co-occurring symptoms to later externalizing-only symptoms.) I tackle each of these specifics in turn.

1) The fact that children with high-chronic symptoms in one domain overwhelmingly had stable symptoms in the other domain through age 11 shows that the near-Lockian or Watsonian idea of young children beginning as an undifferentiated mass of dysregulated impulses that is honed by development into having a single areas of difficulty does not describe most children with early adjustment problems. That is, unless the bulk of differentiation occurs after age 11. Such late-onset differentiation did not seem to be the gist of the theory, as most cognitive pre-conditions for symptom articulation would have been
laid before then. Rather, the small group of high-chronic children may have such intense problems in one domain, that they inevitably interface with and affect the other domain of functioning. This follows Zoccolillo’s (1992) reasoning that the more severe the externalizing problem, the higher the likelihood of co-occurrence with emotional disorders, which can onset at the same time. These children with high-chronic and co-occurring symptoms may be those who are identified for treatment later on and who contribute to the observation that children who present for treatment are overwhelmingly comorbid. (Why differentiation mainly occurred for those with moderately, not highly severe symptoms will be further discussed in a later section.)

2) The finding that differentiation occurred only for boys, but that a similar proportion of girls experienced heterotypic continuity of co-occurring symptoms leading to rising internalizing, indicates that there is a similar phenomena occurring across genders. More research is needed to tease apart why for girls internalizing adds to externalizing and for boys internalizing replaces externalizing.

3) Finally, the finding that differentiation only occurred from co-occurring symptoms to internalizing symptoms, not co-occurring symptoms to externalizing symptoms is especially noteworthy, as it may have interesting implications for developmental aspects of comorbidity theory. Specifically, certain patterns of heterotypic continuity may be more likely at different ages/stages of development. While this study and Gilliom and Shaw’s (2004) study found heterotypic continuity from externalizing to internalizing for young children, in later-childhood and adolescence, researchers commonly find heterotypic continuity from internalizing to externalizing problems (Lilienfeld, 2003; Capakil, 1991). There are many theories that seek to explain such patterns, such as depression impairing
individuals’ concern about the consequences of their behavior, thus increasing their risk of antisocial actions (Capakil, 1991). Of course, there is other evidence of antisocial behavior preceding depression in adolescence, testifying to the interacting vulnerabilities presented by these two symptoms. I am thus merely stating that the existing evidence on heterotypic continuity (albeit scant) suggests that certain directional paths of symptom transformations are more likely than others at certain developmental stages, not that alternate patterns cannot occur. This theory remains to be investigated further, with more longitudinal datasets, ideally including children from preschool through late-adolescence.

*Internalizing Subtypes*. Neither gender evidenced internalizing subtype 1) a pure, elevation in internalizing symptoms from 2-11. In fact, this pattern occurred at rates below chance levels. Although 12% of boys and 3% of girls evidenced internalizing subtype 2) co-occurring stable elevations in both domains, I am not discussing this latter finding as evidence of the “internalizing subtypes” hypothesis. Support for “internalizing subtypes” would seem to necessitate finding more than one type, or there is in effect no “sub” to the type. Hence, I instead discuss the finding of stable, co-occurring elevations among some boys and girls within the “covariation stability” hypothesis.

It is perplexing that this *internalizing subtypes* hypothesis--for which there was the greatest accumulation of empirical backing and theoretical rationale among all of the five hypotheses--was not upheld, whereas the other hypotheses were, to varying extents. There was converging genetic (Gjone & Stevenson, 1997; Puig-Antich et al., 1989) and behavioral (Kovacs et al, 2003; Verhulst et al., 1993; Harrington et al., 1991) evidence to distinguish a less-pernicious, internalizing-only trajectory that showed homotypic continuity over time, from a stable, co-occurring trajectory that showed heterotypic continuity over time. It may be
that such internalizing subtypes only emerge in older children. Alternately, such patterns may emerge only within the clinical range of problem behaviors. Sub-diagnostic threshold, there may always be some degree of symptom co-occurrence and hence no child may appear as “purely internalizing.”

While I did find evidence consistent with a mixture of distinct subtypes of internalizing trajectories, these included subtypes showing heterotypic continuity from co-occurring symptoms to externalizing symptoms, and subtypes showing stable co-occurring symptoms—no pure-internalizing subtypes. If this finding were replicated in other samples of very young children, it would call into question whether the older children identified by Gjone, Puig-Antich and others as having “pure-internalizing” symptoms really always exhibited only internalizing symptoms. They may have experienced differentiation, or subthreshold externalizing, at some earlier point that went undetected.

*Rising Internalizing.* Hypothesis #3 was supported for both girls and boys, but by a very small number of boys, and by five times more girls. Specifically it was supported by both genders’ high-chronic groups, whose internalizing symptoms slightly rose over time while their externalizing elevations remained constant. It was also supported by the more prominent fifth of the girls’ sample who had rising internalizing symptoms starting at age 8. The latter subgroup of girls were highly likely to also have moderate and stable externalizing symptoms. These results replicate the empirical findings of Biederman and colleagues (1995) and the theoretical work of Cicchetti and Schneider-Rosen (19984) and Bird and colleagues (1993). These results indicate that it is possible that some toddler and preschool externalizing problems are stage-specific manifestations of complex later disorders that include both externalizing and internalizing symptoms, especially among girls.
Recently, Zahn-Waxler, Klimes-Dougan, & Slattery (2000) underscored the necessity of “systematic analysis of the comorbidity of internalizing and externalizing problems…to lead to more valid classification of subtypes of internalizing problems and further an understanding of the diverse conditions that constitute internalized distress” (p. 443). Person-oriented methods such as GMM seem both well-suited and necessary for achieving this objective. The subtype of one fifth of girls who have early and continuing externalizing symptoms but whose internalizing starts increasing, not in adolescence, but at age 8 has not been discussed in prior research. These girls likely would not have been identified in a variable-oriented analysis that tracked mean rates of internalizing symptoms. The contextual factors and cognitive developmental milestones occurring around age 8 that could engender such symptom changes in girls have not been sufficiently explored (Grabes & Brooks-Gunn, 1996).

Much research has focused on the biological/hormonal, body image/pubertal, intensified gender-role/interpersonal, negative life event/peer rejection stockpile of diatheses that may precipitate the onset of depression among adolescent girls. Ardent research focus on the risk factors involved in the transition to adolescence has led to unanswered questions about the prior developmental histories of the girls whose internalizing symptoms do reach clinical levels in adolescence. Were they symptom-free as young children? Were they externalizers as young children? Were they comorbid as young children? Knowledge about this history could have implications for better understanding and predicting these girls’ current, and possibly longstanding, emotional, social, and cognitive vulnerabilities. We do not, moreover, know if the girls whose internalizing symptoms rise in adolescence represent
a homogenous group. They may be a combination of early externalizers-turned-internalizers and newly-emerging internalizers.

At the least, the rationale that has been put forth for focusing most internalizing research on adolescents and preadolescents--that “it is in the early teens that the sex differences [in internalizing symptoms] become apparent”--ought to be reframed, as mother-reported sex differences are apparent earlier (Zahn-Waxler, Klimes-Dougan, and Slattery’s, 2000, p. 727).

**Covariation Stability.** Although stable elevations in internalizing and externalizing symptoms from early to middle childhood have been evidenced numerous times with correlations (e.g. Lavigne et al., 1998a), relatively little effort has been made to 1) distinguish whether this stable covariation occurs at all levels of the behavior, or 2) distinguish “how and why constellations of comorbid symptoms might differ in males and females” Crick & Zahn-Waxler, 2003). 1) In support of hypothesis #4, stable covariation occurred for all levels (high, moderate, and low) of behavior, for all functional forms of behavior (decreasing, persisting, or low-stable), and for the majority of children (far exceeding rates that would be expected if these domains were independent). This is notable, as some have theorized that such stable covariation may mainly be a phenomenon specific to the most disturbed children, in the clinical range.

While covariation stability was overwhelmingly the most common covariation pattern evidenced in the data, and was equally prevalent across genders, when we narrow our focus to stable covariation of symptoms only in the elevated range, we see gender differences. More girls than boys showed low, stable, and co-occurring mother-reported symptoms. Four times more boys than girls exhibit elevated co-occurring symptoms from 2-11, even though
overall boys did not have more adjustment problems than girls. This counters Costello and colleagues’ (2003) finding for 9-16 year olds that girls had more concurrent comorbidity than did boys over time. This difference may arise from age/developmental differences, a hypothesis meriting further exploration.

Three other features of this covariation stability finding that were not hypothesized (and thus must be deemed exploratory) are notable.

**Moderate, not Extreme Behavior Modulable.** Covariation stability was most common in the tails of the severity distribution. Children with high-chronic symptoms on one dimension tended to have high-chronic on both, and children with low symptoms on one dimension tended to have low on both, over time. *The patterns of heterotypic continuity* discussed in other sections under the rising internalizing and differentiation hypotheses occurred mainly for children in the moderate range of symptom severity. In other words, the chronic group members seem to have such ingrained problems, and the low group members seem to be so robustly well-functioning, that most of the change takes place among a delimited number of moderate symptom children. The low and high groups of children seem to have a more “stable configuration” of symptoms, if you will, than those in the moderate range.

It may be that those children in the low range of behavior in early childhood typically have genetic/environmental protective factors that outweigh their contextual risk, leading them to have low scores on both domains. The high-chronic groups may have extreme contextual risk or genetic vulnerabilities that far outweigh their protective factors and affect both internalizing and externalizing domains on an interactional, constant basis. The moderately severe groups, most interestingly, may embody children with more specific,
circumscribed risk that articulates into a single problem. Their risks may not so overwhelm their resources to produce consistent comorbidity. Alternately, it may be that mothers’ historical rater biases contribute more so to the continuity of reporting at the behavioral extremes. For example, mothers of toddlers at the behavioral extremes may have labeled their children stable “good kids” or “bad kids” throughout, while mothers of children in the moderate range may have been more flexible and open to integrating new information about their children’s behavior.

This finding would only have been uncovered with a person-oriented analysis such as GMM, and certainly deserves further exploration. Children in the moderate range may in fact show enough plasticity that they are more amenable to intervention efforts than are the high-chronic group, and thus may merit programmatic funding that is usually diverted solely to the most disturbed youth.

‘Tracking’ of Symptom Covariation. Results showed that if a child has stable symptoms that are in both internalizing and externalizing domains, their symptoms usually occur at the same level of severity. This may be partially attributable to rater-effects. Mothers may have an overarching construal of their child as having bad, medium, or good behavior, which then taints the level at which the mothers endorse symptoms on each dimension. It is also possible that rater-effects may not explain this finding entirely. If so, it seems that separate groups of children may inhabit the high ranges, medium ranges, and low ranges of pathology—if their symptoms remain stable. This speaks in favor of the often-criticized categorization between clinical and subclinical ranges of behavior. This finding is somewhat analogous to how students are thought of as “AP students” or “honors students” or “special education students” within schools, and usually are tracked to take all their classes at one
level. Of course, there has been much debate over whether the subclinical range is a discrete or a continuous entity with the most severe region of pathology; this is just one possible finding that supports the discrete viewpoint.

**Implications for Screening Efficiency, Detection.** For both genders, any stable elevation in mother-reported externalizing implies a stable elevation in internalizing, but only very high chronic internalizing implies elevations in externalizing symptoms for boys. This implies that, in the interest of screening efficiency in large prevention or intervention efforts, clinicians could assume that high internalizing problems are present if high externalizing problems are detected in either boys or girls in early childhood. Furthermore since a) there is such a widespread awareness of early externalizing problems—especially physical aggression among boys—and since b) there is so little awareness of early internalizing problems for either gender, it follows that we as clinicians are overlooking high levels of co-occurring internalizing symptoms in this age group that may necessitate treatment. What is apparently manifesting among the chronic group of children is not “masked depression” that appears in the form of pure-externalizing problems. Rather, their typical internalizing difficulties are overlooked perhaps because of the greater disruption caused by their acting out behavior.

**Limitations**

**Response Variable.** I tracked behavior problems over time using a dimensional measure rather than categorical DSM-IV diagnoses. Doing so allowed me to include and model a broader range of behavior, not only the most severe portion. However, doing so had three drawbacks. First, clinical meaningfulness and clinical interpretability was constrained due to the fact that the designations of high, moderate, and low symptom levels indicated severity relative to the sample. Thus, high-chronic group membership did not imply meeting
a clinical cut-off. Mothers of the high-chronic group members did endorse, on average, the 
equivalent of all items at the “sometimes” level, or half of the items at the “always” level. 
This level of endorsing, if extrapolated across the entire CBCL measure, would have 
exceeded the clinical cut-offs. While this practice of partitioning the scale into three 
high/moderate/low segments seems crude, it is not only standard practice among the prior 
GMM studies in the externalizing field, but it represents a small improvement. I used a 
common method for partitioning my scale, and applied it to the other studies I surveyed in 
my introduction, to try to facilitate comparisons.

Repeated measures of DSM diagnoses were not available in the current dataset, but it 
would be interesting to model internalizing and externalizing behaviors using dimensional 
and categorical methods in concert. This would have enabled us to investigate which children 
are getting recognized with diagnostic labels at which points, and would have enabled us to 
discern how the diagnosed and undiagnosed children’s behavioral trajectories differ. It would 
also be important to use early trajectory group membership to predict later DSM diagnoses, 
so we could feel more confident of the clinical relevance of certain high-chronic groups.

Another limitation of my response variable is that it is questionable whether the 14 
CBCL internalizing symptoms and 14 CBCL externalizing symptoms utilized (the only items 
on these dimensions that were invariant across the two versions of the CBCL) adequately 
represent the entire internalizing and externalizing constructs. While these two sets of 14 
items each had high internal consistency, they may have lacked completeness. The only other 
alternative would have been to try to isolate a single symptom within each domain (e.g. 
physical aggression or depression) and select items specific to that symptom to model over 
time. The drawback of this alternative is that we may have then underestimated the similarity
of girls’ and boys’ early externalizing behavior, because behaviors such as oppositionality are more similar for both genders over time than is aggression. I sacrificed specificity and reliability for a more comprehensive snapshot of the development of psychopathology over time. Follow-up studies could use hypotheses generated by this investigation to track single symptom trajectories, or joint trajectories of two single symptoms. This is needed to avoid the “lumping of behaviors that may have different developmental trajectories” (Bongers et al., 2004). It may be that covariation patterns somewhat depend on the specific symptoms chosen from each domain. For example, in community samples ADHD and anxiety co-occur at 3:1 odds, while conduct disorder and depression co-occur at 7:1 odds (Angold et al., 1999).

Covariation Estimates Confounded with Shared-Rater Variance. While using only maternal reports of internalizing and externalizing behavior is standard practice in early childhood, and allowed me to maximize the amount of useable data across the 7-year observational window, it had drawbacks. Informant-specific individual-view components could potentially have inflated my estimates of symptom co-occurrence. Differing thresholds for endorsing behaviors can result in the appearance of symptoms on both domains that are at the same level of severity (i.e. high, moderate, or low), perhaps contributing to the “tracking” phenomenon I referred to earlier. Nonetheless, stable symptom covariation occurred for both genders at rates that were between 2 and 6 times greater than would be expected by chance, so even if individual-view influences had some bearing on evidenced covariation, they were likely not responsible for the full effect. Further, because such covariation has been observed in past studies across child-report, parent-report, respondent-based interviews, and
interviewer-based interviews, it is probable that shared-rater effects are in not large part responsible for symptom covariation (Angold, Costello, & Erkanli, 1999).

**Sample.** The sample used for this study was collected for alternate purposes, and was not a random sample. It is thus questionable to whom these results generalize. The original sample was over-sampled on several indices of risk, because of expected higher attrition among these groups. So I cannot simply generalize to the community at large at the 10 recruitment sites. However, participants excluded from analyses due to missing data were somewhat more high-risk than those included, which could mean that the analysis sample does in fact correspond with a typical community composition. Further, most GMM studies are only done with high-risk populations, and it is possible that without oversampling these high-risk groups, I would not have had enough variability in behavior to replicate high-chronic groups at all (e.g. see Cote et al., 2002). I could also have included participants who were missing data from all timepoints except one. This, however, could have led to less reliable predictions of trajectory group membership than the Nagin-recommended 2 timepoints. It would, however, have allowed the inclusion of more of the high-risk demographic.

**Uncertainty and Art Involved in GMM.** In comparison to variable-oriented methods, and like other person-oriented methods, there are not as comprehensive model testing procedures nor techniques for handling errors of measurement for GMM, and GMM may put higher demands on the quality and the developmental invariance of dependent measures (Bergman & Magnusson, 1997). In this study I tried to ensure developmental invariance by selecting only items that were identical on all versions of the CBCL, from 2-11 years; tried to ensure finding an absolute not local maximum likelihood by testing random sets of starting
values; and tried to isolate clinically meaningful patterns through independent-model tests. But although my models optimized fit criteria, only through replication can I tell if I have overfit my data and identified groups that are nuances of the sample (outliers) instead of clinically meaningful subtypes. There, for example, is no rubric for deciding: how many children need to exhibit a hypothesized pattern for us to consider that hypothesis to be upheld? I did de-emphasized the significance of the 2% of boys who supported the rising internalizing hypothesis, of my own judgment.

In this exploratory analysis, one of few to apply GMM to chart internalizing or internalizing-externalizing covariation patterns, I tended follow Jerome Kagan’s advice to err on the side of considering small subgroups to be potentially clinically meaningful and awaiting replication (Kagan, Snidman, & Arcus, 1998). That is how he noticed and, through replication, identified a unique behaviorally-inhibited group of ≤10% of children.

Future Research Directions

A number of my future research ideas appear elsewhere, peppered throughout the discussion of this study’s findings. I here touch on three other compelling research directions, the importance of which are underscored by the present findings.

What common features underlie the observed stable covariation? If we follow the lead of Weiss, Susser, & Catron (1998) in conceptualizing childhood disorders in terms of three levels of generality and specificity: common features distinguishing internalizing and externalizing disorders from normality; broad-band specific features distinguishing internalizing disorders from externalizing disorders; and narrow-band specific features discriminating syndromes within each domain, we can think about searching for common features to explain the robust covariation stability I observed for internalizing and
externalizing symptoms. Lilienfeld (2003) suggests one common feature—a shared disposition for negative emotionality—which is supported by evidence that children with high levels of negative emotionality have higher maternal-reported rates of comorbidity (Keiley et al., 2003).

*What processes Engender the Differentiation of Pathology?* It also seems pertinent to investigate what processes facilitate the differentiation and heterotypic continuity of psychopathology. In other words, what are the processes occurring within the comorbid constellation which differentially channel negative emotions to allow one behavior to eventually predominate (Zahn-Waxler, Klimes-Dougan, & Slattery, 2000)? Answering this question may involve searching for children who have initially co-occurring symptoms, (hence some *common features*), but who also have more *broad-band specific features* for one domain, say internalizing, and following them over time to see if they later manifest mainly internalizing problems. For example, a child with initial high negative emotionality but also high temperamental unadaptability might be expected to have initially co-occurring internalizing-externalizing symptoms, but later mainly internalizing symptoms (Keiley et al., 2003). (Keiley and colleagues have shown that child unadaptability predicts elevated internalizing symptoms but low externalizing symptoms). Investigations of contextual factors or events (e.g. life-course turning points, Elder, 1985; Nagin, Pagani, Tremblay, & Vitaro, 2003) that coincide or precipitate shifts in symptom constellations are also warranted.

*What are Familial/Social/Genetic Predictors of Trajectory Group Membership?* According to von Eye and Bergman (2003) “groupings that result from disaggregation [e.g. cluster analysis or GMM]…are externally valid if group membership can be predicted from other variables than the ones used to create the groupings or if group membership is a
predictor of differences in parameters or covariance patterns in other variables than the ones used to create the groupings” (p. 571). Other theoreticians have echoed this sentiment, and some researchers have begun to predict probabilistic trajectory group membership using familial risk variables (e.g. Gilliom & Shaw, 2004). These efforts are still underdeveloped. Unfortunately, because of computational limitations, Nagin’s SPGMM method is currently only able to use external covariates to predict trajectories of individual symptoms, not covariation patterns (Nagin and Tremblay 2001).

Relatedly, some efforts are underway to administer interventions to children who were identified by GMM to be following high-chronic trajectories. So far, one intervention has successfully shunted a sizable proportion of high-risk boys from high-level to lower-level externalizing trajectories (Lacourse et al., 2002). While such studies must be careful to emphasize the probabilistic nature of trajectory group membership, their findings are striking and may generate more policy attention than would other variable-oriented methodological approaches. Ongoing longitudinal intervention studies might also benefit from applying GMM techniques to find out precisely what behaviors they are successfully decreasing, and for which children.
REFERENCES


101


Table 1:

SPGMM Trajectory Groups for School-Age to Adolescent, Preschool to School-Age Samples

<table>
<thead>
<tr>
<th>N</th>
<th>LS</th>
<th>L/MD</th>
<th>MS</th>
<th>HD</th>
<th>HS</th>
<th>Symp.</th>
<th>Ages</th>
<th>Sample</th>
<th>Gender</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>13</td>
<td>21</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>PA</td>
<td>6-17</td>
<td>risk</td>
<td>male</td>
<td>B</td>
</tr>
<tr>
<td>17</td>
<td>--</td>
<td>52</td>
<td>--</td>
<td>28</td>
<td>4</td>
<td>PA</td>
<td>6-17</td>
<td>risk</td>
<td>male</td>
<td>N/T</td>
</tr>
<tr>
<td>--</td>
<td>25</td>
<td>46</td>
<td>--</td>
<td>25</td>
<td>5</td>
<td>O</td>
<td>6-17</td>
<td>risk</td>
<td>male</td>
<td>N/T</td>
</tr>
<tr>
<td>--</td>
<td>20</td>
<td>45</td>
<td>--</td>
<td>30</td>
<td>6</td>
<td>H</td>
<td>6-17</td>
<td>risk</td>
<td>male</td>
<td>N/T</td>
</tr>
<tr>
<td>17</td>
<td>11</td>
<td>36</td>
<td>--</td>
<td>--</td>
<td>37</td>
<td>H</td>
<td>6-12</td>
<td>risk</td>
<td>male</td>
<td>C</td>
</tr>
<tr>
<td>36</td>
<td>21</td>
<td>25</td>
<td>--</td>
<td>18</td>
<td>1</td>
<td>H</td>
<td>6-12</td>
<td>risk</td>
<td>female</td>
<td>C</td>
</tr>
<tr>
<td>57</td>
<td>32</td>
<td>--</td>
<td>10</td>
<td>1</td>
<td>3</td>
<td>D</td>
<td>6-12</td>
<td>risk</td>
<td>female</td>
<td>C</td>
</tr>
<tr>
<td>--</td>
<td>--</td>
<td>56</td>
<td>--</td>
<td>38</td>
<td>6</td>
<td>C</td>
<td>2-8</td>
<td>risk</td>
<td>male</td>
<td>S</td>
</tr>
<tr>
<td>--</td>
<td>45</td>
<td>37</td>
<td>15</td>
<td>--</td>
<td>3</td>
<td>PA</td>
<td>2-9</td>
<td>nonrisk</td>
<td>both</td>
<td>N</td>
</tr>
<tr>
<td>19</td>
<td>--</td>
<td>74</td>
<td>--</td>
<td>7</td>
<td>1</td>
<td>PA</td>
<td>6-12</td>
<td>risk</td>
<td>male</td>
<td>BR</td>
</tr>
<tr>
<td>52</td>
<td>33</td>
<td>--</td>
<td>--</td>
<td>12</td>
<td>3</td>
<td>PA</td>
<td>6-12</td>
<td>risk</td>
<td>female</td>
<td>BR</td>
</tr>
<tr>
<td>53</td>
<td>38</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>9</td>
<td>PA</td>
<td>7-13</td>
<td>nonrisk</td>
<td>male</td>
<td>BR</td>
</tr>
<tr>
<td>57</td>
<td>--</td>
<td>43</td>
<td>--</td>
<td>--</td>
<td>PA</td>
<td>7-13</td>
<td>nonrisk</td>
<td>female</td>
<td>BR</td>
<td></td>
</tr>
<tr>
<td>57</td>
<td>32</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>11</td>
<td>PA</td>
<td>7-13</td>
<td>nonrisk</td>
<td>male</td>
<td>BR</td>
</tr>
<tr>
<td>42</td>
<td>48</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>10</td>
<td>PA</td>
<td>7-13</td>
<td>nonrisk</td>
<td>female</td>
<td>BR</td>
</tr>
<tr>
<td>64</td>
<td>29</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>7</td>
<td>PA</td>
<td>6-12</td>
<td>nonrisk</td>
<td>male</td>
<td>BR</td>
</tr>
<tr>
<td>46</td>
<td>44</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>10</td>
<td>PA</td>
<td>6-12</td>
<td>nonrisk</td>
<td>female</td>
<td>BR</td>
</tr>
<tr>
<td>36</td>
<td>33</td>
<td>--</td>
<td>22</td>
<td>--</td>
<td>10</td>
<td>PA</td>
<td>8-11</td>
<td>risk</td>
<td>male</td>
<td>BR</td>
</tr>
<tr>
<td>14</td>
<td>--</td>
<td>53</td>
<td>--</td>
<td>28</td>
<td>4</td>
<td>PA</td>
<td>6-15</td>
<td>risk</td>
<td>male</td>
<td>BR</td>
</tr>
</tbody>
</table>
Note: Trajectory group types: N = none; LS = low-stable; L/MD = low or moderate declining; MS = moderate stable; HD = high stable; HS = high stable. Symptom: PA = physical aggression; O = oppositionality; H= hyperactivity; D = delinquency; C= conduct.

Authors: B = Brame et al., 2001; N/T = Nagin & Tremblay, 2001; C=Cote et al., 2002, 2001; S= Shaw et al., 2003; N=NICHD ECCRN, 2005; BR= Broidy et al., 2003.
Table 2:

SPGMM Trajectory Groups for Late Adolescent Samples

<table>
<thead>
<tr>
<th>N</th>
<th>LS</th>
<th>L/MD</th>
<th>LR</th>
<th>HR</th>
<th>HS</th>
<th>Symp</th>
<th>Age</th>
<th>Sample</th>
<th>Gender</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>--</td>
<td>20</td>
<td>7</td>
<td>43</td>
<td>17</td>
<td>C</td>
<td>13-18</td>
<td>risk</td>
<td>male</td>
<td>C</td>
</tr>
<tr>
<td>23</td>
<td>--</td>
<td>27</td>
<td>10</td>
<td>34</td>
<td>6</td>
<td>C</td>
<td>13-18</td>
<td>risk</td>
<td>female</td>
<td>C</td>
</tr>
<tr>
<td>30</td>
<td>16</td>
<td>38</td>
<td>11</td>
<td>5</td>
<td>--</td>
<td>PA</td>
<td>11-17</td>
<td>risk</td>
<td>male</td>
<td>L</td>
</tr>
<tr>
<td>58</td>
<td>14</td>
<td>17</td>
<td>7</td>
<td>4</td>
<td>--</td>
<td>V</td>
<td>11-17</td>
<td>risk</td>
<td>male</td>
<td>L</td>
</tr>
<tr>
<td>32</td>
<td>24</td>
<td>21</td>
<td>16</td>
<td>6</td>
<td>--</td>
<td>T</td>
<td>11-17</td>
<td>risk</td>
<td>male</td>
<td>L</td>
</tr>
</tbody>
</table>

*Note: Trajectory group types: N = none; LS = low-stable; L/MD = low or moderate declining; LR = low rising; HR = high rising; HS = high stable. Symptoms: C=conduct; PA=physical aggression; V=vandalism; T=theft. Authors: C= Chung et al., 2002; L = Lacourse et al., 2002.*
### Table 3:

Probabilities of Trajectory Membership Corresponding with Hypothesized Patterns

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Hypothesized Joint Probabilities</th>
<th>Hypothesized Conditional Probabilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1: Differentiation with Development</td>
<td>elevated, chronic; desisting</td>
<td>desisting conditional on stable, elevated</td>
</tr>
<tr>
<td>#2: Internalizing</td>
<td>stable, elevated internalizing;</td>
<td>elevated, stable conditional on</td>
</tr>
<tr>
<td>Subtype 1:</td>
<td>stable, elevated externalizing</td>
<td>elevated stable</td>
</tr>
<tr>
<td>Subtype 2:</td>
<td>moderately-elevated; low, stable externalizing</td>
<td>moderately-elevated, stable</td>
</tr>
<tr>
<td>Hypothesis #3:</td>
<td>stable, elevated externalizing;</td>
<td>proportion of children in a rising</td>
</tr>
<tr>
<td>Rising Internalizing</td>
<td>increasing internalizing</td>
<td>internalizing group conditional on membership in a stable, elevated</td>
</tr>
<tr>
<td>Hypothesis #4:</td>
<td>same type of trajectory for both</td>
<td>same type conditional on same type</td>
</tr>
<tr>
<td>Covariation Stability</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4:
Jeffrey’s Evidence Scale for Bayes Factors

<table>
<thead>
<tr>
<th>Bayes Factor</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{ij} &lt; 1/10$</td>
<td>Strong evidence for model $j$</td>
</tr>
<tr>
<td>$1/10 &lt; B_{ij} &lt; 1/3$</td>
<td>Moderate evidence for model $j$</td>
</tr>
<tr>
<td>$1/3 &lt; B_{ij} &lt; 1$</td>
<td>Weak evidence for model $j$</td>
</tr>
<tr>
<td>$1 &lt; B_{ij} &lt; 3$</td>
<td>Weak evidence for model $i$</td>
</tr>
<tr>
<td>$3 &lt; B_{ij} &lt; 10$</td>
<td>Moderate evidence for model $i$</td>
</tr>
<tr>
<td>$&gt; 10$</td>
<td>Strong evidence for model $i$</td>
</tr>
</tbody>
</table>
Table 5:
Invariant Internalizing, Externalizing Symptoms Across CBCL 2/3 and CBCL 4-18

<table>
<thead>
<tr>
<th>Internalizing Domain</th>
<th>Externalizing Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Acts too young for his/her age</td>
<td>8. Can’t concentrate, can’t pay attention for long</td>
</tr>
<tr>
<td>11. Clings to adults or too dependent</td>
<td>15. Cruel to animals</td>
</tr>
<tr>
<td>19. Demands a lot of attention</td>
<td>20. Destroys his/her own things</td>
</tr>
<tr>
<td>25. Doesn’t get along well with other kids</td>
<td>21. Destroys things belonging to his/her family or others</td>
</tr>
<tr>
<td>26. Doesn’t seem to feel guilty after misbehaving</td>
<td>22. Disobedient at home</td>
</tr>
<tr>
<td>45. Nervous, high strung, or tense</td>
<td>27. Easily jealous</td>
</tr>
<tr>
<td>50. Too fearful or anxious</td>
<td>28. Eats or drinks things that are not food</td>
</tr>
<tr>
<td>54. Overtired</td>
<td>37. Gets in many fights</td>
</tr>
<tr>
<td>71. Self-conscious or easily embarrassed</td>
<td>68. Screams a lot</td>
</tr>
<tr>
<td>75. Shy or timid</td>
<td>78. Smears or plays with bowel movements</td>
</tr>
<tr>
<td>86. Stubborn, sullen, or irritable</td>
<td>87. Sudden changes in mood or feelings</td>
</tr>
<tr>
<td>102. Underactive, slow moving, or lacks energy</td>
<td>95. Temper tantrums or hot temper</td>
</tr>
<tr>
<td>103. Unhappy, sad, or depressed</td>
<td>104. Unusually loud</td>
</tr>
<tr>
<td>111. Withdrawn, doesn’t get involved with others</td>
<td>109. Whining</td>
</tr>
</tbody>
</table>
Table 6:
Intercorrelations for the Form-Invariant Internalizing and Externalizing CBCL items

<table>
<thead>
<tr>
<th>Time point (years)</th>
<th>Inter-Correlation for Internalizing</th>
<th>Inter-Correlation for Externalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>.70</td>
<td>.79</td>
</tr>
<tr>
<td>3</td>
<td>.74</td>
<td>.80</td>
</tr>
<tr>
<td>4.5</td>
<td>.73</td>
<td>.80</td>
</tr>
<tr>
<td>6</td>
<td>.75</td>
<td>.82</td>
</tr>
<tr>
<td>7</td>
<td>.73</td>
<td>.81</td>
</tr>
<tr>
<td>9</td>
<td>.77</td>
<td>.79</td>
</tr>
<tr>
<td>10</td>
<td>.77</td>
<td>.78</td>
</tr>
<tr>
<td>11</td>
<td>.77</td>
<td>.81</td>
</tr>
</tbody>
</table>
Table 7:
Boys’ Externalizing Symptom Lag+1 Correlations from 2-11 years (N=598)

<table>
<thead>
<tr>
<th></th>
<th>3 yrs</th>
<th>4.5 yrs</th>
<th>6 yrs</th>
<th>7 yrs</th>
<th>9 yrs</th>
<th>10 yrs</th>
<th>11 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
</tr>
</tbody>
</table>

2 yrs EXT .723*
3 yrs EXT .658*
4.5 yrs EXT .676*
6 yrs EXT .733*
7 yrs EXT .669*
9 yrs EXT .763*
10 yrs EXT .761*

*Note. EXT = externalizing symptoms.

* p<0.01 (2-tailed).
Table 8:
Boys’ Internalizing Symptom Lag +1 Correlations from 2-11 years (N= 598)

<table>
<thead>
<tr>
<th></th>
<th>3 yrs</th>
<th>4.5 yrs</th>
<th>6 yrs</th>
<th>7 yrs</th>
<th>9 yrs</th>
<th>10 yrs</th>
<th>11 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
</tr>
</tbody>
</table>

2 yrs INT       .603*
3 yrs INT       .555*
4.5 yrs INT     .569*
6 yrs INT       .648*
7 yrs INT       .620*
9 yrs INT       .728*
10 yrs INT      .733*

* p< 0.01 (2-tailed).

Note. INT = externalizing symptoms.
Table 9:
Girls’ Externalizing Symptom Lag +1 Correlations from 2-11 years (N= 559)

<table>
<thead>
<tr>
<th></th>
<th>3 yrs</th>
<th>4.5 yrs</th>
<th>6 yrs</th>
<th>7 yrs</th>
<th>9 yrs</th>
<th>10 yrs</th>
<th>11 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
<td>EXT</td>
</tr>
</tbody>
</table>

2 yrs EXT    .663*
3 yrs EXT    .671*
4.5 yrs EXT  .723*
6 yrs EXT    .730*
7 yrs EXT    .753*
9 yrs EXT    .783*
10 yrs EXT   .759*

* p< 0.01 (2-tailed).

Note. EXT = externalizing symptoms.
Table 10:

Girls’ Internalizing Symptom Lag+1 Correlations from 2-11 years (N=559)

<table>
<thead>
<tr>
<th></th>
<th>3 yrs INT</th>
<th>4.5 yrs INT</th>
<th>6 yrs INT</th>
<th>7 yrs INT</th>
<th>9 yrs INT</th>
<th>10 yrs INT</th>
<th>11 yrs INT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 yrs INT</td>
<td>.603*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 yrs INT</td>
<td></td>
<td>.624*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.5 yrs INT</td>
<td></td>
<td>.656*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6 yrs INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.683*</td>
<td></td>
</tr>
<tr>
<td>7 yrs INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.663*</td>
<td></td>
</tr>
<tr>
<td>9 yrs INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.720*</td>
</tr>
<tr>
<td>10 yrs INT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.725*</td>
</tr>
</tbody>
</table>

Note. INT = externalizing symptoms.

* p< 0.01 (2-tailed).
### Table 11:
Concurrent Correlations Between Internalizing and Externalizing Symptoms

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>2</th>
<th>3</th>
<th>4.5</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys</td>
<td>.675*</td>
<td>.692*</td>
<td>.655*</td>
<td>.675*</td>
<td>.688*</td>
<td>.678*</td>
<td>.716*</td>
<td>.696*</td>
</tr>
<tr>
<td>Girls</td>
<td>.686*</td>
<td>.653*</td>
<td>.666*</td>
<td>.762*</td>
<td>.689*</td>
<td>.726*</td>
<td>.736*</td>
<td>.760*</td>
</tr>
</tbody>
</table>

*Note.* *p* < 0.01 (2-tailed).
Table 12:
Boys’ Externalizing Univariate Model Fit Statistics

<table>
<thead>
<tr>
<th>No. of Groups</th>
<th>BIC</th>
<th>Bayes factor Comparison groups</th>
<th>Bayes Factor of Bayes Factor</th>
<th>Interpretation</th>
<th>Probability that model $j$ is the correct model</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-9801</td>
<td>4 versus 5 group model</td>
<td>3.03 X $10^{-32}$ (&gt;.01)</td>
<td>Strong evidence for 5 over 4 group model</td>
<td>8.57 x $10^{-21}$</td>
</tr>
<tr>
<td>6</td>
<td>-9772</td>
<td>5 versus 6 group model</td>
<td>3.86 X $10^{-13}$ (&gt;.01)</td>
<td>Strong evidence for 6 over 5 group model</td>
<td>2.22 x $10^{-8}$</td>
</tr>
<tr>
<td>7</td>
<td>-9755</td>
<td>6 versus 7 group model</td>
<td>2.22 X $10^8$ (&gt;.01)</td>
<td>Strong evidence for 7 over 6 group model</td>
<td>&gt;.99</td>
</tr>
<tr>
<td>8</td>
<td>-9871</td>
<td>7 versus 8 group model</td>
<td>4.15 X $10^{50}$ 8 over 7 group model</td>
<td>Weak evidence for 8 over 7 group model</td>
<td>2.41 x $10^{-21}$</td>
</tr>
</tbody>
</table>
Table 13:
Boys’ Internalizing Univariate Model Fit Statistics

<table>
<thead>
<tr>
<th>No. of Groups</th>
<th>BIC</th>
<th>Bayes factor</th>
<th>Bayes Factor</th>
<th>Interpretation of Bayes Factor</th>
<th>Probability that model ( j ) is the correct model</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>-9619</td>
<td>4 versus 5</td>
<td>( 1.33 \times 10^{-26} )</td>
<td>Strong evidence for 5 over 4 group model</td>
<td>( 9.24 \times 10^{-13} )</td>
</tr>
<tr>
<td>6</td>
<td>-9594</td>
<td>5 versus 6</td>
<td>( 1.75 \times 10^{-10} )</td>
<td>Strong evidence for 6 over 5 group model</td>
<td>.11</td>
</tr>
<tr>
<td>7</td>
<td>-9592</td>
<td>6 versus 7</td>
<td>.12</td>
<td>Moderate evidence for 7 over 6 group model</td>
<td>.86</td>
</tr>
<tr>
<td>8</td>
<td>-9598</td>
<td>7 versus 8</td>
<td>751.12</td>
<td>Weak evidence for 8 over 7 group model</td>
<td>( 1.10 \times 10^{-3} )</td>
</tr>
<tr>
<td>9</td>
<td>-9595</td>
<td>7 versus 9</td>
<td>25.65</td>
<td>Weak evidence for 9 over 7 group model</td>
<td>.03</td>
</tr>
</tbody>
</table>
Table 14:
Girls’ Externalizing Univariate Model Fit Statistics

<table>
<thead>
<tr>
<th>No. of Groups</th>
<th>BIC</th>
<th>Bayes factor Comparison</th>
<th>Bayes Factor</th>
<th>Interpretation of Bayes Factor</th>
<th>Probability that model (j) is the correct model</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>-9346</td>
<td>5 versus 6 group model</td>
<td>(1.03 \times 10^7) (&lt;.01)</td>
<td>Strong evidence for 6 over 5 group model</td>
<td>(2.08 \times 10^{-20})</td>
</tr>
<tr>
<td>7</td>
<td>-9324</td>
<td>6 versus 7 group model</td>
<td>(1.59 \times 10^{-10}) (&lt;.01)</td>
<td>Strong evidence for 7 over 6 group model</td>
<td>(1.68 \times 10^{-10})</td>
</tr>
<tr>
<td>8</td>
<td>-9301</td>
<td>7 versus 8 group model</td>
<td>(1.68 \times 10^{-10}) (&lt;.01)</td>
<td>Strong evidence for 8 over 7 group model</td>
<td>(&gt;.99)</td>
</tr>
<tr>
<td>9</td>
<td>-9313</td>
<td>8 versus 9 group model</td>
<td>(1.13 \times 10^5)</td>
<td>Weak evidence for 9 over 8 group model</td>
<td>(8.83 \times 10^{-6})</td>
</tr>
<tr>
<td>10</td>
<td>-9316</td>
<td>9 versus 10 group model</td>
<td>(3.59 \times 10^6)</td>
<td>Weak evidence for 10 over 8 group model</td>
<td>(2.78 \times 10^{-1})</td>
</tr>
</tbody>
</table>
Table 15: Girls’ Internalizing Univariate Model Fit Statistics

<table>
<thead>
<tr>
<th>No. of Groups</th>
<th>BIC</th>
<th>Bayes factor</th>
<th>Bayes Factor</th>
<th>Interpretation of Bayes Factor</th>
<th>Probability that model $j$ is the correct model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Comparison</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-9412</td>
<td>4 versus 5</td>
<td>$4.49 \times 10^{-18}$</td>
<td>Strong evidence for 5 group model over 4 group model</td>
<td>$3.86 \times 10^{-11}$</td>
</tr>
<tr>
<td>6</td>
<td>-9402</td>
<td>5 versus 6</td>
<td>$2.43 \times 10^{-5}$</td>
<td>Strong evidence for 6 group model over 5 group model</td>
<td>$1.59 \times 10^{-6}$</td>
</tr>
<tr>
<td>7</td>
<td>-9388</td>
<td>6 versus 7</td>
<td>$1.59 \times 10^{-6}$</td>
<td>Strong evidence for 7 group model over 6 group model</td>
<td>.99</td>
</tr>
<tr>
<td>8</td>
<td>-9397</td>
<td>7 versus 8</td>
<td>$3.85 \times 10^{3}$</td>
<td>Weak evidence for 8 group model over 7 group model</td>
<td>$2.60 \times 10^{-4}$</td>
</tr>
<tr>
<td>9</td>
<td>-9407</td>
<td>7 versus 9</td>
<td>$1.52 \times 10^{8}$</td>
<td>Weak evidence for 9 group model over 7 group model</td>
<td>$6.57 \times 10^{-9}$</td>
</tr>
</tbody>
</table>
Table 16:
Diagnostics of Model Adequacy for Univariate Models

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.88</td>
<td>40.36</td>
<td>.88</td>
<td>39.42</td>
</tr>
<tr>
<td>2</td>
<td>.77</td>
<td>15.63</td>
<td>.88</td>
<td>110.32</td>
</tr>
<tr>
<td>3</td>
<td>.82</td>
<td>6.22</td>
<td>.85</td>
<td>50.36</td>
</tr>
<tr>
<td>4</td>
<td>.83</td>
<td>57.31</td>
<td>.80</td>
<td>9.85</td>
</tr>
<tr>
<td>5</td>
<td>.77</td>
<td>61.65</td>
<td>.83</td>
<td>9.65</td>
</tr>
<tr>
<td>6</td>
<td>.85</td>
<td>68.69</td>
<td>.87</td>
<td>171.19</td>
</tr>
<tr>
<td>7</td>
<td>.90</td>
<td>424.73</td>
<td>.98</td>
<td>1730.50</td>
</tr>
</tbody>
</table>

Note. OCC = Odds of Correct Classification (Adequate > 5); Avg. PP = Average Posterior Probability of Group Membership (Adequate > .70).
Table 17:
Degree of Functions Used to Model Trajectory Groups

<table>
<thead>
<tr>
<th>Trajectory</th>
<th>Boys’ Externalizing</th>
<th>Boys’ Internalizing</th>
<th>Girls’ Externalizing</th>
<th>Girls’ Internalizing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Quadratic</td>
<td>Linear</td>
<td>Linear</td>
<td>Linear</td>
</tr>
<tr>
<td>2</td>
<td>Linear</td>
<td>Zero-order</td>
<td>Linear</td>
<td>Linear</td>
</tr>
<tr>
<td>3</td>
<td>Linear</td>
<td>Linear</td>
<td>Quadratic</td>
<td>Quadratic</td>
</tr>
<tr>
<td>4</td>
<td>Linear</td>
<td>Quadratic</td>
<td>Quadratic</td>
<td>Quadratic</td>
</tr>
<tr>
<td>5</td>
<td>Quadratic</td>
<td>Quadratic</td>
<td>Linear</td>
<td>Linear</td>
</tr>
<tr>
<td>6</td>
<td>Linear</td>
<td>Linear</td>
<td>Linear</td>
<td>Linear</td>
</tr>
<tr>
<td>7</td>
<td>Quadratic</td>
<td>Linear</td>
<td>Quadratic</td>
<td>Linear</td>
</tr>
<tr>
<td>8</td>
<td>--</td>
<td>--</td>
<td>Quadratic</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 18:
Comparison of Symptom Patterns exhibited by Children in the Four Univariate Models

<table>
<thead>
<tr>
<th></th>
<th>Low, stable (or desisting) (%)</th>
<th>Moderate-desisting (%)</th>
<th>High-desisting (%)</th>
<th>High-declining (%)</th>
<th>Moderate-Stable (%)</th>
<th>High-Chronic (%)</th>
<th>Moderate-Rising (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boys’ Externalizing</td>
<td>35.8</td>
<td>45.6</td>
<td>6.4</td>
<td>--</td>
<td>9.1</td>
<td>2.3</td>
<td>--</td>
</tr>
<tr>
<td>Boys’ Internalizing</td>
<td>36.3</td>
<td>19.3</td>
<td>8.4</td>
<td>--</td>
<td>33.7</td>
<td>2.3</td>
<td>--</td>
</tr>
<tr>
<td>Girls’ Externalizing</td>
<td>42.9</td>
<td>16.9</td>
<td>4.1</td>
<td>17.9</td>
<td>14.5</td>
<td>3.7</td>
<td>--</td>
</tr>
<tr>
<td>Girls’ Internalizing</td>
<td>44.7</td>
<td>23.0</td>
<td>--</td>
<td>13.1</td>
<td>--</td>
<td>4.1</td>
<td>15.1</td>
</tr>
</tbody>
</table>
Table 19:
Boys’ Joint Membership versus Hypothetical Independent Model Probabilities

<table>
<thead>
<tr>
<th>Externalizing Trajectory Type</th>
<th>Low-stable (or low-desisting)</th>
<th>Moderate-stable</th>
<th>Moderate-desisting</th>
<th>High-desisting</th>
<th>High-Chronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-stable</td>
<td>.34 (.15)</td>
<td>&lt;.01 (.03)</td>
<td>.01 (.12)</td>
<td>&lt;.01</td>
<td>&lt;.01, (&lt;.01)</td>
</tr>
<tr>
<td>Low-desisting</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderate-stable</td>
<td>.06 (.15)</td>
<td>.09 (.03)</td>
<td>.19 (.11)</td>
<td>&lt;.01</td>
<td>.01 (&lt;.01)</td>
</tr>
<tr>
<td>Moderate-desisting</td>
<td>.03 (.08)</td>
<td>&lt;.01 (.02)</td>
<td>.17 (.07)</td>
<td>&lt;.01</td>
<td>&lt;.01 (&lt;.01)</td>
</tr>
<tr>
<td>High-desisting</td>
<td>&lt;.01 (.03)</td>
<td>&lt;.01 (.01)</td>
<td>.02 (.03)</td>
<td>.06</td>
<td>&lt;.01 (&lt;.01)</td>
</tr>
<tr>
<td>High-chronic</td>
<td>&lt;.01 (.01)</td>
<td>.01 (&lt;.01)</td>
<td>&lt;.01 (&lt;.01)</td>
<td>&lt;.01</td>
<td>.01 (&lt;.01)</td>
</tr>
</tbody>
</table>

Notes. Hypothetical-independent probabilities are enclosed in parentheses in each cell, to the right of the observed joint probabilities. A cell is bold if hypothetical probabilities are at least half of the observed probability.
Table 20:
Girls’ Joint Membership versus Hypothetical Independent Model Probabilities

<table>
<thead>
<tr>
<th>Externalizing Trajectory Type (trajectory #)</th>
<th>Low-stable or desisting</th>
<th>Moderate-desisting</th>
<th>High-decreasing</th>
<th>Moderate-stable (6)</th>
<th>High-desisting</th>
<th>High-chronic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-stable or desisting</td>
<td>(1,2,3)</td>
<td>(5)</td>
<td>(8)</td>
<td>(4)</td>
<td>(7)</td>
<td></td>
</tr>
<tr>
<td>Low-stable/low-desist</td>
<td><strong>.43 (.21)</strong></td>
<td>&lt;.01 (.07)</td>
<td>&lt;.01 (.09)</td>
<td>.02 (.07)</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Moderate-desisting</td>
<td></td>
<td><strong>.15 (.04)</strong></td>
<td>.04 (.05)</td>
<td>&lt;.01 (.04)</td>
<td><strong>.03 (.01)</strong></td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Moderate-desisting (5, 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.01)</td>
<td></td>
</tr>
<tr>
<td>Moderate-rising (2)</td>
<td>.01 (.07)</td>
<td>&lt;.01 (.03)</td>
<td>.03 (.03)</td>
<td><strong>.11 (.02)</strong></td>
<td>&lt;.01</td>
<td>.01 (.01)</td>
</tr>
<tr>
<td>High-decreasing (6)</td>
<td>&lt;.01 (.06)</td>
<td>.02 (.02)</td>
<td><strong>.10 (.02)</strong></td>
<td>&lt;.01 (.02)</td>
<td>&lt;.01</td>
<td>.01 (.01)</td>
</tr>
<tr>
<td>High-chronic (7)</td>
<td>&lt;.01 (.02)</td>
<td>&lt;.01 (.01)</td>
<td>.01 (.01)</td>
<td>.01 (.01)</td>
<td>&lt;.01</td>
<td><strong>.02</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(&lt;.01)</td>
<td>(&lt;.01)</td>
</tr>
</tbody>
</table>
Notes. Hypothetical-independent probabilities are enclosed in parentheses in each cell, to the right of the observed joint probabilities. A cell is bold if hypothetical probabilities are at least half of the observed probability.
Figure 1:

Schematic of Hypothesized Externalizing and Internalizing Trajectories

<table>
<thead>
<tr>
<th>Trajectory</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chronic, high, stable externalizing</td>
</tr>
<tr>
<td>2</td>
<td>Moderate/high, desisting externalizing</td>
</tr>
<tr>
<td>3</td>
<td>Moderate, stable externalizing</td>
</tr>
<tr>
<td>4</td>
<td>Moderate declining to low externalizing</td>
</tr>
<tr>
<td>5</td>
<td>Low, stable externalizing</td>
</tr>
<tr>
<td>6</td>
<td>Chronic, high, stable internalizing</td>
</tr>
<tr>
<td>7</td>
<td>Moderate/high, desisting internalizing</td>
</tr>
<tr>
<td>8</td>
<td>Moderate, stable internalizing</td>
</tr>
<tr>
<td>9</td>
<td>Moderate declining to low internalizing</td>
</tr>
<tr>
<td>10</td>
<td>Low, stable internalizing</td>
</tr>
<tr>
<td>11</td>
<td>Rising internalizing</td>
</tr>
</tbody>
</table>

Note. Trajectory 1 = chronic, high, stable externalizing
Trajectory 2 = moderate/high, desisting externalizing
Trajectory 3 = moderate, stable externalizing
Trajectory 4 = moderate declining to low externalizing
Trajectory 5 = low, stable externalizing
Trajectory 6 = chronic, high, stable internalizing
Trajectory 7 = moderate/high, desisting internalizing
Trajectory 8 = moderate, stable internalizing
Trajectory 9 = moderate declining to low internalizing
Trajectory 10 = low, stable internalizing
Trajectory 11 = rising internalizing
Figure 2:
Boys Externalizing Scores versus Age: 7 Group Model

Notes.  ------ Predicted Values for each group

_____ Average Values for each group

Trajectory Group 1 “low-stable”: 18.0% of boys or N=114
Trajectory Group 2 “moderate-desisting”: 20.5% of boys or N= 129
Trajectory Group 3 “moderate-stable”: 9.1% of boys or N= 57
Trajectory Group 4 “low-stable”:  26.0% of boys or N= 164
Trajectory Group 5 “moderate-desisting”: 17.8% of boys or N= 112
Trajectory Group 6 “high-desisting”:  6.4% of boys or N= 40
Trajectory Group 7 “high-chronic”:  2.3% of boys or N= 15
Figure 3:
Boys’ Internalizing Scores versus Age: 7 Group Model

Notes. ------ Predicted Values for each group
_____ Average Values for each group

Trajectory Group 1 “low-stable”: 12.4% of boys or N=78
Trajectory Group 2 “moderate-stable”: 23.8% of boys or N=150
Trajectory Group 3 “low-stable”: 23.9% of boys or N=151
Trajectory Group 4 “moderate-stable”: 9.9% of boys or N=62
Trajectory Group 5 “moderate-desisting”: 19.3% of boys or N=122
Trajectory Group 6 “high-desisting”: 8.4% of boys or N=53
Trajectory Group 7 “high-chronic”: 2.3% of boys or N=15
Figure 4:

Girls’ Externalizing Scores versus Age: 8 Group Model

Notes: ------ Predicted Values for each group

_____ Average Values for each group

Trajectory Group 1 “low-stable”: 20.3% of girls or N=120
Trajectory Group 2 “low-stable”: 17.9% of girls or N=106
Trajectory Group 3 “low-stable”: 4.7% of girls or N=28
Trajectory Group 4 “high-desisting”: 4.1% of girls or N=24
Trajectory Group 5 “moderate-desisting”: 16.9% of girls or N=100
Trajectory Group 6 “moderate-stable”: 14.5% of girls or N=86
Trajectory Group 7 “high-chronic”: 3.7% of girls or N=22
Trajectory Group 8 “high-declining”: 17.9% of girls or N=106
Figure 5:
Girls’ Internalizing Scores versus Age: 7 Group Model

Notes.------- Predicted Values for each group

_____ Average Values for each group

Trajectory Group 1 “low-stable”: 13.7% of girls or N=81

Trajectory Group 2 “moderate rising”: 15.1% of girls or N= 89

Trajectory Group 3 “low-stable”: 31.0% of girls or N= 183

Trajectory Group 4 “moderate-desisting”: 3.5% of girls or N= 21

Trajectory Group 5 “moderate-desisting”: 19.5% of girls or N= 115

Trajectory Group 6 “high-decreasing”: 13.1% of girls or N= 77

Trajectory Group 7 “high-chronic”: 4.1% of girls or N= 24
Figure 6:
Probability of Externalizing Membership Conditional on Internalizing Membership: Boys.
Figure 7:

Probability of Internalizing Membership Conditional on Externalizing Membership: Boys
Figure 8:

Probability of Internalizing Membership Conditional on Externalizing Membership: Girls.
Figure 9:
Probability of Externalizing Membership Conditional on Internalizing Membership: Girls.