
Rachael Elizabeth Wagner
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

Part of the Developmental Psychology Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/2382

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Refining the Characterization of Causation in Early Childhood Neuropsychiatric Conditions:
Nature, Nurture, and Time
by
Rachael E. Wagner

A dissertation presented to
The Graduate School
of Washington University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

January, 2021
St. Louis, Missouri
Table of Contents

List of Figures ........................................................................................................................... v
List of Tables .......................................................................................................................... vi
Acknowledgments ................................................................................................................... vii
Chapter 1: Introduction ........................................................................................................... 1
  1.1 The Significance of Understanding Causal Risk Factors for Psychopathology ........ 1
  1.1.1 Improvement of prevention and intervention efforts ........................................ 1
  1.1.2 Challenges to determining causality ................................................................. 7
  1.2 Approaches to Refining the Characterization of Causality .................................... 19
  1.2.1 Nurture ............................................................................................................. 19
  1.2.2 Time ................................................................................................................ 20
  1.2.3 Nature .............................................................................................................. 20
  1.3 Conclusions .............................................................................................................. 21
  1.4 References ................................................................................................................ 23
Chapter 2: Characterizing Toxic Stress in Early Childhood: Maternal Depression, Maltreatment, and HPA-Axis Variation in a Pilot Intervention Study ....................................................... 32
  2.1 Abstract .................................................................................................................. 33
  2.2 Introduction .......................................................................................................... 34
  2.3 Results .................................................................................................................... 38
    2.3.1 Baseline Analysis ........................................................................................... 38
    2.3.2 Effects of Pilot TIY Intervention ................................................................. 39
  2.4 Discussion .............................................................................................................. 40
  2.5 Methods .................................................................................................................. 42
    2.5.1 Stage 1: Validation of Hair Cortisol ............................................................... 43
    2.5.2 Stage 2: Pilot Intervention of The Incredible Years ..................................... 46
  2.6 Acknowledgements ............................................................................................... 51
  2.7 Figures .................................................................................................................... 53
  2.8 References .............................................................................................................. 61
Chapter 3: Autism-Related Variation in Reciprocal Social Behavior: a Longitudinal Study ...... 66
  3.1 Abstract .................................................................................................................. 67
  3.2 Introduction .......................................................................................................... 68
  3.3 Results .................................................................................................................... 69
    3.3.1 Inter-rater Agreement ................................................................................... 69
5.3.2 Why longitudinal stability matters..........................................................150
5.3.3 The importance of identifying genetic risk..............................................152
5.3.4 Implications for rehabilitation science..................................................153
5.4 References................................................................................................. 155
List of Figures

Chapter 2
Figure 1: Subject flow ........................................................................................................66
Figure 2: Parental attendance at EHS group meetings .......................................................67

Chapter 3
Figure 1: Maternal and teacher SRS relationship ...............................................................96
Figure 2: Maternal baseline SRS scores .............................................................................97
Figure 3: Stability of the SRS ............................................................................................98
Figure 4: Individual SRS trajectories ................................................................................99
Figure 5: AVR over the life course .....................................................................................100

Supplemental Figures
Supplemental Figure 1: Individual subject span: baseline to final follow-up ..................107
Supplemental Figure 2: Recruitment and enrollment .........................................................108
Supplemental Figure 3: Subjects per age group as a function of evaluation type ..........109
Supplemental Figure 4: Median duration of follow-up .....................................................110
Supplemental Figure 5: Number of baseline evaluations per year per age group ..........111
Supplemental Figure 6: Stability of SCI subdomain .........................................................112
Supplemental Figure 7: Stability of RRB subdomain ........................................................113
Supplemental Figure 8: Trajectory of SCI over the life course .......................................114
Supplemental Figure 9: Trajectory of RRB over the life course .......................................115

Chapter 4
Figure 1: Distribution of relative frequency of PRS scores, stratified by sex ....................143
Figure 2: Distribution of relative frequency of SRS scores, stratified by sex ....................144
Figure 3: Distribution of relative frequency of PRS scores, stratified by ASD status ..........145
Figure 4: Distribution of relative frequency of SRS scores, stratified by ASD status ..........146
Figure 5: Correlation of PRS with SRS scores, stratified by sex .......................................147
Figure 6: Correlation of PRS with SRS scores, stratified by ASD status .........................148
Figure 7: Correlation of PRS with SRS scores, stratified by sex and ASD status ..............149
List of Tables

Chapter 2
Table 1: Correlation and ICC matrix of HPA-axis biomarkers .............................................68
Table 2: Descriptive summary statistics of the sample.............................................................69
Table 3: Correlation matrix of quantitative measures............................................................70
Table 4: Comparison of subjects on measures as a function of CAN group status ..................71
Table 5: Linear regressions of maternal depression and CAN on psychopathology ...............72
Table 6: Pre- and post- descriptive statistics of treatment and control groups ......................73

Chapter 3
Table 1: Maternal SRS measures ..........................................................................................101
Table 2: Teacher SRS measures .........................................................................................102
Table 3: Influences on baseline maternal SRS scores .........................................................103
Table 4: Mother and teacher test-retest correlations ............................................................104
Table 5: Model-fitting of longitudinal data ..........................................................................105
Table 6: Fixed effects estimates of model 2 .......................................................................106

Supplemental Tables
Supplemental Table 1: Correlations of SRS with ADOS-2 and ADI-R.................................117
Supplemental Table 2: Test-retest correlations of SCI and RRB subdomains ....................118

Chapter 4
Table 1: AGRE sample characteristics ..................................................................................150
Table 2: Odds ratios of ASD across quartiles of polygenic load ..........................................151
Table 3: Mean PRS scores ....................................................................................................152
Table 4: Linear mixed model of fixed and random effects .....................................................153
Acknowledgments

This dissertation reflects the time, effort, and support of so many people, only a few of whom I will be able to acknowledge here. First, I would like to give thanks to all of the members of my dissertation committee who have sacrificially given of their time to invest in my training: Ryan Bogdan, Lisa Connor, Nico Dosenbach, Chris Gunter, Allison King, and my mentor, John Constantino. These researchers have expertise across a diverse range of fields and have given me invaluable guidance and support throughout my graduate career.

Additionally, I want to thank several other scientists whose mentorship has shaped my understanding of science and strong leadership. Scott Frey, my first mentor in graduate school, cultivated my critical thinking and passion for scientific rigor, doing so with humility and deep care. Joseph Dougherty and Susan Maloney graciously allowed me the opportunity to train in the Dougherty lab to round out my education with some molecular biology and animal behavior experience. All of these individuals are exemplars of scientists who balance intellectually rigorous science with compassion and humanity, and their leadership serves as an inspiration.

The journey of graduate school would not have been possible without members of my family and the friends I have made who have supported and encouraged me, pushing not only my intellectual, but also my personal growth. Those who have enriched my life and kept me sane throughout the process include, but are not limited to Julia Erker Kesarwani, Marina Gross, Anna Abbacchi, Zoë Hawks, Ben Seitzman, Alex Hatoum, Shelly Cooper, Rebecca Davis, Johanna Burr, and Andrew Wagner.

Above all, I want to express my deepest gratitude to Dr. John Constantino, my mentor, who has provided me with a rich intellectual foundation and incredible opportunities for growth. At our first meeting, I knew I had found someone who shared my obsession with the question of
why, someone who believed that the answer to that question was critical to rational and profound treatments of psychiatric conditions. This work would not have been possible without the hours he spent conversing with me, his boundless support and encouragement, and his commitment to my growth. I am tremendously humbled and grateful.

Rachael E. Wagner

Washington University in St. Louis

January 2021
Dedicated to Marina Gross
ABSTRACT OF THE DISSERTATION


by

Rachael E. Wagner

Doctor of Philosophy in Rehabilitation and Participation Science

Washington University in St. Louis, 2021

Professor John N. Constantino, Ph.D. Chair

Understanding causal factors in the development of early childhood neuropsychiatric conditions is essential for an understanding of disease mechanisms and therapeutic approaches. Yet the lack of objective classification of psychiatric diagnoses, phenotypic and genetic heterogeneity, pleiotropy, and extensive comorbidity have posed immense challenges to the acquisition of knowledge regarding neuropsychiatric disease etiology. While it is unequivocally established that nearly all psychiatric conditions are substantially heritable, non-genetic factors do play a role in the development of psychopathology. This thesis explores both genetic and environmental contributors to neuropsychiatric conditions in an attempt to refine the characterization of some of these risk factors. In Part 1 (comprising Chapter 2), we examined putative environmental contributors to early childhood psychopathology broadly. We found that none of the factors were significantly correlated with one another, nor did they predict child behavioral and functional outcomes. These findings suggest that current measures of risk and outcome require further development and that genetically-informative study design should be employed in future interrogations of environmental contributors to psychopathology. In Part 2 (comprising Chapters 3 – 4) we focused on autism spectrum disorder (ASD) with the ultimate goal of refining current risk prediction efforts. To do so, we employed a quantitative measure of autistic traits, the Social
Responsiveness Scale (SRS), since a strong body of evidence demonstrates that ASD is the pathological tail of a continuous distribution of autism-related variation in reciprocal social behavior (AVR). As a first step, we established the longitudinal stability of the SRS to ensure its reliability as a measure over the life course. The SRS demonstrated very high measurement stability from early childhood through adulthood, as well as an ability to clearly delineate the long-term trajectory of cases and controls. Having determined the longitudinal stability of AVR, we then were able to employ the SRS in a genetic study of risk prediction. Using polygenic risk scores (PRS) derived from a genome-wide association study (GWAS) of ASD, we examined if ASD-PRS could explain AVR in a familial sample of individuals with and without a diagnosis of ASD. Despite the small sample size of our discovery GWAS dataset, ASD-PRS explained a significant, albeit modest, proportion of the variance in AVR—critically, in both affected and unaffected individuals. This work lays the foundation for future studies characterizing polygenic risk of early childhood neuropsychiatric disorders.
Chapter 1: Introduction

1.1 The Significance of Understanding Causal Risk Factors for Psychopathology

A fundamental objective of the biomedical, behavioral, public health, social, and rehabilitation sciences—in theory—is to determine causal risk and protective determinants of an array of germane phenotypes and behaviors, psychological and neurodevelopmental disorders included. Yet while the axiom that correlation does not imply correlation is undisputed, in practice, many fields largely fail to interrogate the question of causality or to design studies in such a way so as to yield the ability to judge whether or not the variable in question is playing a causal role.

Instead, in a rush to implement interventions or build public and social policy, causality is often either ignored or assumed—often without evidence, or even substantial evidence to the contrary that a given presumptive variable is determining the outcome of interest. While many observational studies may provide a substantial amount of data on the correlations of exposures (i.e., environmental factors) and psychopathology outcomes, they are frequently plagued with confounding variables and reverse causality.

1.1.1 Improvement of prevention and intervention efforts

Ignoring causality in the pursuit of an understanding of risk and protective factors for psychological or neurodevelopmental disorders results in not only theoretical and scientific compromises, but also in ethical and pragmatic ones. Finite resources can be misdirected to preventions or interventions that target variables assumed to be playing a causal role in a given disorder, but in reality are merely correlated with the outcome of interest. A long-standing question in fields from biology to epidemiology is whether phenotypic variation, including psychiatric disorders, stems from genetic or environmental factors—otherwise known as the
nature-nurture debate. Disentangling genetic and environmental influences on associations between phenotype and environment is critical, because if genetic variation mediates these relationships, assuming environmental causation is incorrect, yielding serious implications for intervention. For example, cannabis use is widely touted as a risk factor for schizophrenia (or psychosis), with many advising against cannabis consumption as a means to prevent schizophrenia (Gage, Hickman, and Zammit 2016). Yet a substantial body of genetically-informative literature suggests (conservatively) that a very minimal causal role of cannabis in the etiology of schizophrenia, with gene-environment confounding and problems of reverse causality primarily driving the association (Gillespie and Kendler 2021). And thus, interventions with the goal of preventing schizophrenia by reduction of cannabis consumption will likely fail due to the minimal variation explained.

Another important way that understanding causal risk factors is clinically significant is the ability to use this information—in particular, genetic information—for early diagnosis. Below the example of autism spectrum disorder (ASD) will illustrate this point, though the argument holds for other psychiatric disorders. The fact that ASD is a highly heritable, heterogeneous neurodevelopmental disorder has critical implications for early diagnosis and intervention. Due to the developmental nature of ASD, the general consensus is that interventions should be administered as early as possible, even before the emergence of symptoms, if possible, to shape the course of development. Yet unfortunately, according to the Centers for Disease Control and Prevention (CDC), the median age of ASD diagnosis in the United States is 51 months (Maenner et al. 2020). As work by our laboratory and others has demonstrated, by this age, the individual’s severity level will stay relatively stable throughout the rest of his or her life (Wagner et al., 2019). As demonstrated in a recent large and rigorous meta-
analysis, it is an unfortunate reality that the *majority* of current intervention approaches (with some exceptions) demonstrate very modest effects on core and related ASD symptoms; in fact, when limiting studies to RCT designs and outcomes without risk of detection bias, the authors found that there were no outcomes which demonstrated significant effects (Sandbank, Bottema-Beutel, Crowley, Cassidy, and Dunham 2019). One notable exception is an intervention entitled, “parent-mediated social communication therapy for young children with autism” (PACT), conducted in children aged 2-4, which did demonstrate significant effects of intervention, with effects lasting six years later (Pickles et al. 2016). Another recent study demonstrated improvement of school skills in children with ASD and intellectual disability (ID), though no significant effects on ASD core symptoms (Saint-Georges et al., 2020). Though there are some interventions with promise, substantial work remains to improve intervention effectiveness. One possibility is that creating developmentally-appropriate interventions and administering them in infancy, capitalizing on early development when the brain is most plastic, may prove to reduce symptom severity to a greater extent or prevent the full development of the disorder. Yet for most children with ASD, we currently lack the capability to identify them this early in life, due to the lack of established biomarkers, symptom emergence, and our underdeveloped understanding of the genetic architecture. Another potentially important factor which may be influencing intervention efficacy is the heterogeneity of ASD. Individuals diagnosed with ASD differ considerably with respect to the traits, or phenotypes (e.g., cognitive, social, emotional, behavioral, functional deficits), and actual genetic mechanisms of their particular form of the disease. Evidence has emerged that there exist different, separable strands of causal risk factors within individuals within diagnoses (Hawks et al., 2019). This heterogeneous nature may require a precise approach to optimize risk assessment, diagnosis, prevention, and intervention tailored
to the individual’s characteristics and needs. Recent studies suggest that individuals’ genetic risk for neurodevelopmental disorder (hence, genetic heterogeneity) moderates intervention effectiveness (Li et al. 2020, Tammimies et al. 2019). Therefore, using genetic information to identify individuals who respond better to various interventions; as well as efforts to uncover the genetic architecture of not only disease categories broadly, but also specific phenotypes that pose risk for disability, is an important area of research for precision health. It has been suggested that using sub-phenotypes of ASD will not provide us with greater understanding of the disorder’s genetic architecture: one small GWAS of simplex families (one affected individual per family) found that a reduction of phenotypic heterogeneity had only a modest impact on genetic homogeneity (Chaste et al. 2014). However, for a number of reasons, this study should not be considered conclusive: 1) the sample was simplex families, generally regarded to have far less common variant (inherited) contribution to ASD, and greater de novo (non-inherited) influence—yet GWAS examines only common variants; 2) the sample size was far too small to uncover meaningful effects: the most recent ASD GWAS (sample size= 18,381 individuals with ASD and 27,969 controls) revealed only a SNP-heritability of ~12%, suggesting that we need far greater sample sizes to capture the full common variant contribution to ASD. Moreover, GWAS of more deeply phentyped traits often show greater heritability than those of sparsely-phenotyped ones; for example, a recent GWAS of executive function demonstrated double the SNPheritability (20%) in deeply phenotyped subjects than in ones with only sparse phenotyping (~10%) (Hatoum et al. 2020). For these reasons, more substantial research will be needed before determining if sub-phenotyping will be informative, but current research suggests that it holds promise.
Given the largely unmet need in autism for early diagnosis, the pressing question remaining is what is the best means to address this precision medicine approach needed to advance the field. Because ASD is a highly heritable disorder, with a high degree of causality residing in genetic factors, a genetic approach is optimal for pre-symptomatic identification of the disorder. The inaccessibility and intricacy of human brain tissue and subsequent lack of understanding of disease mechanisms, have constrained the ability to identify biomarkers. Phenotypic, or behavioral assessment approaches, are very challenging to conduct in an infant population, time-consuming to administer, and most importantly, unable to provide the essential pre-symptomatic diagnostic information necessary for early identification. In contrast, genetic testing is cost-effective; non-invasive; easily-implementable across diverse populations; and critically, able to identify individuals at risk of developing heritable disorders, as well as predict other heritable traits which confer risk or protection for a given disorder. Neuroimaging is the only other modality with potential for early, pre-symptomatic identification of disease, yet in addition to being extremely costly and time-consuming (preventing its implementation among diverse populations with disparities in access to healthcare and resources), no neuroimaging study to date has been able to identify a neural signature for ASD in the general population or capture neural biomarkers of specific phenotypic trait liabilities increasing risk for neurodevelopmental disorders. Eye movement tracking holds potential as a promising biomarker, but impairment in this domain does not definitively establish the presence of ASD in an individual (Constantino et al., 2017). Therefore, incorporating a genetic approach is a promising path forward, and is key to precision medicine and targeted, individualized, early interventions, particularly for mental health disorders which are predominantly genetic in etiology. If the field of rehabilitation science is going to take seriously the prevention of and early intervention in
neurodevelopmental disorders, understanding etiology is essential to enable pre-symptomatic diagnosis and identify critical periods in development (i.e., potential windows in development in which treatment could theoretically be the most effective)—a process in which genetics plays an essential role. However, it is important to note that genetic approaches will not be effective for all ASD patients, at least not in the near future. Moreover, genetics will very rarely be the only variable employed in estimations of ASD risk; genetic information will need to be combined with other variables, such as family history, deep phenotyping of family members, and perhaps eye-tracking in infancy, to generate meaningful risk estimates. Further, genetic risk is merely that: *risk*. Our language is that of probability, not of determinism. Stochastic variation in the process of neurodevelopment limits the accuracy of the genotype-phenotype relationship, and ensures the probabilistic rather than deterministic nature of genetic risk estimates (Mitchell, 2015).

The power of genetics poses enormous ethical challenges. This is of particular importance, given the history of eugenics and its relationship with genetic science. Currently, embryonic screening based on genetic information is a common practice for parents undergoing in-vitro fertilization (IVF), as is abortion of fetuses diagnosed with various genetic conditions. As our understanding of genetics grows, so will the capacity to select on the basis of genetic information, resulting in profound technical, ethical, and even legal considerations. While these incredibly important ethical considerations are outside the scope of this dissertation, others have articulated the issues which we as a society need to consider as our understanding of genetic prediction increases, as well as our capacity for precision genome editing (e.g., the innovation of CRISPR) (Lázaro-Muñoz, et al. 2019).
1.1.2 Challenges to determining causality
Although philosophers have debated the very meaning of causality for centuries, disagreeing on necessary and sufficient criteria constituting the construct, the idea that causal relationships can be established has been unquestioned. Yet determining causality is a challenging, slow, and complex process—particularly when it is ethically and feasibly impossible to assign individuals to the condition of interest. Below are outlined some specific challenges to understanding causal determinants of neuropsychiatric conditions.

Gene-Environment Confounding
A large body of literature has established that both genes and environment are causal agents in the development of psychiatric disorders. As many have pointed out, what is intriguing about behavior genetic studies is that in addition to providing genetic information, they provide just as much information about causal environment factors (Johnson et al., 2010). With the employment of family, twin studies, and adoption studies, variation in the etiology of psychopathology can be decomposed into genetic and environmental factors: additive genetic variance (i.e., heritability), common (or shared) environmental factors, and unique (non-shared) environmental experiences specific to an individual (this estimate also includes measurement error). Heritability is the variance in a measured phenotype in a population apparently explained by inherited genetic differences. Depending on the disorder, the heritability of psychiatric conditions ranges from 30-85% (Bai et al. 2019; Pettersson et al., 2018; Xie et al. 2020). The common environment shared by individuals explains very little of the variance in psychiatric phenotypes; instead, the majority of environmental variance stems from unsystematic, random experiences. In other words, what drives sibling similarity is genetic influence, not their shared environment (Plomin, DeFries, Knopik, & Neiderhiser 2016). Contrary to what is assumed in much of child development literature (and by extension, rehabilitation efforts based on this work), differences in the family
environment within the range of normal have minimal effect on a child’s development (Plomin 2011; Plomin, DeFries, Knopik, & Neiderhiser 2016; Turkheimer 2000). A notable exception to this general principle (of shared environment accounting for a negligible amount of variation in psychiatric conditions) is severe trauma and deprivation, such as child abuse and neglect, where opportunities for species-normal development are significantly limited, resulting in the development of psychiatric conditions (even controlling for genetic risk) (Scarr, 1992; Jaffee, Caspi, Moffitt, & Taylor 2004).

The conventional assumption is that the correlation between the environment and the individual is unidirectional: environments influence variation in phenotypes, serving as the cause rather than the effect of the trait or behavior in question. However, a large body of evidence has demonstrated the untenability of this notion. Investigations of a plethora of widely-used ostensibly environmental measurements, ranging from parenting characteristics to socioeconomic status, have revealed that they are marked by pervasive and substantial genetic influence (Butcher and Plomin, 2008; Kendler and Baker, 2007; Krapohl et al., 2017; Plomin and Bergeman, 1991; Vinkhuyzen et al., 2010). As theorized in the 1980s by Scarr and McCartney, and now reinforced with substantial behavioral and molecular genetic support, throughout development genotypes at least in part drive what environment is experienced. According to their theory which now has received substantial empirical support, this occurs in three ways: passively, evocatively, and actively. Passively gene-environment correlation occurs via parental provision of both genes and environment to a child. An example of this is the correlation of the number of books in the home with the child’s IQ. While it may be easy to speculate that the books are a causal factor in the child’s IQ, the two are correlated merely because parents with high IQs tend to have more books in their home, and their children inherit their IQ (which is
around 80% heritable). A second form of gene-environment correlation is evocative. This occurs when different genotypes evoke different responses in the individuals’ physical and social environments; these responses iteratively mold the individuals’ development in a manner correlated with their genotype. For example, a child with a sweet and happy disposition will likely evoke more positive reactions from the people around her as compared to the child with an angry temperament. Active shaping of the environment by genotype occurs when individuals select the environments which are aligned with elements of their intellect, personality, and emotional makeup, such as those with athletic abilities choose to involve themselves in sports and succeed, reinforcing their genetically-driven phenotype (Jaffee 2016; Scarr and McCartney, 1983). Understanding this mechanism of genotype shaping environment is an important distinction when attempting to understand causal and modifiable risk factors for psychiatric conditions. The shared environment does not function as the primary source of the phenotypic variation, although it can play a role in shaping the phenotype (Jaffee 2016). These findings have important implications for the target of interventions, suggesting that shared environment will likely not be a fruitful area for intervention. Given the key role of genetics in all psychiatric conditions and other psychological traits, failing to account for genetic influence and gene-environment confounding can result in misleading findings and misallocated resources.

It should be emphasized that these realities do not justify a philosophy of genetic determinism; it is the combination of both genotype and environment that shape individual differences in complex traits. But an important mechanism by which environment influences the individual is by means of the individual’s genotype; it is possible that a deeper exploration of how this process occurs could be informative for our understanding of development. The reality of this phenomenon has serious implications for studies of development, brain and behavior, and
public health and intervention research. Despite this fact, acknowledging the reality of the powerful role of genetics in shaping individual differences is often an unpopular and controversial stance for a broad array of reasons. Many of these concerns are valid, founded in highly unethical historical movements such as eugenics or examples of individuals seeking to discriminate on the basis of genetic differences. As mentioned above, genetic scientists have a responsibility to carefully address these ethical issues in the communication and interpretation of their science. Another reason (of particular relevance to rehabilitation science) is the flawed assumption that heritability implies immutability. While it is true that non-random environmental factors are likely more amenable to modification, intervening on heritable conditions is possible. But the larger point is that what we may wish to be true about the etiology and thus by extension, the remediation, of psychiatric disorders has no bearing on reality. Here, there is also an ethical responsibility on the part of the scientist, and it is to accurately represent the cause of disorder. When scientists suggest or argue that intervening on a variety of environmental factors will prevent a condition or substantially remediate it, despite a strong body of evidence that said environmental variables do not play a role in the etiology of a condition, this is also unethical. In scenarios such as these, failing to acknowledge causal mechanisms can impede the development of interventions which have a profound impact on disease course and severity. In summary, both extremes of genetic determinism and also a misattribution of environmental factors as causal in a given condition are fraught with ethical and scientific problems.

**The Nature of Psychopathology**

*Psychiatric Nosology*

Another challenge facing a greater understanding of influences on psychopathology is the very nature of psychopathology itself. Perhaps unsurprisingly, it is now widely accepted that current
categorical nosology of psychopathology as outlined by the DSM-5 and the International
Classification of Disability (ICD) fail to represent natural kinds, a reality which drove the
National Institute of Mental Health (NIMH) to create a new transdiagnostic, dimensional
approach to examining psychopathology, Research Domain Criteria (RDoC) (Insel et al., 2010).
The underlying invalidity of the modern DSM disease classifications was evident even before the
advent of advanced genetic and biological techniques. The observation is now ubiquitous that
comorbidities with other forms of psychopathology and core symptom overlap are the rule rather
than the exception in psychiatry, and finding homogenous groups of psychopathological
variation with clearly demarcated boundaries is rare (Angold et al., 1999; Caron and Rutter,
1991; Krueger and Piasecki, 2002). While these comorbidities are problematic for a categorical
classification system, they are informative about the nature and etiology of psychopathology
(Angold and Costello, 2009; Lahey et al., 2011, 2017b). It was 1967 when Gottesman and
Shields observed that psychiatric disorders, and schizophrenia in particular, failed to segregate as
unitary symptom complexes (Gottesman and Shields, 1967). Since then, discoveries from
genetics, systems neuroscience, and behavioral science have served to substantiate their prescient
argument and further undermine current nosology.

**Quantitative Nature of Psychiatric Conditions**

Moreover, not only do the categories fail to represent natural kinds, but it has also become
apparent that the majority of conditions are dimensional and not categorical phenomena, with
psychiatric disorders representing pathological tails of continuous distributions. The fifth edition
of the DSM reflects this general consensus in the field, yet provides diagnostic cut points of
severity for pragmatic clinical purposes. Since individual psychiatric disorders are dimensional
in nature as well as marked by comorbidity with other diagnoses, it is perhaps no surprise that
research has uncovered that the majority (if not all) psychiatric disorders represent pathological tails of not merely of discrete diagnoses as we currently understand them, but rather the extreme end of continuous distributions of trait vulnerabilities (e.g., general cognition, reciprocal social behavior, cooperativeness, impulsivity, anhedonia, emotion regulation) in the population (Broman-Fulks et al., 2006; Crome et al., 2010; Hankin et al., 2005; R F Krueger et al., 2004; Markon and Krueger, 2005; Pickles and Angold, 2003; Prisciandaro and Roberts, 2009). Not a trivial finding, this paradigm shift has had profound implications for the exploration of neural mechanisms underlying behavior. Any study relating phenotype to endophenotype to neural mechanism to gene needs to account for the effects of variation in a variety of traits (Beauchaine and Constantino, 2017). Given the above, all available evidence indicates that categorical diagnoses are not pathophysiological entities for which there is evidence of biological coherence or an underlying single definable pathophysiology (Kendler, 2013). In the words of Meehl, they are not “carving nature at its joints” (Meehl, 2001).

**Heterogeneity**

This overwhelming consensus in the psychopathology literature that psychobiological constructs do not align neatly with categorical mental disorders, but rather cut across them transdiagnostically (Beauchaine and Constantino, 2017; Zisner and Beauchaine, 2016) has important implications. One is that the etiology of a given categorical mental disorder is heterogeneous and typically represents dysfunction in a variety of psychobiological constructs which lend vulnerability to neuropsychiatric conditions (Insel et al., 2010). This heterogeneity has been examined at both genetic epidemiological and molecular levels. One way of examining the heterogeneity of categorical psychiatric conditions is by conducting fine-grain analyses of subsets of symptoms within diagnoses. Behavior genetic analyses of individual DSM criteria
have consistently found evidence for more than one underlying dimension of genetic liability in a number of psychiatric disorders (Kendler et al., 2012, 2013). These analyses indicate differing etiological influences on different subsets of symptoms within each disorder, suggesting different neurobiological mechanisms (Kendler, 2005).

Molecular genetic analyses reveal evidence supporting genetic epidemiological and factorial analyses of categorical mental disorders, as well. Despite hopes of researchers who wished to “find a gene for [a specific trait]”, candidate gene studies, linkage studies, and genome wide association studies (GWAS) have revealed that individual gene variants of large effect appear to have a small role in the etiology of the majority of major, common psychiatric conditions (Visscher et al. 2017). In the case of ASD, a percentage of individuals possess gene variants of large effects, such as de novo structural variants, like copy number variants (CNVs) (Brandler et al., 2016; Sebat et al., 2007) and protein-altering point mutations (Iossifov et al., 2012; Neale et al., 2012; O’Roak et al., 2012; Sanders et al., 2012). Some research suggests that individuals with large CNVs experience greater disability than those without (LaBianca et al. 2020; Tammimies et al. 2019). These variants of large effect are typically accompanied by an intellectual disorder, whereas common variants for ASD overlap with that of elevated educational attainment and IQ (primarily stemming from individuals with ASD without an intellectual disorder) (Brainstorm Consortium et al., 2018; Grove et al., 2019; Iakoucheva, Muotri, & Sebat 2019). Rare inherited variants with incomplete penetrance accounts for a small percentage of ASD risk, as well. Yet not all carriers of these large-effect variants meet ASD diagnostic criteria, suggesting the multifactorial nature of genetic risk (Iakoucheva, Muotri, & Sebat 2019). A range of studies have revealed that distinct disease processes cause very few forms of psychopathology; instead, what were previously considered to be identical categorical
diseases are in reality occurrences of very similar psychiatric conditions representing indistinguishable diagnostic phenocopies which possess discrete pathological genetic mechanisms / biological processes (Bogdan et al., 2017; Coryell and Schlesser, 2001).

Heterogeneity riddles the pathways of causation within diagnosis, even within families, and even in conditions such as ASD, one of the most heritable of all forms of psychopathology (Yuen et al., 2015).

**Pleiotropy of Genetic Risk Factors**

A logical conclusion of extensive psychopathological comorbidity and heterogeneity is the pleiotropy of genetic risk factors. The extreme comorbidities among DSM-defined mental disorders, as well as somewhat arbitrary categorizing factors, prompted researchers to interrogate the taxonomy of psychopathology, seeking to organize distinct elements in a hierarchy according to their common and varying properties to reveal higher order relations among them. This extensive analysis revealed extreme overlap at all levels of the hierarchy, a construct known as pleiotropy. This pleiotropy suggests at least partially shared origins of seemingly discrete disorders (Lahey et al., 2017a; Pettersson et al., 2015).

Support for pleiotropy in the etiology of psychopathology exists at several levels of analysis. Behavior genetic studies and factor analyses have replicated the finding that these phenotypic correlations among diagnoses, termed “first-order dimensions,” vary in magnitude, with stronger correlations among some dimensions, which then result in second-order factors, categorized as externalizing and internalizing dimensions—this distinction has long been recognized (Achenbach et al., 1989). However, much more recent analyses have determined that substantial correlations exist between these second-order factors, a correlation that results from the loading of every first-order dimension on a general factor of psychopathology, otherwise
referred to as the “p-factor” (not dissimilar from the g-factor for intelligence) (Caspi et al., 2014; Lahey et al., 2012, 2017b). Extensive work has been done to establish that this p-factor is not just reflecting measurement noise. First, it is correlated with constructs central to its validity yet distinct from its definition: negative emotionality and cognitive ability. Second, it has “real-world” significance, correlating significantly with social and behavioral adaptive functioning (Lahey et al., 2017a, 2017b). Third, the p-factor is moderately heritable (Waldman et al., 2016), suggesting that contributing to the majority of phenotypic correlations of first-order dimensions are pleiotropic genetic influences (Pettersson et al., 2013, 2015) and that a considerable portion of genetic factors non-specifically escalate risk for all forms of psychopathology (Lahey et al., 2017a). As Lahey and co-authors argue, “dimensions of psychopathology are too highly correlated and there is too much sharing of genetic and environmental influences . . . not to hypothesize that variations in some neurobiological systems non-specifically underlie multiple dimensions of psychopathology” (p. 143; Lahey et al., 2017b).

Plentiful evidence at a molecular genetic level exists for pleiotropy: rare copy number variants and other structural variants (Bergen et al., 2012; Ionita-Laza et al., 2014; Levinson et al., 2011; Malhotra et al., 2011; Sanders et al., 2011) as well as common SNPs have been found to be connected with numerous psychiatric conditions (Bipolar 2018; Brainstorm 2018; Cross-Disorder 2013; Gratten et al., 2014; Lee et al., 2018). In summary, etiological heterogeneity and pleiotropy are inextricably linked because while each categorical psychiatric condition is etiologically heterogeneous with underlying dysfunction in a variety of psychobiological mechanisms, any given psychobiological mechanism may underlie numerous categorical disorders (Lahey et al., 2017a, 2017b). Further complicating our understanding of genetic function is penetrance and expressivity, which can determine whether loci are pleiotropic under
certain conditions. This pleiotropic confounding raises critical issues: 1) what uniquely differentiates highly correlated phenotypes; 2) whether the expansion of the search for variants by employing deep phenotyping of traits and disorders which covary would yield a greater insight into the genetic architecture of psychiatric disorders and uncover additional, separable genomic loci with more functional specificity. One example of the importance of careful phenotyping is ASD alone vs. ASD plus intellectual disability. Genetic studies have revealed different genetic architecture when intellectual disability accompanies ASD. Around 30-50% of individuals diagnosed with ASD have an accompanying diagnosis of intellectual disability or another NDD (Pinto et al. 2010), leading to an overlap of genes related to both disorders. But a recent exome-sequencing study conducted analyses to determine genes which are more frequently disrupted in ASD vs. NDD and found preliminary evidence supporting some functional distinctions (Satterstrom et al. 2020). These findings suggest that a careful dissection of the various genes influencing different elements of neurodevelopment may yield insight into the biological pathways differentiating ASD from NDD.

**Polygenic Nature of Psychopathology**

This genetic architecture of psychiatric disorders has only recently begun to emerge. In the last decade, our understanding of the architecture of psychiatric conditions has been drastically altered by increasingly larger genome-wide association studies (GWAS). It has become apparent that complex traits and common psychiatric disorders such as ASD are highly polygenic in nature, comprised of single-nucleotide polymorphisms (SNP) of extremely small effect sizes (odds ratios <1.3) across the genome, none of which individually would engender a given disorder, yet in aggregate account for a substantial portion of the variance of total risk (Bogdan et al., 2018). Further evidence supporting the polygenic nature of ASD includes its highly
heritable nature and the aggregation of subclinical quantitative autistic traits (QAT) and related phenotypes in first-degree family members. These findings regarding the important role of common variants in ASD support the theory that clinical diagnosis is often the pathological tail of a continuous distribution of quantitative heritable traits (Taylor et al., 2018). Even when there are highly deleterious mutations, the evidence suggests that they are operating on the background of high polygenic risk (Weiner et al., 2017). Further, an elucidation of the polygenic risk (i.e., common variants) for neuropsychiatric conditions is important not only for quantifying disease risk for individuals without known de novo mutations, but also for exploring the nature of the interactions between rare and common variants. Determining these interactions and their potential to affect the course of disease, ultimate phenotype, and response to treatment will be possible only with a complete characterization of polygenic risk (Sestan and State, 2018). Yet characterizing this polygenic risk has been more challenging than the field initially hoped. Given the initial underwhelming amount of variance explained by GWAS and relatively few genome-wide significant loci uncovered (for example, only in the past year have GWAS uncovered loci reaching the threshold for statistical significance for ASD and ADHD (Demontis et al., 2018; Grove et al., 2019)), the field of quantitative genetics developed polygenic risk scores (PRS), which aim to capture the aggregate effect of genetic variants (polygenicity), and as such, are similar to heritability ($h^2$). Researchers recognized that insufficient sample sizes in early studies produced few robust associations, but the aggregation of many loci below the genome-wide significance threshold could significantly predict disease risk in new studies. These analyses were consistent with a polygenic mode of inheritance from variants tagging causal risk (Bogdan et al., 2018). To ensure the validity of these scores, it is essential that the effect sizes are estimated in an independent cohort. A cumulative index of measured genetic liability to a
disorder, PRS quantifies within an individual the aggregate effect of common variants for a given trait, typically calculated as the sum of trait-associated alleles across the genome, weighted by effect size (Bogdan et al., 2018; Gandal et al., 2016). If the estimates of the effects of the SNPs at a population-level were free of error, then the PRS could predict the phenotype of individuals in the target data with variance explained equal to SNP-\( h^2 \) (the proportion of phenotypic variance explained by genotype). Yet due to insufficient GWAS sample sizes, imperfect effect size estimates, and variation in discovery and target datasets, PRS generally yield predictive power lower than that of SNP-\( h^2 \) (Choi et al., 2018).

PRS can be employed for the probabilistic estimation of disorder risk, identification of disease trajectories, generation of mechanistic insight into disorders and related phenotypes, stratification for clinical trials, and selection of individualized treatments targeted at underlying pathophysiology and developmental disease trajectories (Gandal et al., 2016). As PRS estimates improve with greater sample size and inclusion of much greater ancestral diversity, they will allow for early identification of vulnerability traits in an individual (up until recently, prediction of a disease or trait could only be at the level of family risk). For the first time, the possibility of implementing very early intervention (before disease onset) is within reach—and critically, not only for broad categories of disease but also for an individual’s probability of developing specific traits leading to impairment. Yet it is critically important to note the following: 1) PRS must be employed only in the population upon which the discovery dataset is based or else there can be inaccurate estimates of risk in individuals; and 2) currently, PRS are more predictive in European populations than in all other ancestries, due to the overrepresentation of Europeans in GWAS (Martin et al. 2019). Given the clinical utility of PRS for a variety of biomedical conditions, this lack of diversity in genetic studies could lead to an even greater accentuation of
health disparities between individuals of European descent and underserved non-European populations (Martin et al. 2019). Therefore, it is of the utmost importance that concerted efforts be made to diversify genetic studies.

1.2 Approaches to Refining the Characterization of Causality
In this dissertation, we have employed a variety of approaches in an attempt to refine the characterization of causal risk factors for the development of psychopathology, seeking to integrate our understanding of some of the aforementioned challenges.

1.2.1 Nurture
In Chapter 2, we examine what the field currently considers to be some of the greatest non-genetic risk factors for psychopathology, some of which prior genetically-informative studies have demonstrated to incur risk for the development of psychiatric conditions, above and beyond genetic risk: child maltreatment and maternal depression in an impoverished sample. This sample was enriched for what are presumed to be toxic stressors, and yet surprisingly, we found no association of the measures with one another, nor did they predict child psychopathology. Even the children who experienced maltreatment did not demonstrate elevated rates of emotional and behavioral problems or impairment in adaptive functioning, assuming the measurements were sensitive enough and that caregiver report in this sample is valid. This study demonstrated the challenges of capturing risk factors and outcomes in this highly complex milieu comprised of genetic and environmental influences, and demonstrated the criticality of employing genetic information to more fully understand risk, resilience, and outcomes. My role in this study included data analysis and interpretation, consultation with co-authors on data analyses and
interpretation, literature review, and writing the entire manuscript (incorporating edits suggested by co-authors).

### 1.2.2 Time
Given this complexity and the obstacles to parsing environmental elements of the etiology of psychopathology in a sample which theoretically was enriched for risk factors, I turned the focus of my efforts to a more scientifically tractable, highly heritable neuropsychiatric condition: ASD. As outlined above, given that clinical ASD is merely the pathological tail of a continuous distribution of autism-related variation in reciprocal social behavior (AVR), employing a quantitative measure of AVR was critical. If the long-term goal of being able to predict AVR risk in infancy by using genomic data was to be realized, AVR stability throughout the life course needed to first be established. Yet no study had examined the longitudinal course of AVR across the entire range of variation. Therefore, in Chapter 3, we analyzed the longitudinal stability of a quantitative measure of AVR, the Social Responsiveness Scale (SRS). The SRS exhibited trait-like stability from preschool to young adulthood across the entire range of variation in which it manifests in childhood. My role in this study included data analysis and interpretation, consultation with co-authors on data analyses and interpretation, literature review, and writing the entire manuscript (incorporating edits suggested by co-authors).

### 1.2.3 Nature
Having established the enduring nature of AVR, the utility of incorporating it into molecular genetic analyses for PRS could begin to be interrogated. As outlined above, given the substantial common variant burden contributing to ASD and the disorder’s quantitative nature, in Chapter 4 we employed genome-wide association data to generate polygenic risk scores (PRS) of AVR. Specifically, we examined the contribution of ASD-PRS to AVR in a multiplex sample (families
in whom there is more than one child with ASD), given the elevation of quantitative AVR among unaffected members in multiplex families. ASD-PRS significantly explained variance in AVR, critically in both affected and unaffected subjects. The unique nature of this study’s multiplex familial sample allowed it to yield a novel demonstration that common polygenic variation in ASD contributes to familial, subclinical QAT. My role in this study included study design, decision-making about data analyses and interpretation, consultation with co-authors on data analyses and interpretation, literature review, and writing the entire manuscript (incorporating edits suggested by co-authors).

1.3 Conclusions
In summary, disentangling causal factors in the development of psychopathology is challenging, yet critically important for both prevention and intervention efforts. In this dissertation I have utilized three different approaches—environmental, longitudinal, and genetic—in an attempt to refine the field’s understanding of measurable causal influences on neuropsychiatric conditions. I began in Chapter 2 by uncovering that putative risk factors for psychopathology were not correlated with one another in our sample, nor did they significantly predict child psychopathology. These findings raised questions about current measurements of presumed toxic stressors, as well as the very nature of what constitutes toxic stress. In Chapter 3, we then turned our efforts to a disorder that is nearly entirely genetic in origin, autism, with the hopes that its genetic nature would allow a more fruitful identification of disease risk factors. The ultimate goal of early prediction of autism risk necessitated that we first establish the long-term stability of the SRS, the only comprehensive quantitative measure of AVR. We found remarkable stability in AVR in every sub-domain of the measure. Having established that measurement of early subclinical variation in AVR remains stable through early adulthood, in Chapter 4, we then
utilized ASD-PRS to predict variation in this phenotype. Although the variance explained was modest, it was in line with other PRS studies of QAT and will be improved upon with larger discovery GWAS. More importantly, this study served as a critical proof of concept that ASD-PRS predicted phenotypic variation in both unaffected and affected individuals. Future research by our group and others will examine if conducting a GWAS on the quantitative phenotype itself will provide greater SNP-heritability estimates than GWAS based on categorical diagnosis, and ultimately greater predictive power for ASD risk.
1.4 References


Cross-Disorder Group of the Psychiatric Genomics Consortium. (2013). Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nature Genetics, 45(9), 984–994. https://doi.org/10.1038/ng.2711


Kendler, Kenneth S. (2013). What psychiatric genetics has taught us about the nature of psychiatric illness and what is left to learn. *Molecular Psychiatry, 18*(10), 1058–1066. https://doi.org/10.1038/mp.2013.50


National Academy of Sciences, 114(44), 11727–11732. https://doi.org/10.1073/pnas.1707178114


Chapter 2: Characterizing Toxic Stress in Early Childhood: Maternal Depression, Maltreatment, and HPA-Axis Variation in a Pilot Intervention Study

Rachael E. Wagner, Melissa Jonson-Reid, Brett Drake, Patricia Kohl, Laura Pons, Yi Zhang, Robert Fitzgerald, Mark L. Laudenslager, and John N. Constantino
2.1 Abstract
Adverse experiences superseding a child’s capacity to sustain regulation of emotion and adaptive function are theorized to constitute “toxic stressors” when they induce a deleterious biological response within an individual. We ascertained presumptive parameters of toxic stress, including hair cortisol concentration (HCC), a promising biomarker of hypothalamic-pituitary-adrenal (HPA) activation, among 164 impoverished infants and toddlers (ages 4-48 months) from 132 families enrolled in Early Head Start (EHS). We randomized a subset of these families into a pilot intervention arm of parenting education (the Incredible Years, TIY), which supplemented the EHS curriculum. Official report child abuse and neglect (CAN) and child behavior were serially ascertained over the course of the study. We observed relatively low associations among maternal depression, CAN, caregiver-child relationship quality, HCC, and adverse child behavioral outcomes. Despite the poverty and high prevalence (51%) of CAN in the sample, the frequency of internalizing and externalizing disorders among the children did not exceed that of the general population. The pilot supplementation of EHS with TIY did not significantly reduce adverse behavioral outcomes or CAN. This study revealed marked independence of standard indices of toxic stress (child maltreatment, maternal depression, caregiver emotional unavailability) which are presumed to be severe risk factors for the development of psychopathology. Not only were these variables weakly inter-correlated, but furthermore, they were only minimally predictive of child behavioral outcomes in this sample. These findings caution against presumptions about the toxicity of individual stressors, highlight the importance of ascertaining risk (and compensatory influences) comprehensively, and demonstrate the need for greater understanding of what comprises parameters of resilience in early childhood.
2.2 Introduction
The National Scientific Council on the Developing Child (NSCDC) developed the construct of toxic stress, defining it as “strong, frequent, or prolonged activation of the body’s stress response system” (National Scientific Council on the Developing Child, n.d.), typically occurring in the absence of the buffering protection of a supportive caregiver (Shonkoff 2016). Examples of stressors that may induce a toxic stress response include physical, sexual, or emotional abuse; chronic neglect; caregiver substance abuse or mental illness; exposure to violence; and/or the accumulated burdens of family economic hardship. It is theorized that these stressors are much more likely to evoke a pathological response of the stress response system if the child lacks buffering protection from adult caregivers, such as often occurs with maternal depression (Essex, Klein, Cho, & Kalin 2002; Lupien, King, Meaney, & McEwen 2000; Lupien, King, Meaney, & McEwen 2001). An essential characteristic of this phenomenon is a deleterious biological response to a stressor during sensitive developmental periods. According to the NSCDC, potential biological consequences include the disruption of developing brain structure and function; changes in gene expression; and a lowered threshold for stress system activation, which can lead to greater risk of physical diseases, psychopathology, and cognitive impairment—not only during childhood and adolescence, but also into adulthood (Jonson-Reid et al. 2010; Jonson-Reid, Kohl, & Drake 2012; Lanier, Jonson-Reid, Stahlschmidt, Drake, & Constantino 2010).

The last two decades of research on the deleterious effects of adverse early life experiences have generally supported a stress-diathesis conceptualization, in which a) individuals vary in their capacity to mount an adaptive response to a given stressor; and b) higher-acuity stressors—at the extreme of typical environmental variation—are much more likely than lower-acuity stressors to adversely affect development (Shonkoff et al. 2012;
Constantino 2018). However, it is difficult to know whether a given environmental stressor is exceeding a young child’s capacity to manage it safely. Overt behavioral indices of emotion dysregulation may represent useful markers, especially before the time when most children are able to verbalize their subjective experience. Yet their absence is no guarantee of freedom from later-stage adverse consequences of accumulated adversity. For this reason, biological markers—particularly those which can be assayed early in life—have been invoked as key indices of a pathological response to stress.

One such biomarker that has received a substantial amount of investigation is the stress hormone, cortisol. Levels of cortisol in blood, urine, and saliva have historically been used to index activation of the hypothalamic-pituitary-adrenal (HPA axis in response to stress (Hellhammer, Wüst, & Kudielka 2008; McCallister, Smith, & Elwood 2004; Papadimitriou & Priftis 2009)). However, cross-sectional measurements using these biomarkers have been inconsistent in their respective associations with carefully measured environmental stressors and concomitant profiles of behavior (Adam & Kumari 2009; Gunnar, Talge, & Herrera 2009). One possible reason for these inconsistencies relates to the dynamic nature of cortisol release and the possibility that cross-sectional assays at single points in time do not adequately reflect chronic perturbations of the HPA axis (Adam & Kumari 2009; Segerstrom, Boggero, Smith, & Sephton 2014; Trickett, Noll, Susman, Shenk, & Putnam 2010). Prior studies in non-human primates have suggested that hair cortisol concentration (HCC) levels might reliably assess HPA activity over an extended period and provide greater predictive power in determining whether environmental challenges are chronically exceeding an individual’s adaptive capacity (Fairbanks, Jorgensen, Bailey, Breidenthal, Grzywa, & Laudenslager 2011).
One environmental risk factor for the development of psychopathology—to date, the most influential and modifiable—is child maltreatment. Child maltreatment is particularly prevalent in low-income populations, yet its proper ascertainment is challenging. Retrospective measures have the potential to be affected by recall and response bias (Widom, Raphael, & DuMont 2004), but they capture a greater number of individuals affected by maltreatment than prospective measures (Sedlak et al. 2010). In contrast, prospective measures, such as official report, are regarded as having more validity (i.e., specificity), and may tend to capture the more severe cases (Baldwin, Reuben, Newbury, & Danese 2019). In this study, we capitalized upon a unique opportunity to collect official report data (i.e., a prospective measure) from the States of Missouri and Illinois to ensure valid measurement of child maltreatment.

Critically, there is evidence that the effects of even the most severe psychosocial stressors may be mitigated by means of interventions that provide supportive caregiving, nurturing, and engagement (National Scientific Council on the Developing Child n.d.; Shonkoff 2016). Early Head Start (EHS) is a federal program which provides comprehensive support services to families of infants and young children who are at or below the national poverty line, a risk factor associated with a number of presumed toxic stressors. This study involved families enrolled in EHS, a subset of which received enhanced parenting education using the Incredible Years (TIY) curriculum. Substantial work has been done to establish the efficacy and effectiveness of TIY, an intervention which seeks to decrease harsh, critical parenting and cultivate supportive and responsive caregiving, elements essential for children’s emotional, social and behavioral development (Barlow 2007; Foster, Olchowski, & Webster-Stratton, 2007; Gardner, Burton, & Klimes, 2006). There is also evidence supporting the effectiveness of TIY for high-risk populations, such as among families characterized by child welfare (Linares, Montalto, Li, &
Oza 2006) and in particular EHS families (Gross et al. 2003; Hurlburt, Nguyen, Reid, Webster-Stratton, & Zhang 2013).

We undertook this multi-stage study with the following goals: 1) to validate HCC as a biomarker of HPA-axis variation; 2) to understand the relationships of presumptive toxic stressors with one another, as well as child behavioral outcomes; and 3) to pilot the supplementation of a parental education intervention within an EHS curriculum. We hypothesized that theorized risk factors for adverse child outcomes, ones which have been presumed to elicit a toxic stress response—poverty, maternal depression, deficits in emotional availability of parents, and CAN—would be associated with one another, HPA axis abnormality, and behavioral deviation in early childhood. Further, we predicted that their effects might be buffered by intensive levels of parent training via TIY.
2.3 Results

2.3.1 Baseline Analysis
Table 1 presents descriptive statistics of key characteristics of the sample at baseline. Of note, only 33.6% (N=45) of the mothers who completed a CES-D (N=134) met the clinical cut-off for depression (as determined by a score of ≥16, per standard guidelines). CAN data were accessible for 79.7% of the 136 families in our sample: of these families for whom we obtained CAN data, 50.8% of the 120 children lived in families with at least one Child Protective Services report of CAN.

Table 2 presents a correlation matrix of selected quantitative risk and outcome indices employed in the study, none of which were statistically significant, except for the expected intercorrelation of internalizing and externalizing scales of the CBCL. Of note, neither hair nor salivary cortisol was significantly correlated with any of the other risk measures (nor with each other). Since child maltreatment has been demonstrated to incur additional risk for psychopathology, we analyzed presumptive measures of toxic stress as a function of official-report CAN (see Table 3 for results of independent samples t-tests). No significant differences existed between groups (official-report maltreatment versus no maltreatment). As summarized in Table 4, linear regression analyses failed to identify main effect predictors of child behavioral outcome as measured by the CBCL.
2.3.2 Effects of Pilot TIY Intervention

Socialization Group Attendance

Figure 2 presents attendance in three time blocks: 1) the 12 scheduled socialization groups prior to enrollment in BTS (“Pre”); 2) the pilot study interval during which 12 sessions of the TIY intervention were administered, along with continued socialization groups-as-usual for the controls (“During”); and 3) the 12 socialization groups following the randomized intervention (“Post”). The reported percentages reflect average attendance rate for the enrolled subjects. During TIY intervention interval, attendance at socialization groups was approximately three times higher in TIY groups than in the control treatment as usual (TAU) groups.

Child Abuse and Neglect

Among those without prior CAN reports, TIY group had 23.9% at follow-up compared to 27.9% of TAU; this difference failed to reach statistical significance, but exhibited a trend in the expected direction. It is important to note that among those with prior CAN reports, TIY group had 39.9% later reports compared to 60% of controls. However, there were only 5 controls, rendering our analysis without power to detect an effect. There was no greater likelihood of missing official report CAN data (not ever matched) between TIY and TAU groups. However, when limiting to children for whom we obtained a state match, all of the cases that had family level prior reports only (not the index child) were in TIY group, suggesting that the intervention group may have been at higher risk at the time of enrollment.

Internalizing and Externalizing Behavior

We conducted repeated measures ANOVAs to ascertain if there was a change in CBCL scores between time points across TIY and TAU. There was no significant difference in CBCL score change between the groups. Pre- and post- descriptive statistics are outlined in Table 5.
2.4 Discussion
In this study involving measurement of key indices of putative toxic stressors among young impoverished children participating in EHS, we observed a lower-than-expected frequency of maternal-report childhood behavioral abnormalities, and high rates of official-report CAN. We observed strikingly low associations among maternal depression, CAN, caregiver-child relationship quality, HCC, and adverse child behavioral outcomes. That the proportion of families with official reports of CAN was exceptionally high, and that CAN was not associated with variation in maternal reports of child behavioral outcome was surprising, as CAN has been found to exert deleterious causal influences on child development (R2~20%) (see Constantino, 2018; Jonson-Reid, Presnall, Drake, & Fox, 2010). In this study Tiy did not significantly augment effects of EHS in mitigating child psychopathology or CAN. Although there were trends for improvement in the expected direction, attrition rendered this analysis without sufficient power to derive a definitive determination of the intervention’s effectiveness.

The observation of negligible associations between putative “toxic stressors” (child maltreatment, poverty, maternal depression), HCC, and behavioral outcomes in this sample of young children in poverty was unexpected. These findings could be accounted for by a number of factors: 1) the maternal-report measures of child psychopathology may not have been sensitive or reliable enough to faithfully reflect early manifestations of psychopathology; 2) the effects of these stressors on enduring psychopathology may manifest later in development; 3) the role of these presumed stressors in child development may not prove toxic in the context of supportive intervention paradigms such as EHS; and 4) inferences about the associations of these variables from studies of lower risk samples may not hold true under conditions of extreme high risk. Moreover, it is clear from these data that any one major risk factor (e.g., maternal
depression) is not adequate to serve as a proxy for others, such as the direct measurement of parental emotional responsiveness or the occurrence of CAN.

Limitations of the study include the following: 1) attrition over the course of the intervention, resulting in constraints on statistical power, and without sufficient sample size to test interactions; 2) a low number of subjects in the clinical range of child psychopathology, thus reducing ability to ascertain associations between child psychopathology and risk and protective factors with small effect sizes; 3) absence of an epidemiologic sampling frame; 4) variable time periods families were enrolled in EHS prior to the start of our intervention, which may have attenuated relationships among risk variables and outcomes. We note that child behavioral abnormalities were ascertained by parent-report and therefore could have been systematically underestimated.

In this study, variation in HCC was not associated with the putative toxic stressors of maternal depression or official report CAN, nor did HCC predict adverse behavioral outcomes. Recent genetically-informative studies have revealed that HCC is substantially heritable, around 72% (in line with our highly stable ICC estimates of HCC), and more reliably index the effects of genotype than within-family environment—importantly, including socioeconomic status (Rietschel et al. 2016; Tucker-Drob et al. 2017). In a study of older twins, age 12-21 years, no significant phenotypic or genetic correlation existed between HCC and perceived stress, depressive symptoms, and neuroticism (Rietschel et al. 2017). Further, despite the measurement reliability of HCC, its usefulness in the present study may have been mitigated by the sample size and its complex relationships to other factors. It should also be noted that cortisol is only a part of a complex stress response.
Testing theories about the biological effects of stress on development and psychopathological outcomes—particularly in humans—is challenging. These data indicate that caution should be exercised in drawing inferences about what constitutes “toxic stress,” especially when inferred from studies that are not designed to critically test causal associations and their potential confounds, one of which is unmeasured gene-environment correlation. Here, for example, a marker originally presumed to relate to toxic stress (HCC), and subsequently traced to genetic influences, proved to be minimally associated with major risks and outcomes related to overwhelming stress in early childhood. Moreover, the most notable finding of this study was that in a sample enriched for putative toxic stress factors (e.g., poverty, child maltreatment, maternal depression), the risk factors were only weakly correlated with one another, and the children were characterized by lower-than-expected levels of internalizing and externalizing psychopathology by their parents’ report. These findings suggest that our current understanding of “toxic stress” is under-developed, and that some presumed risk factors may not be as deleterious as currently hypothesized in the field—at least not within the population of young children such as the one studied here. This research reinforces the need in future research to control for genetic variation (and its effects) in order to operationalize what constitutes “toxic stress.”

2.5 Methods
The study design was divided into two stages, as depicted in Figure 1: Stage 1) validation of HCC as an index of HPA-axis variation, and Stage 2) a pilot intervention study of risk and outcome among impoverished children in EHS. Figure 1 depicts the study design and subject flow from enrollment to analysis.
2.5.1 Stage 1: Validation of Hair Cortisol

Subjects
We recruited 27 subjects (9 female) in the greater St. Louis, Missouri area from local early childhood centers and collaboration with a local research study, Early Childhood Connections (ECC; R34MH083871). This sample of children had a mean age of 34.37 months (SD 12.15, range 10-58 months); and racial distribution was 40.7% African American, 33.3% Caucasian, and 25.9% other.

Procedure
We collected samples of overnight urine, hair, and saliva acquired from the subjects precisely 30 minutes after awakening (see below) and at 4pm on a single day throughout which we measured their heart rate variability. This procedure was repeated at three-month follow-up (T2) for 25 of the 27 original subjects.

Measures
Hair Cortisol Concentrations. Per established published protocols (Hoffman et al. 2017), approximately 50 strands of hair were snipped at the base of the hair shaft from the posterior vertex during the baseline assessment and post-treatment. The first three centimeters of hair length from the base were assayed for HCC using a commercially available enzyme immunoassay kit (Salimetrics, LLC) following a previously published protocol (Arch et al. 2014; D’Anna-Hernandez, Ross, Natvig, & Laudenslager 2011; Hoffman, D’Anna-Hernandez, Benitez, Ross, & Laudenslager 2017). Cortisol concentration analysis and assay validation were conducted as previously described (Hoffman, D’Anna-Hernandez, Benitez, Ross, & Laudenslager 2017). Briefly, each hair sample was placed in a pre-weighed 2 ml cryovial (Wheaton, Millville, NJ, USA), washed three times in 100% isopropanol and dried. After washing, drying, and re-weighing samples on a high-sensitivity electronic balance (Mettler Toledo Model MS105, Greifense, Switzerland) to determine individual hair mass in these small...
samples, hair was ground in the same cryovial using a ball mill (Retsch, Haan, Germany) after adding a 4.76 mm carefully cleaned stainless steel ball bearing. Specially milled aluminum cassettes were designed to hold three cryovials. The cassettes, containing the cryovials, were submerged in liquid nitrogen for 3 to 6 minutes to freeze hair samples to facilitate grinding. Samples subsequently were ground for 4 to 5 minutes. Powdered hair was extracted in the same cryovial in 0.33-1.0 ml (depending on sample mass at a ratio of 5 pg hair/100 microliters (µl) methanol) high pressure liquid chromatography (HPLC) grade methanol for 24 hours at room temperature on a side-to-side shaker platform. Following methanol extraction, cryovials were spun for three minutes in a centrifuge at 15700g to pellet the hair. Then 133 µl of the extraction supernatant was removed, placed into a microcentrifuge tube, and dried under a stream of nitrogen in a drying rack in a fume hood at room temperature. The dried extracts were then reconstituted with assay diluent based on hair weight.

Cortisol levels were determined using a commercial high sensitivity Enzyme Immunoassay (EIA) kit (Salimetrics LLC, State College, PA, USA) per manufacturer’s protocol. Methods for assay cross validation with other laboratories using liquid chromatograph-mass spectrometry (LC/-MS/MS) and cross reactivity were described previously by (Russell et al. 2015). Cross validation entailed assaying 10 identical samples by four laboratories by EIA and/or LC/MS and comparing the resulting levels. Correlations across laboratories of r^2>0.9 were noted for both EIA and LC/MS indicating excellent consistency and comparability across several laboratories. Intra- and inter-assay coefficients of variation were less than 10 and 5% respectively for all assays using this commercial assay, regardless of cortisol matrix.

**Salivary Cortisol Concentrations (SCC).** Saliva samples were assayed for cortisol concentration using a commercially available enzyme immunoassay kit (Salimetrics, LLC) following a
previously published protocol (Arch et al. 2014). For morning (AM) saliva samples, parents were instructed to call the research team immediately upon the child’s awakening; saliva samples were collected, on average, 32 minutes (SD=5.7 minutes) after the child awoke. Afternoon (PM) saliva samples were obtained as close to 4:00 pm as possible and assayed in the same manner.

**Urine Cortisol.** Urine samples were obtained from an overnight collection preceding the morning in which AM saliva samples were obtained. Urine samples were diluted in assay buffer and assayed for cortisol concentration using a commercially available enzyme immunoassay kit (Salimetrics, LLC) following the protocol described above.

**Heart Rate Variability.** Heart rate recordings were obtained using a Polar rs400 heart rate monitor applied to the child’s chest on the day of each biomaterials collection. Resting heart rate variability was calculated over intervals in which the child’s heart rate was less than 140 beats per minute.

**Statistical Analyses**
HCC was log transformed to approximate a normal distribution. HCC outliers were removed using standard interquartile range (IQR) formulations (1.5x the IQR) after the log transformation. The skewness and kurtosis of the non-log-transformed data were 7.81 and 68.8, respectively; after transformation, skewness was .84 and kurtosis was .626. Pearson’s correlation coefficients were computed for bivariate associations among the variables, and intra-class correlation (ICC) coefficients were computed to estimate each measure’s temporal stability.

**Results of Initial HCC Validation**
HCC exhibited marked temporal stability (ICC 0.70, \( p=2.52e-05 \)), greater than all other HPA-axis biomarkers measured, as detailed in Table 6. HCC was modestly correlated with PM SCC at baseline \( (r=.47) \).
2.5.1 Stage 2: Pilot Intervention of The Incredible Years

Feasibility Study
In order to ensure the feasibility of implementing TIY within an EHS curriculum, we enrolled 13 families from the Youth in Need EHS program, comprising 19 children aged 12-48 months. We supplemented their families’ biweekly group visits (see below) with TIY curriculum for a six-month period. After completion of TIY, families returned to the usual schedule of socialization visits. Curriculum was administered by two trained and certified TIY group leaders under the employment of Washington University School of Medicine (WUSM). These individuals completed all measures to ensure fidelity to TIY model.

Pilot Study

Subjects
Having established the feasibility of our study design, we recruited 145 subjects from 125 families enrolled in Early Head Start programs in Eastern Missouri (Youth in Need) and Madison County, Illinois (Riverbend), over calendar years 2015-2017. Both the Youth in Need and Riverbend sites participated in EHS using the Home-Based program option, in which the Parents as Teachers (PAT) home visitation model and Creative Curriculum Teaching Strategies served as the core curricula in use by family educators during weekly home visits. Consistent with the EHS Program Standards for the Home-Based program option, biweekly group meetings were a core component of the EHS intervention (see below). At the time of their respective enrollments, the children’s ages ranged from 5-48 months. Primary caregiver ages ranged from 17-45 years; in our sample, all but four primary caregivers were mothers; two primary caregivers were fathers; one was an aunt; and one was a grandmother.

Randomization
In this pilot intervention study, EHS families were randomized 2:1 to supplementation with the Incredible Years (TIY) curriculum according to EHS family educator, following determination
of study eligibility and the obtaining of consent. Families were informed that they would be assigned to a supplemental education track (TIY) or control condition (treatment as usual, TAU) on the basis of whether their assigned EHS family educator was randomized to one or the other condition. During the study years, the family educators carried caseloads of 2-5 EHS families as well as 5-10 families enrolled in Head Start (HS), to ensure continuity of support when children age out of EHS at age 3. Randomization was conducted at the level of family educator in order to minimize “contagion effects” (i.e., knowledge gained by family educators through co-participation with their clients in TIY (see below) being transferred to families assigned to the control condition over the 6 months of biweekly intervention). This arrangement also afforded alignment of the intervention trial with the programmatically-embedded continuity of relationships between family educators and their respective clients as EHS children cross the threshold of eligibility for HS services at age 3.

For the intervention group, TIY was implemented in the context of biweekly group meetings. Each month, as a core element of the EHS curriculum itself, all families are encouraged to participate in one socialization group meeting and one education group meeting on an alternating biweekly schedule. Socialization groups serve to expose both parents and children to developmentally appropriate learning opportunities in a group setting, and to provide social opportunities for parents who are often isolated in their role of caring for young children. For families randomized to TIY, the usual curriculum for biweekly EHS group meetings was substituted with TIY curriculum over a period of 6 consecutive months (12 manualized meetings of TIY), with babysitter services provided to afford parents opportunity to fully engage and participate in the intervention according to its manualized protocol. After completion of TIY, families resumed the usual schedule of EHS group meetings. In both the intervention and control
groups families continued to receive weekly home visits from their family educators before, during, and after the intervention period.

**The Incredible Years (TIY) Curriculum**

TIY differs from the curricular content of EHS group meetings in that it is a structured, evidence-based program of parenting education in which video vignettes, role play scenarios and active discussion of foundations of developmentally-appropriate parent-child interaction are discussed with trained group leaders (who were certified by the developers of TIY) and among participating parents. The curriculum has been previously described in an extensive body of previously published research (see above). In this randomized controlled trial embedded in two EHS populations, TIY was delivered in five waves, each of 6 months duration, by two group leaders.

In summary, following enrollment of age-eligible EHS children residing within a given service area, clusters of enrolled children sharing a given family educator were identified, sets of three comparably-sized clusters were defined (each assigned to their own family educator) and two of the three clusters (categorized by family educator) were selected at random for assignment to the intervention condition. The intervention group adopted a separate location for group visits on the same schedule as their control counterparts, and the standard EHS group meeting content (for alternating socialization and education group visits) was replaced with TIY for 6 consecutive months.

**Outcomes Monitoring**

Data were collected at enrollment and at the end of TIY intervention period (6 months following baseline). Subjects were evaluated on pencil-and-paper measures completed by the child’s primary caregiver, clinician observation, HCC, SCC, and cross-referencing with State administrative records for official report child maltreatment.
**Measures**

**Hair Cortisol Concentrations (HCC).** HCC was assayed in the same manner as outlined above in Stage 1 of the study.

**Salivary Cortisol Concentrations (SCC).** SCC was assayed in the same manner as outlined above in Stage 1 of the study.

**Official Report Child Abuse and Neglect (CAN).** As a condition of participation in this research program, families consented to cross-referencing their identifying information and all research data with official state administrative records, including reports of CAN during and after the pilot intervention. All of the research information except the identifiers was encrypted. The encrypted information was submitted with identifiers to the Department of Social Services (Missouri) and the Department of Health Services (Illinois), where it was cross-referenced with official report records at an individual level. The linked dataset was subsequently stripped of individual identifiers before being returned to the research team. Official reports included placement in foster care (FC), substantiated reports of maltreatment (SRM), and unsubstantiated reports of maltreatment (URM). It should be emphasized that according to the harm/evidence model of substantiation, an SRM is not equivalent to verifying the presence of maltreatment, but rather a label used when sufficient evidence and/or risk of harm exists to permit family court intervention if needed (Drake & Jonson-Reid 2000). In Missouri, both SRM and URM have been eligible to receive in-home or FC intervention for many years. It was not possible to specify with confidence which child or how many children in a family were subjected to maltreatment; nor could we ascertain the age at which maltreatment first occurred, since a) date of first available report is not the same as the time when maltreatment began; and b) some of the reports
represented recurrences of unsubstantiated maltreatment have been purged. However, we note that abuse and neglect likely affect all children in a family.

**Center for Epidemiological Studies Depression Scale (CES-D).** The CES-D is a freely-available, highly utilized brief depression screener (20 items), with established reliability and validity across a wide variety of health conditions and demographics, both nationally and internationally (Radloff 1977). The symptomology assessed by the CES-D include the following sub-domains: depressed mood, feelings of guilt and worthlessness, feelings of helplessness and hopelessness, psychomotor retardation, loss of appetite, and sleep disturbance.

**Caregiver-Child Social Emotional Rating Scale (CCSERRS).** The CCSERS is designed to assess the quality of caregiver–child socioemotional interactions and relationships (McCall, Groark, & Fish, 2010). Characterized by established validity and reliability, this clinician-observation scale examines the caregiver’s emotional availability, affect of child and caregiver, the mutual nature of caregiver-child interactions, and the caregiver’s responsivity to the child’s lead. Parents and their children were directly observed in free play at baseline and follow-up by trained clinicians who completed the 15-minute assessment at each time point.

**Child Behavior Checklist (CBCL).** The CBCL (i.e., Achenbach Scales of Empirically Based Assessment, Achenbach & Rescorla 2000) is parent- and/or teacher-report instrument assessing emotional and behavioral problems during preschool and childhood. The widely-used CBCL has undergone extensive establishment of its validity, reliability, and psychometric properties (Verhulst & Ende 1995; Verhulst & VanderEnde 1992). The various syndrome scores of the CBCL can be combined to form Internalizing and Externalizing composites, which we utilized in the current study as outcome measures of child psychopathology.
Data Analysis

During preliminary analyses, we log transformed HCC to approximate a normal distribution. We then conducted two sets of primary analyses in order to examine two lines of questioning: 1) baseline data analysis and 2) baseline-to-follow-up analysis. In our baseline analyses, we tested associations among quantitative measures in a correlation matrix, conducted t-tests to compare measures as a function of the presence or absence of CAN, and performed linear regressions to determine if theoretically important variables—maternal depression and CAN—predicted our dependent variables: internalizing and externalizing child psychopathology. We utilized a repeated-measures ANOVA to determine the presence of an intervention effect by ascertaining if scores on the pre- and post-measures changed significantly between baseline and follow-up, for both TIY and TAU groups. The data analysis for this paper was generated using SAS/STAT, Version 9.4 of the SAS System for Microsoft Windows Copyright © 2002-2012. SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

2.6 Acknowledgements
This study was funded by the U.S. Department of Health and Human Services, Administration for Children and Families, as one of a set of six grants constituting the Buffering Toxic Stress Consortium (#90YR0054). The authors gratefully acknowledge the collaboration of Pat Holterman-Hommes, Chief Executive Officer of Youth in Need, and the administrative and family education teams of Youth in Need and Riverbend Head Start. We wish to recognize the dedicated effort of Sandra Polanc, who led TIY groups with co-author LP and whose devotion to this research program and the families was an inspiration to all. Finally we express the very deepest gratitude to the parents who graciously participated in the study in the context of their
busy lives caring for young children under conditions of socioeconomic stressors that defined their children’s eligibility for enrollment in Early Head Start.

**Informed Consent**
All subjects provided informed consent.

**Ethics Approval**
Conducted at Washington University School of Medicine in St. Louis as a collaboration by investigators at the School of Medicine, Department of Psychiatry, and the Center for Violence and Injury Prevention at the George Warren Brown School of Social Work, the study was approved by the Human Research Protection Office at Washington University School of Medicine (WUSM). The authors certify that the study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments.
2.7 Figures

![Diagram showing the subject flow through the study's stages]

Figure 1: Subject flow. A depiction of the subject flow through the study's stages, from enrollment to analysis. Note: FU = follow-up.

- Stage 1: Cortisol Validation
  - Enrollment: N=27
  - Allocation: N=27
  - Follow-Up: N=17
  - Validation Analyses: N=27

- Stage 2: Pilot Study
  - Preliminary Feasibility Study: N=19
  - Randomized: N=145
    - Lost to FU: N=55
    - TAU: N=28
      - N=17
      - N=35
    - TIY: N=62
      - N=52

Baseline Associations: N=164
Figure 2: Parental attendance at EHS group meetings. This graph depicts parental attendance at EHS group meetings in relation to implementation trial interval. “During” = TIY Groups Substitute Socialization Visits.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of Subjects</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>123</td>
<td>--</td>
</tr>
<tr>
<td>Urban</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Missing Data</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>99</td>
<td>--</td>
</tr>
<tr>
<td>Black</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>Caregiver Level of Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No HS degree</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>HS degree or GED</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Some post-HS</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Caregiver Marital Status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>56</td>
<td>--</td>
</tr>
<tr>
<td>Married</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>Living with Partner</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Divorced or Separated</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Separated</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>82</td>
<td>--</td>
</tr>
<tr>
<td>Male</td>
<td>82</td>
<td></td>
</tr>
<tr>
<td>Caregiver CES-D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical&lt;sup&gt;a&lt;/sup&gt;</td>
<td>45</td>
<td>13.17 (9.60)</td>
</tr>
<tr>
<td>Normal</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>CBCL: Internalizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>83</td>
<td>46.63 (10.61)</td>
</tr>
<tr>
<td>Clinical&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>CBCL: Externalizing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal</td>
<td>85</td>
<td>47.74 (10.36)</td>
</tr>
<tr>
<td>Clinical&lt;sup&gt;b&lt;/sup&gt;</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CCSERRS</td>
<td>138</td>
<td>16.98 (4.76)</td>
</tr>
<tr>
<td>Hair Cortisol Concentration</td>
<td>135</td>
<td>1.64 (.67)&lt;sup&gt;c&lt;/sup&gt; pg/mg</td>
</tr>
<tr>
<td>Salivary Cortisol Concentration</td>
<td>146</td>
<td>0.15 (.24) ug/mg</td>
</tr>
<tr>
<td>CAN&lt;sup&gt;d&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No-CAN</td>
<td>59</td>
<td>--</td>
</tr>
<tr>
<td>CAN</td>
<td>61</td>
<td></td>
</tr>
</tbody>
</table>

Subject counts for demographic variables and all measures employed in the study, as well as mean and standard deviation (SD) for relevant variables. Unless otherwise indicated, variables reflect child-level data. <sup>a</sup> Clinical cutoff for CES-D utilized in this sample was ≥16. <sup>b</sup> Clinical cut-off for CBCL-Internalizing and Externalizing was ≥64. <sup>c</sup> Log transformed values. <sup>d</sup> CAN status is operationally defined as an official Child Protective Services report of maltreatment for any child in a family.
### Table 2: Correlation matrix of quantitative measures

<table>
<thead>
<tr>
<th></th>
<th>CCSERRS</th>
<th>CES-D</th>
<th>CBCL-I</th>
<th>CBCL-E</th>
<th>HCC</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCSERRS</td>
<td>Pearson’s r</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CES-D</td>
<td>0.071</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CBCL-I</td>
<td>0.272</td>
<td>0.226</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>CBCL-E</td>
<td>0.157</td>
<td>0.267</td>
<td>0.705</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>HCC</td>
<td>-0.06</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>SCC</td>
<td>0.032</td>
<td>0.007</td>
<td>0.099</td>
<td>-0.010</td>
<td>0.118</td>
<td>—</td>
</tr>
</tbody>
</table>

A matrix of Pearson two-tailed correlations between measures. *Note:* Bold values are statistically significant after Bonferroni correction for multiple comparisons.
Table 3: Comparison of subjects on measures as a function of CAN group status

<table>
<thead>
<tr>
<th>Measure</th>
<th>Number of Subjects (No-CAN/CAN)</th>
<th>Mean (SD)</th>
<th>t</th>
<th>Degrees of Freedom</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CES-D</td>
<td>58/61</td>
<td>11.89 (9.00) / 14.18 (10.07)</td>
<td>-1.30</td>
<td>117</td>
<td>0.19</td>
</tr>
<tr>
<td>CBCL: Internalizing</td>
<td>40/38</td>
<td>41.5 (20.35) / 40.39 (18.82)</td>
<td>0.25</td>
<td>76</td>
<td>0.80</td>
</tr>
<tr>
<td>CBCL: Externalizing</td>
<td>40/38</td>
<td>42.15 (20.48) / 40.89 (19.28)</td>
<td>0.28</td>
<td>76</td>
<td>0.78</td>
</tr>
<tr>
<td>CCSERRS</td>
<td>56/59</td>
<td>17.11 (4.74) / 17.54 (4.80)</td>
<td>-0.49</td>
<td>113</td>
<td>0.63</td>
</tr>
<tr>
<td>Hair Cortisol Concentration (HCC)</td>
<td>51/55</td>
<td>291.1 (1252.5) / 581.5 (2528.5)</td>
<td>-.76*</td>
<td>104</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Independent Samples t-tests as a function of group (No-CAN vs. CAN). No-CAN includes only subjects for whom there was a documented absence of a report. *Satterthwaite due to folded F-value.
Table 4: Linear regressions of maternal depression and CAN on psychopathology

<table>
<thead>
<tr>
<th>N</th>
<th>R²</th>
<th>Root MSE</th>
<th>Model F</th>
<th>Model p-value</th>
<th>Degrees of Freedom</th>
<th>Coefficient</th>
<th>β (SE)</th>
<th>t</th>
<th>Coefficient p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>0.039</td>
<td>19.8</td>
<td>1.30</td>
<td>0.278</td>
<td>2.68</td>
<td>CES-D</td>
<td>0.32 (0.25)</td>
<td>1.24</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAN</td>
<td>-6.73 (5.05)</td>
<td>-1.33</td>
<td>0.19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>R²</th>
<th>Root MSE</th>
<th>Model F</th>
<th>Model p-value</th>
<th>Degrees of Freedom</th>
<th>Coefficient</th>
<th>β (SE)</th>
<th>t</th>
<th>Coefficient p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>0.008</td>
<td>19.79</td>
<td>0.30</td>
<td>0.74</td>
<td>2.68</td>
<td>CES-D</td>
<td>0.14 (0.25)</td>
<td>0.57</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CAN</td>
<td>-3.30 (4.97)</td>
<td>-0.66</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Independent variables included the following: maternal depression (CES-D) and CAN, including prior CAN.
Table 5: Pre- and post- descriptive statistics of treatment and control groups

<table>
<thead>
<tr>
<th>Measure</th>
<th>The Incredible Years</th>
<th>Treatment at Usual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline N</td>
<td>Baseline Mean (SD)</td>
</tr>
<tr>
<td>CBCL: Internalizing</td>
<td>62</td>
<td>40.11 (20.34)</td>
</tr>
<tr>
<td>CBCL: Externalizing</td>
<td>62</td>
<td>40.56 (20.54)</td>
</tr>
</tbody>
</table>

This table presents the baseline and follow-up descriptive statistics on CBCL scores for TIY and TAU groups.
<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>HRV</th>
<th>Hair Cortisol</th>
<th>AM Salivary Cortisol</th>
<th>PM Salivary Cortisol</th>
<th>Time-Rated Change in Salivary Cortisol</th>
<th>Urine Cortisol</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td>( r )</td>
<td>0.10</td>
<td>-0.25</td>
<td>0.07</td>
<td>-0.04</td>
<td>-0.08</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.65</td>
<td>0.26</td>
<td>0.75</td>
<td>0.85</td>
<td>0.72</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td><strong>HRV</strong></td>
<td>( r )</td>
<td>0.48</td>
<td>-0.07</td>
<td>0.19</td>
<td>0.45</td>
<td>0.01</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.01</td>
<td>0.007</td>
<td>0.77</td>
<td>0.38</td>
<td>0.96</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>27</td>
<td>25</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td><strong>Hair Cortisol</strong></td>
<td>( r )</td>
<td>-0.12</td>
<td>0.70</td>
<td>-0.12</td>
<td>0.19</td>
<td>0.19</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.54</td>
<td>0.60</td>
<td>0.57</td>
<td>0.38</td>
<td>0.40</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>27</td>
<td>27</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td><strong>AM Salivary Cortisol</strong></td>
<td>( r )</td>
<td>-0.13</td>
<td>-0.01</td>
<td>0.09</td>
<td>-0.14</td>
<td>-0.92</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.52</td>
<td>0.97</td>
<td>0.33</td>
<td>0.53</td>
<td>&lt;.0001</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>27</td>
<td>27</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td><strong>PM Salivary Cortisol</strong></td>
<td>( r )</td>
<td>0.01</td>
<td>-0.10</td>
<td>0.08</td>
<td>0.38</td>
<td>0.52</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.98</td>
<td>0.61</td>
<td>0.70</td>
<td>0.03</td>
<td>0.01</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>23</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td><strong>Time-Rated Change in Salivary Cortisol</strong></td>
<td>( r )</td>
<td>0.12</td>
<td>-0.09</td>
<td>-0.85</td>
<td>0.45</td>
<td>-0.08</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.55</td>
<td>0.66</td>
<td>(&lt;.0001)</td>
<td>0.02</td>
<td>0.64</td>
<td>0.69</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>27</td>
<td>25</td>
<td>23</td>
</tr>
<tr>
<td><strong>Urine Cortisol</strong></td>
<td>( r )</td>
<td>-0.28</td>
<td>-0.03</td>
<td>-0.16</td>
<td>-0.24</td>
<td>-0.11</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td>( p )</td>
<td>0.16</td>
<td>0.89</td>
<td>0.93</td>
<td>0.58</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>( n )</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>26</td>
<td>24</td>
<td>24</td>
</tr>
</tbody>
</table>

*Note:* Pearson’s Coefficients of Correlation between the various measurements at baseline (light gray) and three-month follow-up (dark gray). Time 1-Time 2 intra class correlations (ICC) for each measure are presented along the diagonal. Time rated change in salivary cortisol is the change in salivary cortisol from morning to afternoon, divided by the time in hours. Statistically significant correlations (\( p<.0001\), adjusted for multiple comparisons using a Bonferroni correction) are in bold.
2.8 References


Chapter 3: Autism-Related Variation in Reciprocal Social Behavior: a Longitudinal Study

Rachael E. Wagner, Yi Zhang, Teddi Gray, Anna Abbacchi, Deporres Cormier, Alexandre Todorov, John N. Constantino

From:
3.1 Abstract
Deficits in reciprocal social behavior are a characterizing feature of autism spectrum disorder (ASD). Autism-related variation in reciprocal social behavior (AVR) in the general population is continuously-distributed and highly-heritable—a function of additive genetic influences that overlap substantially with those which engender clinical autistic syndromes. This is the first long-term prospective study of the stability of AVR from childhood through early adulthood, conducted via serial ratings using the Social Responsiveness Scale (SRS), in a cohort-sequential study involving children with ASD, other psychiatric conditions, and their siblings (N=602, ages=2.5-29). AVR exhibits marked stability throughout childhood in individuals with and without ASD.
3.2 Introduction

Autism spectrum disorder (ASD), characterized primarily by early-onset deficits in reciprocal social behavior, is one of the most severe, heritable, and enduring of all neuropsychiatric syndromes, with ASD diagnoses being lifelong (Constantino & Charman, 2016). It is now well established that autism-related variation in reciprocal social behavior (AVR), as measured by the Social Responsiveness Scale (SRS; Constantino & Gruber, 2012), is continuously distributed in the general population (Constantino & Todd, 2003; Constantino et al., 2012); highly heritable throughout the range observed from unaffected to sub-clinically affected to fully ASD-affected individuals (Constantino & Todd, 2000; 2003; 2005), and that the common genetic susceptibilities of subclinical autistic traits exhibit near-complete overlap with those of ASD itself (Constantino et al., 2006; Robinson et al., 2011). Further, when present, subclinical AVR (i.e., SRS scores which are elevated above population average, yet not severe enough to meet the arbitrary cut-off for clinical diagnosis) is associated with exacerbation in the severity of nearly any psychiatric condition with which it co-occurs (Constantino, 2017; Constantino & Frazier, 2013; Hawks et al., 2018; Lichtenstein, Carlström, Råstam, Gillberg, & Anckarsäter, 2010), which has clinical implications for enduring patterns of social adaptation in individuals with other diagnoses.

Given that deficits in this endophenotype modify the severity of an array of other forms of psychopathology, tracking the course of AVR across its entire range of variation over the course of early childhood to young adulthood is of critical importance. Therefore, we conducted a novel examination of the longitudinal course of AVR in both affected and unaffected individuals. To quantify AVR in clinical and general population samples, we utilized the SRS, a research standard for the characterization of inherited traits that comprise ASD. Since a) the
genetic causes of ASD substantially overlap with those underlying subclinical autistic traits; b) confirmatory factor analyses of SRS ratings exhibit measurement invariance across ages (Frazier et al., 2014); and c) the very high monozygotic twin correlation for SRS scores in unaffected individuals (~.80, Constantino & Todd, 2003) establishes relatively narrow constraints on the possible influence of error on an individual measurement, we hypothesized that AVR would exhibit stability not only in clinically-affected individuals, but also across its entire range of variation. If the longitudinal trajectories of AVR are stable throughout life, it is conceivable that measurements of AVR in early childhood could make strong predictions over the course of an individual’s life, particularly with respect to social interaction and adaptation in the context of any neuropsychiatric condition of childhood.

3.3 Results

3.3.1 Inter-rater Agreement

Very substantial agreement existed between maternal-report and teacher-report SRS scores, as demonstrated by the homology in mean scores. For all available pairings of maternal and teacher-report data (N=1,401 pairs), the correlation between maternal and teacher ratings was .71 (p<.01), depicted by the linear relationship in Figure 1 (which overwhelmingly reflects independent observations since serial ratings of any given subject in successive years were completed by different teachers). Given the strong maternal-teacher inter-rater reliability and higher consistency of maternal raters over time, all primary analyses were conducted on maternal data.

Inter-rater agreement was assessed starting from the 1,401 pairs of maternal and teacher ratings that were taken within one year of each other. The overall correlation, ignoring for now the fact that multiple such pairs can be found for one subject, is 0.70 (exact: 0.69612). We then
randomly sampled one pair per participant, yielding 496 independent mother-teacher pairs. The correlation in this random subset was 0.70 (exact: 0.69886). The age at which the assessments were taken did not matter. For example, the mother-teacher correlation is 0.73 when focusing on assessments taken after age 15 (the upper quartile of our sample) and 0.60 for those taken before age 7 (the lower quartile). For the RRB subdomain, the corresponding mother-teacher correlation is 0.67 overall, 0.68 for those over 15 and 0.67 for those under 7. For SCI, the values remain high, at 0.69 overall, 0.75 in children over 15 and 0.63 for children under 7. Mothers exhibited a pattern consistent with subtle rater contrast effects in comparison to teacher ratings (Constantino et al., 2010), rating their unaffected children somewhat lower than did teachers, and rating children with ASD somewhat higher than did teachers.

3.3.2 AVR at Baseline
Consistent with numerous previously published studies, the mean baseline maternal SRS score for each of the ASD-affected groups differed by approximately three standard deviations from that of unaffected children and by approximately two standard deviations from psychiatric controls. The maternal baseline SRS scores of individuals who met DSM–IV diagnostic criteria for Autistic Disorder averaged an additional 12.7 points (SD = 3.6; p < .01) above the scores of individuals who met criteria for DSM–IV Asperger Disorder or Pervasive Developmental Disorder Not Otherwise Specified. No significant differences existed in baseline scores between male ASD subjects from simplex families and multiplex families, or between male and female ASD subjects, regardless of familial category (p values between .57 and .92). Baseline scores of male and female unaffected siblings, as well as nonaffected psychiatric condition siblings, exhibited no significant differences (p > .75). Figure 2 depicts maternal SRS scores as a function of group.
Table 3 summarizes relative influences of selected subject characteristics on baseline maternal SRS scores. As expected, possessing a psychiatric diagnosis other than ASD elevated maternal SRS scores significantly \( (p < .01) \). In keeping with prior observations in multiplex family samples, unaffected siblings of individuals with ASD demonstrated mild aggregations of QAT, exhibiting a mean SRS score difference of 6.5 points higher than siblings of psychiatric controls; however, multiplex families comprised only a fraction of our total sample, in which this difference did not reach statistical significance. Age of child influenced maternal baseline SRS score \( (p = .02) \). Whether or not the child was capable of phrase speech also affected baseline SRS; specifically, nonverbal individuals \( (N = 24) \) were characterized by lower SRS scores.

Correlations between relevant domains of ADOS-2 and ADI-R and the SRS, matched by closest time point of evaluation, are provided in Supplemental Table 1. Since the ADI-R scores used were referable to level of functioning in the age range of 4-5 years, they do not reflect current symptom burden, and as such, reflect lower correlations with the SRS.

### 3.3.3 Stability of AVR over Time

When considering the entire sample, there was pronounced preservation of inter-individual variation in AVR over time. Table 4 summarizes test-retest correlations of successive maternal and teacher total SRS scores of all subjects from baseline through the final follow-up measurement (Supplemental Table 2 provides test-retest correlations of SRS subdomains). Test-retest correlations were extremely high for successive maternal ratings \( (~.90) \) and strong \( (~.70) \) for successive teacher ratings (which almost always involved different teachers for any given subject) (Table 4). None of the individuals with an ASD diagnosis experienced a magnitude of reduction in maternal SRS scores over the longitudinal period that would have been consummate with loss of diagnosis.
The scatterplot in Figure 3 depicts 527 randomly-selected pairs of successive maternal total SRS scores, one pair per subject, with linear regression (blue) and non-parametric loess (black) superimposed. For total SRS stability analysis (Figure 3), as well as subdomain analyses (Supplemental Figures 6 and 7), we utilized randomly-selected pair of SRSs per subject in order to prevent using repeated measures on the same subject and biasing the analysis. For all pairs, we used baseline as first measurement. The linear and non-parametric curves are at an extremely high level of agreement, and the correlation between ratings is .91. A linear regression with SRS at first assessment and time between measurements achieves $R^2=.85$, indicating that baseline SRS is an extraordinarily strong predictor of all future measurements. In this subset, average age at first measurement is 9.4 years (SD 4.8, range 1.6–28.7) and the average time between measurements is 2.0 years (SD 1.27, range 0.04–10.2). There is a slight suggestion that at the higher values, maternal ratings underestimate those of teachers.
measurements is 2.0 years (SD 1.27, range 0.04–10.2). There is a slight suggestion that at the higher values, maternal ratings underestimate those of teachers.

3.3.4 AVR over the Life Course

A spaghetti plot depicting the individual childhood trajectories of maternal-report total SRS scores for study subjects with more than one maternal SRS (N=527) is depicted as a function of ASD diagnostic status in Figure 4, revealing a marked distinction between affected and unaffected individuals throughout the course of childhood and adolescence (Supplemental Figures 8 and 9 provide SRS subdomain spaghetti plots).

To explore possible differences in longitudinal course for the seven groups, we developed growth curve models, specifying each group’s trajectory of maternal SRS rating as a function of age, utilizing loess smoothing with approximate 95% CI (Figure 5). There was no effect of control group origin (F_{2,247} = 0.53, p=.59), and only a mild suggestion of a trend for decreasing SRS scores with age (F_{1,396} = 2.86, p=.09); and no evidence of an age by control group interaction (F_{2,396} = 1.23, p=.29). The growth curves reveal strong preservation of inter-individual differences, by relatively consistent CIs over the full range of ages, except for later ages, in which sample size eroded statistical power to specify the precise trajectory. The data indicate a relative absence of differences among subgroups within the respective ASD-affected and ASD-unaffected populations. ASD-affected females exhibited trends for improvement. IQ scores failed to predict the longitudinal course of SRS scores within any group.

We fitted fixed-effects models as well as random effects models (intercept, slopes) to the data, with group, age, age-squared, and interaction terms between group and age and age-squared (Table 5). The model with random intercepts but fixed slopes (Model 2) fit the data the best, based on BIC (Table 6). There is strong evidence of a group effect (F_{4,587} = 48.91, p< .01), age
(F_{1,1402} = 11.05, p<.01), and curvilinearity (F_{1,1402} = 26.75, p<.01), as well as evidence that the
time-courses vary by group (interaction with age F_{1,1402} = 2.90, p<.02; interaction with age-
squared F_{1,1402} = 4.7, p<.01). The group effect is driven primarily by the differences between
typically developing children (TDC) and the three ASD groups (all p<.01); and that the
difference between TDC and psychiatric controls is less pronounced (an increase of ~31 SRS
points, p=.05). We also find that the age by group interaction is primarily driven by differences
in slopes in the female ASD group, which is the only group presenting with a sharp decrease and
substantial curvilinearity (t_{1402} = 3.28, p <.01 and t_{1402} = -4.08, p<.01, respectively).

3.4 Discussion
This cohort-sequential study demonstrated that AVR exhibits trait-like stability from preschool
to young adulthood across the entire range of variation in which it manifests in childhood. Inter-
individual variation in SRS scores was highly preserved, and growth curve modeling confirmed a
marked degree of measurement reliability over time. Further, the course of individuals’ scores
reflected distinct separations in symptom burden for deficiency in AVR, overwhelmingly
differentiating controls from ASD-affected individuals. To the extent that cohort-sequential
designs are a valid proxy for longitudinal measurement of an individual over the course of a
given period of time, these results demonstrate that throughout the entire distribution observed in
nature, AVR exhibits stability from childhood to early adulthood. In this sense, subclinical
variations in AVR are as stable as autism itself.

With stability of .91, AVR exceeds that of IQ (~.63) (Plomin & Deary, 2015), a core
construct in differential psychology, behavioral genetics, and human development; and other
psychopathological traits measured from adolescence to early adulthood, such as conduct
disorder (~.50) (Murray and Farrington, 2010) and borderline personality disorder (~.52) (Bornovalova et al., 2009; Chanen et al., 2004).

Other prior studies have explored the trajectories of ASD symptomologies. One found that the number and severity of repetitive sensorimotor behaviors either persisted or somewhat improved (Richler, Huerta, Bishop, & Lord, 2010). A recent ten-year prospective study demonstrated two trajectories in the Vineland Adaptive Behavior Scale: one with modest gains and another with stable impairments (Baghdadli et al., 2011). Other work has described six developmental trajectories for core autism symptom domains, with most showing a stable to slight improvement in functioning, such that a high-functioning individual generally remained as such longitudinally. Most low-functioning individuals also maintained stable levels of functioning, with the exception of a group of “bloomers” who throughout time achieved levels of functioning similar to higher-functioning individuals (Fountain, Winter, & Bearman, 2012). Yet critically, no studies have examined the course of AVR across the entire range of variation, a knowledge gap this study sought to address.

There are several limitations of this study. First, the SRS does not capture the entirety of the complex construct of reciprocal social behavior; however, the field of social development is still grappling with how to define the construct, assign its parameters, and fully quantify it. Second, due to the cohort-sequential design, no individual was studied from early childhood to adulthood. Third, the affected subjects were receiving a variety of forms of therapy and while their trajectories could have been influenced by the interventions, their SRS scores still remained stable (see also Marrus et al., 2014). Finally, this prospective longitudinal study lacked statistical power to fully define specific trajectories of sub-groups of subjects. The study was, however, adequately powered to detect very modest time-rated changes in standardized scores of AVR—
on the order of less than one-half of a standard deviation over a decade—for major groupings of affected and unaffected subjects. Supporting our current results of longitudinal stability in typically-developing individuals, the means and standard deviations of SRS scores of our unaffected sibling sample match epidemiological, cross-sectional SRS data acquired from cross-cultural studies in Europe (Bolte et al., 2008) and Asia (Kamio et al., 2013). Thus, this study offers the first-ever population-wide appraisal of the stability of AVR using normed quantitative measurements over the entire course of childhood.

Beginning early in infancy, children become socially specialized (Johnson, 2001). The essential role of sociality in human development highlights the significance of inherited determinants of social variation, one of which has potential to be indexed by measurements of AVR as deployed in this longitudinal study. It will require a next generation of studies to determine the nature of AVR’s interactions with other genetic and environmental influences to confer susceptibility or resilience to maladaptive social and emotional outcomes. Characterization of AVR stands to enhance the precision of future biomarker research exploring neural and psychophysiological signatures of social and emotional development, the promise of which is to advance understanding of the mechanisms by which genes influence behavior and to elucidate specific targets for preventive intervention of inherited syndromes of social and neurodevelopmental impairment.

3.5 Methods

3.5.1 Study Design
The study utilized a longitudinal cohort-sequential design, a method which reduces some of the practical constraints of classic longitudinal designs (Nesselroade and Baltes, 1979). Limited repeated measurements of independent age cohorts yields temporally overlapping assessments of
the different groups, allowing for the determination of a developmental trend, or growth curve, through the linkage of adjacent portions of circumscribed longitudinal data (Nesselroade and Baltes, 1979). These temporally overlapping assessments can be seen in Supplemental Figure 1, which depicts the span from baseline evaluation to final follow up of every subject with more than one maternal SRS (N=527). Subjects of all ages were enrolled on a rolling basis over the course of the study (Supplemental Figure 2). We administered most SRSs within two consecutive 5-year phases during 2003-2013, including in the dataset a small number (85) of SRSs for 60 subjects, representing only 2.37% of the total sample, which existed prior to 2003.

3.5.2 Sample
This report is a continuation of a previously published study (Constantino et al., 2009), an ongoing longitudinal evaluation of AVR. The subject group comprised 602 socio-economically diverse, predominantly Caucasian individuals. To be eligible for the study, the subject needed to be either 1) an individual with an ASD diagnosis (henceforth referred to as “ASD subject”); 2) the sibling of an ASD subject; 3) an individual with a non-ASD psychiatric diagnosis (“psych subject”); 4) or the sibling of a psych subject. Subjects were excluded from the study if any ambiguity existed regarding a singular primary ASD diagnosis.

Diagnostic Procedures
Diagnoses of ASD and non-ASD disorders were conducted via clinical examination by psychiatrists with substantial expertise in the diagnosis and treatment of autism and other psychiatric disorders. To be classified as an “ASD subject” for the purposes of this study, the individual must have received one of the following DSM-IV pervasive developmental disorder (PDD) diagnoses: ASD (299.0), Asperger’s Disorder (AS; 299.80), or PDD-Not Otherwise Specified (NOS; 299.80). Subjects who had an SRS score at or above the 85th percentile who
had been recruited by means other than a psychiatrist were administered the ADI-R and/or ADOS by a certified mental health professional to rule in or out a diagnosis of ASD.

**Sample characteristics**

Of the total of 602 subjects, the majority were ASD subjects (305), 99 of which were from 49 multiplex families, and 206 of which were from 206 simplex families. Our sample—which was clinically ascertained—was principally composed of individuals with ASD and their male siblings. The distributions of SRS scores for children within ASD-affected families are distinctly different for simplex and multiplex families, with multiplex families exhibiting a fully unimodal distribution, and simplex families demonstrating a bimodal distribution (Virkud et al., 2009; Frazier et al., 2010). Therefore, the mode of inheritance holds significant implications for unaffected siblings. This phenomenon reflects the established difference between the groups in the proportion to which rare, highly penetrant mutations versus cumulative additive genetic influences are responsible for trait variation. Yet given the differences in mode of transmission, the stability of trait variation was the same, regardless of inheritance mode.

The broad psychiatric diagnoses represented in the 43 psych subjects are as follows: attention deficit hyperactivity disorder (ADHD) (25), mood and anxiety disorders (15), and developmental delay (3). None of the psych subjects had an ASD diagnosis.

During this study, most subjects did not receive a change in clinical diagnosis, with the following exceptions: 1) Nine of the undiagnosed siblings of index ASD subjects were later diagnosed with ASD and included in the appropriate ASD-affected subgroup; 2) nine undiagnosed siblings of index psych subjects were later diagnosed with a psych condition. Notably, none of the psych siblings were later diagnosed with ASD.
3.5.3 Measures

The Social Responsiveness Scale
The SRS is a 65-item measure of AVR, which capitalizes on observations of children in naturalistic social contexts, by parent and/or teacher report. Its internal consistency is very high (alpha = ~.95), and it distinguishes ASD-affected individuals from controls with a Cohen’s d effect size of ~2.7 and from individuals with other psychiatric conditions with an effect size of ~1 (Constantino & Gruber, 2007; Constantino, Przybeck, Friesen, & Todd, 2000). It characterizes variation in the two DSM-5 domains of ASD: social communication and interaction (SCI) and restricted interests and repetitive behaviors (RRB). Prior studies of the SRS in clinical and epidemiological populations have established that these two subdomains encompass a unitary factor structure (Constantino et al., 2004), with the exception of one analysis of 14,744 subjects, which suggested a minimally better fit for a two-factor structure (Frazier et al., 2012). Given this established factor structure and that restriction of the number of SRS items erodes its ability to distinguish between clinical groups (Constantino & Todd, 2003), we elected to use the empirically-supported total score as the metric of analysis; however, subdomain analyses are provided in Supplemental Materials 7, 9, and 10.

Autism Diagnostic Observation Schedule-2 (ADOS-2) and Autism Diagnostic Interview-Revised (ADI-R)
To provide an independent measurement of autistic traits, ADOS-2 and ADI-R assessments were conducted in a research reliable manner by a certified individual on a subset of ASD-affected subjects (N=68).

3.5.4 Procedures
Subjects were recruited via a number of means: 1) their physicians from either (a) the Washington University Child and Adolescent clinic or (b) from outpatient child psychiatry practices in the greater St. Louis metropolitan area, all of which were general clinics, not ASD-
specific; 2) advertising on National Public Radio (NPR); 3) Autism Speaks; 4) local television programming; 5) other studies in which the laboratory was engaged; and 6) the laboratory’s website. The psychiatric clinics serve children from 0-18, and in some cases follow the individuals past age 18 to provide continuity of care for the individuals and their families. Some participants were presenting for an initial diagnosis and others for follow-up.

We mailed SRSs to all parents, and to teachers of school-age subjects. Teacher reports were a secondary measure, with different teachers reporting on a given subject in successive years. The teacher who completed the SRS had known the child for a minimum of two months and was selected by the parent as most knowledgeable of the child. SRSs were requested annually.

We enrolled a total of 1,026 subjects, of which 268 did not continue (overall attrition rate=26.1%). Of the remaining 758 subjects, 602 met inclusion criteria, 530 of whom had more than one SRS. The 72 subjects who had only one SRS were included in the dataset for analyzing inter-rater agreement. Because of the low number of paternal or undefined SRSs (148 and 90, respectively), we excluded paternal and undefined reports from the analysis. For the 602 subjects included in the analysis, 1,997 maternal SRSs were performed for 592 subjects, and 1,586 teacher measures were performed for 515 subjects, with 10 subjects having only teacher measures and 87 subjects having only maternal measures.

Tables 1 and 2 provide descriptive statistics of maternal and teacher assessments, respectively, including selected sample characteristics as a function of specific groupings of subjects. Some of the planned annual follow-up assessments were not returned by informants (missing); the mean duration of follow-up for each group compared to the mean number of SRSs
reflects the average number of annual assessments missing. Supplemental Figures 2 – 5 provides further depictions of the sample.

3.5.5 Data Analysis
In order to analyze the agreement between maternal and teacher ratings and between successive ratings by mothers and teachers, we used non-parametric LOESS curve fitting and regression analysis. We employed fixed effects and mixed models to analyze the longitudinal course of serial maternal ratings as a function of group, age, and age-squared; and used Bayesian Information Criterion (BIC) as the model selection tool. The primary contrasts are between the three ASD groups, the psychiatric comparison group, and unaffected controls (with group subsuming gender). Since we found only modest differences in SRS scores between the unaffected siblings of both ASD subjects and psych subjects, we pooled these control groups. Raw SRS scores were used for all analyses to allow for consistent comparison of results across published studies, the majority of which have utilized raw scores.

There are no effects of age at first assessment on the later value of AVR, \( p = 0.3819 \), from a linear regression). There is no association between baseline SRS score and length of observation (number of years between first and last assessment, \( p = 0.9357 \)). By group, ASD subjects were observed for 4.9 years on average, compared to 4.6 years for non-ASD participants \( (p = 0.2370) \). Further, ASD were as likely as non-ASD participants to have multiple measurements \( (p = 0.29, \) from a logistic regression). Missing data were accounted for by using full information maximum likelihood (FIML). All analyses were conducted using SAS 9.4, with the exception of Supplemental Figures 1 and 3 – 5, which were conducted in R (R Studio, 2018).
3.6 Acknowledgements
This work was supported by Grant HD042541 from the National Institute of Child Health and Human Development (to J.N.C.). The study protocol was approved by Washington University in St. Louis School of Medicine Human Research Protection Office. We gratefully acknowledge the parents and families participating in the Washington University Social Developmental Studies program (sdslab.wustl.edu), for their contribution to this research effort. Without their willingness to participate, this research would not have been possible, and we greatly appreciate their dedication and contribution to deepening our understanding of ASD.
3.7 Figures

**Figure 1: Maternal and teacher SRS relationship.** Linear relationship of all available pairings of maternal and teacher SRSs for individual subjects (N=1,401 pairs). The shaded light blue around the darker blue linear regression line is the 95% confidence interval (CI).
**Figure 2: Maternal baseline SRS scores.** Mean raw maternal baseline total SRS scores as a function of group (N=592). Whiskers represent standard deviation. Horizontal lines inside boxes represent the median scores and the diamonds represent the means.
Figure 3: Stability of SRS. Stability of randomly-selected pairs of successive maternal SRS measurements (N=527 pairs). The dotted line is the 45-degree line.
Figure 4: Individual SRS trajectories. Individual childhood trajectories of maternal-report total SRS scores, as a function of ASD diagnostic status (N=527).
Figure 5: AVR over the life course. The pattern of symptom severity over the life course of subjects portrayed as a function of each study group (N per group provided in Table 1). All subjects with at least two maternal assessments are represented (N=527), with the number of subjects per follow-up interval provided in Table 4.
Table 1: Maternal SRS measures

<table>
<thead>
<tr>
<th>Subject Type</th>
<th>Number of subjects</th>
<th>Mean age (SD): baseline</th>
<th>Age range: baseline</th>
<th>Mean Age (SD): latest follow-up</th>
<th>Age range: latest follow-up</th>
<th>Total number SRS</th>
<th>Mean baseline SRS (SD)</th>
<th>Mean duration of Follow-Up (SD)</th>
<th>Mean number of SRS (SD)</th>
<th>Range of follow-up years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male ASD simplex family</td>
<td>170</td>
<td>8.0 (3.9)</td>
<td>2.6-18.0</td>
<td>12.4 (4.9)</td>
<td>16.1-23.8</td>
<td>646</td>
<td>101.8 (29.0)</td>
<td>3.80 (2.09)</td>
<td>0 - 14.0</td>
<td></td>
</tr>
<tr>
<td>Male ASD multiplex family</td>
<td>89</td>
<td>6.1 (2.8)</td>
<td>3.0-15.7</td>
<td>11.2 (4.0)</td>
<td>16.1-23.8</td>
<td>378</td>
<td>94.0 (32.6)</td>
<td>4.25 (1.94)</td>
<td>0 - 12.6</td>
<td></td>
</tr>
<tr>
<td>Female ASD simplex / multiplex family</td>
<td>40</td>
<td>8.3 (3.7)</td>
<td>3.1-16.2</td>
<td>12.2 (5.2)</td>
<td>10.6-17.0</td>
<td>133</td>
<td>101.5 (29.1)</td>
<td>3.32 (1.66)</td>
<td>0 - 12.0</td>
<td></td>
</tr>
<tr>
<td>Male unaffected siblings</td>
<td>149</td>
<td>8.1 (4.2)</td>
<td>2.6-22.2</td>
<td>12.0 (5.2)</td>
<td>17.9-23.1</td>
<td>412</td>
<td>25.6 (19.9)</td>
<td>2.77 (1.28)</td>
<td>0 - 11.7</td>
<td></td>
</tr>
<tr>
<td>Female unaffected siblings</td>
<td>64</td>
<td>9.8 (4.7)</td>
<td>2.9-21.7</td>
<td>12.3 (5.0)</td>
<td>12.2-14.4</td>
<td>116</td>
<td>21.0 (24.9)</td>
<td>1.81 (.77)</td>
<td>0 - 10.3</td>
<td></td>
</tr>
<tr>
<td>Male psych condition</td>
<td>43</td>
<td>11.4 (3.5)</td>
<td>4.3-21.5</td>
<td>17.0 (4.2)</td>
<td>9.6-12.7</td>
<td>196</td>
<td>51.4 (34.1)</td>
<td>4.56 (1.82)</td>
<td>0.1 - 15.1</td>
<td></td>
</tr>
<tr>
<td>Male unaffected siblings, psych-affected family</td>
<td>37</td>
<td>10.6 (5.0)</td>
<td>3.0-28.7</td>
<td>15.5 (5.5)</td>
<td>15.3-24.1</td>
<td>116</td>
<td>19.1 (20.0)</td>
<td>3.14 (1.29)</td>
<td>0 - 9.4</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>592</td>
<td>8.4 (4.2)</td>
<td>2.6-28.7</td>
<td>12.6 (5.1)</td>
<td>9.5-22.7</td>
<td>1,997</td>
<td>63.9 (45.6)</td>
<td>3.37 (1.85)</td>
<td>0 - 15.1</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2: Teacher SRS measures

<table>
<thead>
<tr>
<th>Subject Type</th>
<th>Number of Subjects</th>
<th>Mean age (SD): baseline</th>
<th>Age range: baseline</th>
<th>Mean age (SD): latest follow-up</th>
<th>Age range: latest follow-up</th>
<th>Total number of SRS</th>
<th>Mean baseline SRS (SD)</th>
<th>Mean Duration of Follow-Up</th>
<th>Mean number of SRS (SD)</th>
<th>Range of follow-up years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male ASD simplex family</td>
<td>155</td>
<td>7.7 (3.7)</td>
<td>2.8-17</td>
<td>11.6 (3.9)</td>
<td>14.5-17.9</td>
<td>547</td>
<td>92.6 (30.8)</td>
<td>3.9</td>
<td></td>
<td>3.53 (1.56)</td>
</tr>
<tr>
<td>Male ASD multiplex family</td>
<td>87</td>
<td>6.1 (2.8)</td>
<td>2.5-13</td>
<td>10.4 (3.8)</td>
<td>10.8-17.4</td>
<td>316</td>
<td>95.4 (32.7)</td>
<td>4.3</td>
<td></td>
<td>3.64 (1.56)</td>
</tr>
<tr>
<td>Female ASD simplex / multiplex family</td>
<td>35</td>
<td>7.4 (3.2)</td>
<td>3.1-16</td>
<td>10.4 (3.2)</td>
<td>9.7-18.0</td>
<td>104</td>
<td>95.3 (32.8)</td>
<td>3.0</td>
<td></td>
<td>2.97 (1.12)</td>
</tr>
<tr>
<td>Male unaffected sibling</td>
<td>124</td>
<td>8.0 (3.9)</td>
<td>2.7-16</td>
<td>11.3 (4.0)</td>
<td>9.1-17.4</td>
<td>314</td>
<td>32.0 (26.4)</td>
<td>3.3</td>
<td></td>
<td>2.54 (1.81)</td>
</tr>
<tr>
<td>Female unaffected sibling</td>
<td>40</td>
<td>8.0 (4.0)</td>
<td>3.2-17</td>
<td>10.9 (3.8)</td>
<td>7.9-16.3</td>
<td>89</td>
<td>24.9 (27.6)</td>
<td>2.8</td>
<td></td>
<td>2.23 (1.42)</td>
</tr>
<tr>
<td>Male psych condition</td>
<td>41</td>
<td>11.6 (3.1)</td>
<td>5.1-16</td>
<td>14.8 (2.9)</td>
<td>13.7-17.0</td>
<td>133</td>
<td>55.0 (31.0)</td>
<td>3.4</td>
<td></td>
<td>3.24 (1.27)</td>
</tr>
<tr>
<td>Male unaffected siblings, psych-affected family</td>
<td>33</td>
<td>9.7 (3.6)</td>
<td>3.0-15</td>
<td>13.2 (3.7)</td>
<td>10.0-22.7</td>
<td>82</td>
<td>24.8 (24.3)</td>
<td>3.5</td>
<td></td>
<td>2.48 (1.91)</td>
</tr>
<tr>
<td>Total</td>
<td>515</td>
<td>7.9 (3.8)</td>
<td>2.5-17</td>
<td>11.5 (3.9)</td>
<td>10.0-16.0</td>
<td>1,585</td>
<td>66.1 (42.6)</td>
<td>3.6</td>
<td></td>
<td>3.08 (1.35)</td>
</tr>
</tbody>
</table>
Table 3: Influences on baseline maternal SRS scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>N</th>
<th>T</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher score</td>
<td>0.22</td>
<td>505</td>
<td>11.9</td>
<td>&lt; .01*</td>
</tr>
<tr>
<td>ASD</td>
<td>63.50</td>
<td>592</td>
<td>15.2</td>
<td>&lt; .01*</td>
</tr>
<tr>
<td>Other psych diagnosis</td>
<td>28.30</td>
<td>592</td>
<td>6.1</td>
<td>&lt; .01*</td>
</tr>
<tr>
<td>Unaffected sibling: simplex and multiplex families (^a)</td>
<td>2.90</td>
<td>592</td>
<td>0.7</td>
<td>.48</td>
</tr>
<tr>
<td>Age</td>
<td>-0.36</td>
<td>592</td>
<td>-2.3</td>
<td>.02*</td>
</tr>
<tr>
<td>Gender</td>
<td>1.38</td>
<td>592</td>
<td>0.5</td>
<td>.59</td>
</tr>
<tr>
<td>Verbal(^b)</td>
<td>-11.50</td>
<td>592</td>
<td>-2.3</td>
<td>.02*</td>
</tr>
<tr>
<td>IQ(^c)</td>
<td>-0.21</td>
<td>228</td>
<td>-2.8</td>
<td>.06</td>
</tr>
</tbody>
</table>

Note. Relative influences of selected subject characteristics on baseline maternal SRS scores for all subjects.
\(^a\) elevated in comparison to psychiatric sibling controls and population norms
\(^b\) capable of phrase speech
\(^c\) Supplement 8 provides IQ tests utilized
Table 4: Mother and teacher test-retest correlations

<table>
<thead>
<tr>
<th>Interval</th>
<th>N</th>
<th>Avg. time lag</th>
<th>Test-retest correlation</th>
<th>N</th>
<th>Avg. time lag</th>
<th>Test-retest correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>527</td>
<td>1.95</td>
<td>.90</td>
<td>508</td>
<td>1.78</td>
<td>.74</td>
</tr>
<tr>
<td>2</td>
<td>347</td>
<td>1.64</td>
<td>.91</td>
<td>312</td>
<td>1.54</td>
<td>.69</td>
</tr>
<tr>
<td>3</td>
<td>239</td>
<td>1.73</td>
<td>.93</td>
<td>112</td>
<td>1.92</td>
<td>.74</td>
</tr>
<tr>
<td>4</td>
<td>137</td>
<td>1.86</td>
<td>.92</td>
<td>71</td>
<td>1.97</td>
<td>.71</td>
</tr>
<tr>
<td>5</td>
<td>86</td>
<td>1.80</td>
<td>.93</td>
<td>49</td>
<td>1.71</td>
<td>.63</td>
</tr>
<tr>
<td>6</td>
<td>55</td>
<td>1.61</td>
<td>.95</td>
<td>15</td>
<td>1.35</td>
<td>.56</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>1.45</td>
<td>.95</td>
<td>5</td>
<td>1.34</td>
<td>.60</td>
</tr>
</tbody>
</table>

*Note.* Correlations between successive measurements showing the stability of both maternal and teacher total SRS scores.
Table 5: Model-fitting of longitudinal data

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>df&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>--</td>
<td>17915.9</td>
<td>17798.6</td>
<td>17937.6</td>
</tr>
<tr>
<td>BIC</td>
<td>--</td>
<td>17924.7</td>
<td>17846.8</td>
<td>17985.8</td>
</tr>
<tr>
<td>Group</td>
<td>4,587</td>
<td>22.18</td>
<td>48.91</td>
<td>39.55</td>
</tr>
<tr>
<td></td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1,1402</td>
<td>2.72</td>
<td>11.05</td>
<td>7.66</td>
</tr>
<tr>
<td></td>
<td>p=.09</td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Age squared</td>
<td>1,1402</td>
<td>6.73</td>
<td>26.75</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Age by Group</td>
<td>4,1402</td>
<td>1.19</td>
<td>2.90</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>p=.31</td>
<td>p=.02</td>
<td>p&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Age squared by Group</td>
<td>4,1402</td>
<td>1.57</td>
<td>4.70</td>
<td>5.27</td>
</tr>
<tr>
<td></td>
<td>p=.18</td>
<td>p&lt;.01</td>
<td>p&lt;.01</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>For the corresponding Type III test of fixed effect.

<sup>Note.</sup> Model 1: Fixed effects only; Model 2: Random intercept, fixed slope; Model 3: Random intercept and slope. All models assume a spatial exponential residual correlation structure and include group, age, age-squared, and interaction between group and age and age-squared.
Table 6: Fixed effects estimates of model 2

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>25.76</td>
<td>3.85</td>
<td>6.68</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Male ASD (Simplex)</td>
<td>69.84</td>
<td>6.16</td>
<td>11.33</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Male ASD (Multiplex)</td>
<td>62.94</td>
<td>7.08</td>
<td>8.88</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Female ASD</td>
<td>47.51</td>
<td>1.78</td>
<td>4.03</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Male Psych</td>
<td>31.94</td>
<td>6.25</td>
<td>1.97</td>
<td>.05</td>
</tr>
<tr>
<td>age</td>
<td>-0.13</td>
<td>0.64</td>
<td>-0.20</td>
<td>.84</td>
</tr>
<tr>
<td>age*age</td>
<td>-0.02</td>
<td>0.04</td>
<td>-0.62</td>
<td>.53</td>
</tr>
<tr>
<td>age* Male ASD (Simplex)</td>
<td>1.06</td>
<td>0.99</td>
<td>1.07</td>
<td>.29</td>
</tr>
<tr>
<td>age* Male ASD (Multiplex)</td>
<td>1.81</td>
<td>1.23</td>
<td>1.46</td>
<td>.14</td>
</tr>
<tr>
<td>age* Female ASD</td>
<td>5.95</td>
<td>1.82</td>
<td>3.28</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>age* Male Psych</td>
<td>0.27</td>
<td>2.17</td>
<td>0.12</td>
<td>.90</td>
</tr>
<tr>
<td>age<em>age</em> Male ASD (Simplex)</td>
<td>-0.04</td>
<td>0.04</td>
<td>-0.94</td>
<td>.34</td>
</tr>
<tr>
<td>age<em>age</em> Male ASD (Multiplex)</td>
<td>-0.12</td>
<td>0.06</td>
<td>-2.09</td>
<td>.04</td>
</tr>
<tr>
<td>age<em>age</em> Female ASD</td>
<td>-0.27</td>
<td>0.07</td>
<td>-4.08</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>age<em>age</em> Male Psych</td>
<td>-0.04</td>
<td>0.07</td>
<td>-0.61</td>
<td>.54</td>
</tr>
</tbody>
</table>

*Note.* Model 2: Random intercept, fixed slope. This model, which fit the data best, assumes a spatial exponential residual correlation structure and includes group, age, age-squared, and interaction between group and age and age-squared.
Supplemental Figure 1: Individual subject span: baseline to final follow-up
**Supplemental Figure 2: Recruitment and enrollment.** Depicts the rolling recruitment design of the study, along with the number of enrollments (i.e., new baseline assessments) per year.
Supplemental Figure 3: Subjects per age group as a function of evaluation type. Depicts total number of subjects per age group as a function of maternal SRS evaluation type (N=592). FU=Follow-up
Supplemental Figure 4: Median duration of follow-up. Depicts median duration of follow-up using maternal SRS data (N=592) as a function of year of enrollment.
Supplemental Figure 5: Number of baseline evaluations per year per age group. *Note.* Maternal baseline evaluations are depicted (N=592) as a function of age group.
Supplemental Figure 6: Stability of SCI subdomain. Note. Stability of randomly-selected pairs of successive maternal SCI measurements (N=527 pairs). The dotted line is the 45-degree line. There are no effects of age at first assessment on the later value of SCI ($p = 0.5798$, from a linear regression). The correlation between successive SCIs is 0.91, with an $R^2 = 0.82$. 
Supplemental Figure 7: Stability of RRB subdomain. Note. Stability of randomly-selected pairs of successive maternal RRB measurements (N=527 pairs). The dotted line is the 45-degree line. There is a mild suggestion of an association between age at first assessment and RRB ($p = 0.0121$, from a linear regression). The correlation between successive RRBs is 0.90, with an $R^2 = 0.81$. 
Supplemental Figure 8: Trajectory of SCI over the life course. Note. Individual childhood trajectories of maternal-report SCI scores, as a function of ASD diagnostic status (N=527).
Supplemental Figure 9: Trajectory of RRB over the life course. Note. Individual childhood trajectories of maternal-report RRB scores, as a function of ASD diagnostic status (N=527).
Supplemental Table 1: Correlations of SRS with ADOS-2 and ADI-R

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Measurement Type</th>
<th>SRS SCI</th>
<th>SRS RRB</th>
<th>SRS Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADOS-2: Social Affect (N=21)</td>
<td>Social Affect</td>
<td>.526*</td>
<td>---</td>
<td>.546*</td>
</tr>
<tr>
<td>ADI-R: Reciprocal Social Interaction Domain (N=64)</td>
<td>Reciprocal Social Interaction Domain</td>
<td>.332**</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>ADI-R: Repetitive and Stereotyped Patterns of Behavior (N=64)</td>
<td>Repetitive and Stereotyped Patterns of Behavior</td>
<td>---</td>
<td>0.124</td>
<td>---</td>
</tr>
</tbody>
</table>

* correlation is significant at the .05 level
** correlation is significant at the .005 level
Supplemental Table 2: Test-retest correlations of SCI and RRB subdomains

<table>
<thead>
<tr>
<th>Interval</th>
<th>RRB Maternal</th>
<th>RRB Teacher</th>
<th>SCI Maternal</th>
<th>SCI Teacher</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90</td>
<td>0.73</td>
<td>0.89</td>
<td>0.72</td>
</tr>
<tr>
<td>2</td>
<td>0.88</td>
<td>0.69</td>
<td>0.90</td>
<td>0.67</td>
</tr>
<tr>
<td>3</td>
<td>0.91</td>
<td>0.74</td>
<td>0.92</td>
<td>0.74</td>
</tr>
<tr>
<td>4</td>
<td>0.89</td>
<td>0.70</td>
<td>0.91</td>
<td>0.70</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>0.62</td>
<td>0.92</td>
<td>0.64</td>
</tr>
<tr>
<td>6</td>
<td>0.91</td>
<td>0.40</td>
<td>0.95</td>
<td>0.55</td>
</tr>
<tr>
<td>7</td>
<td>0.96</td>
<td>0.62</td>
<td>0.95</td>
<td>0.50</td>
</tr>
</tbody>
</table>

*Note.* Correlations between successive measurements showing the stability of both maternal and teacher SRS subdomain scores. The values for time lag and sample size remain the same from Table 4 in the main text.
3.8 Lists

The list below provides all of the IQ tests utilized with our subjects.

- Bayley Scales of infant development, 2nd edition (BSID-II)
- Comprehensive Test of Nonverbal Intelligence (CTONI)
- Developmental Assessment of Young Children (DAYC)
- Developmental Profile-II (DP-II)
- Differential Ability Scales (DAS)
- Differential Ability Scales-II (DAS-II)
- Kauffman Assessment Battery for Children-II (KABC-II)
- Kauffman Brief Intelligence Test-II (KBIT-2)
- Leiter International Performance Scale-Revised (LIPS-R)
- Mullen Scales of Early Learning (MSEL)
- Raven’s Matrices
- Stanford-Binet Intelligence Scales-4th and 5th editions (SBIS-4,5)
- Slosson Intelligence Test-Revised (SIT-R)
- Test of Nonverbal Intelligence-3 (TONI-3)
- Wechsler Abbreviated Scale of Intelligence-II (WASI-II)
- Wechsler Adult Intelligence Scale-III,IV (WAIS-III, IV)
- Wechsler Intelligence Scale for Children-III,IV (WISC-III,IV)
- Wechsler Nonverbal Scale of Ability (WNV)
- Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III)
- Woodcock Johnson Tests of Cognitive Abilities – III (WJ-III)
3.9 References


http://doi.org/10.1176/appi.ajp.2010.10020223

http://doi.org/10.1001/jamapsychiatry.2014.476

https://doi.org/10.1089/cap.2014.0055


http://doi.org/10.1016/j.chiabu.2015.04.010


Chapter 4: The Contribution of Autism-Related Common Variants to the Familial Aggregation of Quantitative Autistic Traits

Rachael E. Wagner, William B. Howells, Emma Johnson, Arpana Agrawal, Jennifer K. Lowe, Daniel H. Geschwind, and John N. Constantino
4.1 Abstract

Autism spectrum disorder (ASD) genome-wide association studies (GWAS) have used polygenic risk scores (PRS) to predict ASD diagnosis in the general population (Nagelkerke’s $R^2=2.45\%$). Critically, large genetic epidemiological studies have demonstrated that family members of affected individuals exhibit subclinical quantitative autistic traits (QAT), suggesting that diagnosed ASD is the pathological tail of a continuous distribution of QAT. Recent molecular genetic studies have reinforced these findings by uncovering a genetic correlation between ASD and QAT in the general population, emphasizing the importance of employing QAT in GWAS. Yet no study to date has examined the extent to which common variant burden contributes to variance in QAT in a familial multiplex sample—one enriched for inherited ASD risk. Given the elevation of QAT among unaffected members in multiplex families, we examined the contribution of ASD-PRS to QAT, as measured by the Social Responsiveness Scale, in 1491 subjects from multiplex families in the Autism Genetic Resource Exchange (AGRE)—with and without an ASD diagnosis. Using the iPSYCH-ASD-GWAS comprised of cases (N=8,605) and controls (N=19,526) as our discovery dataset, we estimated how much variance in QAT ASD-PRS explained in our target dataset, AGRE. We also examined if there were interaction effects of diagnosis and sex. The ASD-PRS explained 0.34% of the variance in QAT ($p=0.02$). There was no significant interaction effect of disorder status ($p=0.51$), but there was a significant interaction of sex ($p=0.03$) and a significant interaction effect of ASD and sex ($p=0.031$). Our results reveal that autism-related polygenic load, as measured via the ASD-PRS, significantly predicts QAT—critically, in both affected and unaffected family members. Further, the association of ASD-PRS with QAT in females was modified by diagnosis, indicating that this relationship was the strongest among affected females. The unique nature of this study’s
multiplex familial sample enables a novel demonstration that common polygenic variation in ASD contributes to familial QAT.

4.2 Introduction

The aggregation of autism spectrum disorder (ASD) in families is well established (Sandin et al., 2014), but its genetic architecture remains poorly understood. Three classes of genetic risk have been identified as contributing to ASD: de novo mutations, rare inherited variants, and common polygenic variation. Several models, none of which are mutually exclusive, have been proposed to explain this aggregation. A major gene model hypothesizes that either one highly penetrant rare mutation or a restricted number of moderately to highly penetrant mutations contribute to the familial nature of ASD (O’Roak et al., 2011; Schaaf et al., 2011; Torre-Ubieta et al., 2016). In contrast, polygenic models assume that many common inherited variants (minor allele frequency > 0.05), each of very small effect, collectively account for a substantial portion of the variance of total ASD risk and contribute to the disorder’s familial aggregation (Anney et al., 2012; Cross-Disorder, 2018; Gaugler et al., 2014; Grove et al., 2019; Klei, Sanders, Murtha, Hus, Lowe, Willsey, Moreno-De-Luca, Timothy, et al., 2012). Further, single-nucleotide polymorphism (SNP) data heritability estimates suggest that common variants explain a large proportion of the population-wide variance of ASD (Gaugler et al., 2014; Grove et al., 2019; Klei, Sanders, Murtha, Hus, Lowe, Willsey, Moreno-De-Luca, Yu, et al., 2012) as in many other common, complex, psychiatric disorders (Cross-Disorder, 2018; Ripke et al., 2014; Smoller et al., 2018; Wray et al., 2018). In fact, the latest genome-wide association study (GWAS) of autism has revealed five statistically significant loci and a SNP-heritability of ~12% (Grove et al., 2019). Some evidence suggests that even highly penetrant de novo mutations are operating on a background of high polygenic risk burden for ASD (Weiner et al., 2017). A recent large
exome sequencing study of *de novo* and rare inherited coding variation uncovered 102 risk genes for ASD; an examination of the overlap of these genes with GWAS common variants revealed a significant association with GWAS results of schizophrenia and educational attainment (genetically correlated with ASD), as well as demonstrated a significant association of the gene KMT2E with ASD by both common and rare risk variation (Satterstrom et al. 2020).

Characterizing common variant risk has been more challenging than the field initially hoped. Given the underwhelming amount of variance explained by GWAS and relatively few genome-wide significant loci uncovered, the field of quantitative genetics developed polygenic risk scores (PRS), which aim to capture the aggregate effect of genetic variants (polygenicity), and as such, are similar to heritability ($h^2$). To ensure the validity of these scores, it is essential that the effect sizes are estimated in an independent cohort. A cumulative index of measured genetic liability to a disorder, PRS attempts to quantify within an individual the aggregate effect of common variants for a given trait, typically calculated as the sum of trait-associated alleles across the genome, weighted by effect size (Bogdan et al., 2018; Gandal et al., 2016). The discovery GWAS and the target dataset in which the PRS are calculated must be restricted to the same ancestral population to avoid spurious associations due to population stratification. Currently, PRS are more predictive in European populations than in all other ancestries, due to the overrepresentation of Europeans in GWAS (Martin et al. 2019). Given the clinical utility of PRS for a variety of biomedical conditions, this lack of diversity in genetic studies could lead to an even greater accentuation of health disparities between individuals of European descent and underserved non-European populations (Martin et al. 2019). Therefore, it is of the utmost importance that concerted efforts be made to diversify genetic studies.
Both rare and common genetic variation act additively to contribute to ASD risk in simplex (one clinically-affected individual) and multiplex (more than one clinically-affected individual) families, though evidence also indicates that there exist at least partially disparate genetic pathways in simplex versus multiplex families. A growing body of literature suggests that in simplex families, relatively rare, *de novo*, more highly penetrant and deleterious genetic mutations are present, whereas in multiplex families, evidence suggests a more highly polygenic, common variant burden (Bernier et al., 2011; Connolly et al., 2019; Klei, Sanders, Murtha, Hus, Lowe, Willsey, Moreno-De-Luca, Yu, et al., 2012). Supporting a polygenic (additive genetic) model in multiplex families is the fact that relatives of children with ASD demonstrate elevated quantitative autistic traits. (Lowe et al., 2015; Lyall et al., 2014; Virkud et al., 2009) In fact, first-degree relatives without an ASD diagnosis in multiplex families exhibit elevated quantitative autistic traits (QAT; subthreshold autistic-like deficits in social interaction and communication, behavioral rigidity, restricted interests and repetitive behaviors not severe enough to meet a relatively arbitrary cut-off for clinical diagnosis) (Lyall et al., 2014; Virkud et al., 2009), yet this pathological shift has not been found in first-degree relatives of probands in simplex families (Virkud et al., 2009). Family and twin literature supports the theory that clinical diagnosis of ASD is often the pathological tail of a continuous distribution of heritable quantitative autistic traits (Constantino and Todd, 2003, 2005; Robinson et al., 2011).

Given the clinical and epidemiological support for an additive, polygenic model of ASD, molecular genetic studies have begun to interrogate the relationship between diagnosed ASD and QAT to determine their genetic overlap. Employing a general population sample, Robinson and colleagues (Robinson et al., 2016) found a genetic correlation of ~.3 between ASD and a quantitative measure of autistic traits, the Social Communication Disorders Checklist (SCDC).
However, the SCDC captures only deficits in social communication and interaction and does not evaluate the other established domain of ASD symptomology: restricted interests and repetitive behaviors. Given this genetic overlap, St. Pourcain and colleagues found in a UK population-based study that ASD-PRS are higher in ASD cases than in pseudo-controls and were associated with variation in SCDC scores, but only at 8 years (PGC-ASD: adjusted R$^2_{\text{max}} = 0.13\%$, $P_{\text{min}} = 0.0042$) (2017). Bralten and colleagues employed a novel questionnaire of autistic traits in a Dutch general population sample to evaluate genetic overlap of ASD with each of the individual traits assessed by the questionnaire as well as employ an ASD PRS to predict phenotypic variance in autistic traits. They found small, yet statistically significant overlap with and prediction of some of the autistic traits ($R^2$ values ranging from .17-.54% for some subscales) (2017). Yet to our knowledge, no study to date has employed ASD-PRS in a multiplex familial sample to predict variation in a psychometrically-established valid and reliable phenotype that captures the continuous variation in trait distribution encompassing social communication and interaction and restricted interests and repetitive behaviors. A multiplex familial sample is an ideal sample in which to interrogate this question, as these individuals likely carry the highest common variant burden. To that end, we employed the Social Responsiveness Scale (SRS) (Constantino and Gruber, 2012), a comprehensive measure of QAT, in a multiplex family sample.

We hypothesized that common variants of small individual effect sizes contribute to the familial aggregation of QAT. To examine this question, we constructed a polygenic risk score (PRS) from the iPSYCH ASD GWAS, comprised of cases (N=8,605) and controls (19,526) (Gandal et al., 2018). We then investigated the contribution of common polygenic variation to QAT in our multiplex family cohort. Our primary question was to determine how much variance
in QAT phenotype an ASD-PRS derived from a case-control GWAS could explain in a multiplex familial sample. Finally, we conducted secondary analyses to determine if there were interaction effects of diagnosis and sex in our sample (Carter and Evans, 1969; Werling, 2016).

4.3 Results
The sample’s descriptive statistics are outlined in Table 1. The distributions of the PRS and SRS observed in the sample stratified by sex are shown in Figures 1 and 2, respectively; and the distributions of PRS and SRS stratified by affected status are depicted in Figures 3 and 4, respectively. The unadjusted mean PRS scores of the total sample, stratified by sex and diagnostic status are reported in Table 2.

4.3.1 ASD as Outcome
Prior to QAT analyses, we conducted a preliminary set of analyses with ASD diagnosis as our outcome measure to verify foundational hypothesized effects. To determine if odds ratios (OR) for ASD diagnosis increased with greater polygenic load, we separated individuals from the AGRE cohort into quartiles of PRS risk. Using a logistic regression model (adjusted for sex, PC1 – PC3, and family variance) to test for association between the PRS quartiles and diagnostic status, we found that OR in each PRS quartile increased with greater polygenic risk. As outlined in Table 3, compared to the lowest <25% percentile group, OR for individuals in the 75th percentile were statistically significant 1.450 (95% CI: 1.084, 1.941); \( p=0.012 \). The test for trend across PRS quartiles was significant, OR=1.235 (95% CI: 1.016, 1.245, \( p=0.023 \)), demonstrating a dose response effect of PRS on risk for ASD. As expected, the main effect of sex was large: the OR for males was 4.12 (95% CI: 3.27, 5.19, \( p=5.30E-32 \)).
4.3.1 QAT as Outcome

**Association between PRS and SRS**
As a preliminary step in our analyses, we conducted descriptive statistics to visualize the data.

The unadjusted Pearson correlation coefficient of PRS with SRS for the entire sample was \( r = 0.065 \) (95%CI: 0.014, 0.115, \( p = 0.013 \)). The correlation between PRS and SRS was higher for females \( r = 0.122 \) (95%CI: 0.036, 0.206; \( p = 0.005 \), n=521) than for males \( r = 0.029 \) (95%CI: -0.034, 0.119, \( p = 0.092 \), n=970) (Figure 5). When stratified by ASD status, the correlation for affected individuals \( r = 0.031 \) (95% CI: -0.033, 0.108; \( p = 0.342 \), n=921) was slightly lower than that of unaffected individuals \( r = 0.054 \) (95% CI: -0.033, 0.137; \( p = 0.214 \), n=538) (Figure 6). When stratified by ASD status and sex, the correlation of PRS with SRS in unaffected females was \( r = 0.055 \) (95% CI: -0.059, 0.167, \( p = 0.345 \), n = 299) and for affected females it was \( r = 0.186 \) (95% CI: 0.052, 0.311, \( p = .006 \), n = 215). For unaffected males, the correlation was \( r = 0.055 \) (95% CI: -0.072, 0.181, \( p = 0.397 \), n = 239) and for affected males, the correlation was \( r = -0.018 \) (95% CI: -0.092, 0.056, \( p = 0.634 \), n = 736) (Figure 7).

**Variance in QAT Explained by Common Variation in ASD Risk**
In our primary model, Model 1, to estimate the variance explained by PRS, we compared the pseudo-\( R^2 \) of a full model (PC1 – PC3, sex, and PRS) to the reduced model (eliminating PRS) (Table 3, Model 1). The variance explained (delta \( R^2 \)) by the PRS was 0.342% (\( p=0.023 \)). In Model 2, there was a significant interaction of sex with PRS, with the variance explained (delta \( R^2 \)) being 0.142% (\( p=0.027 \)). This is the statistical test of the different slopes depicted in Figure 5. In Model 3, we examined the interaction of PRS with diagnosis; here the interaction term was not significant (delta \( R^2 = 0.012\% \), \( p=0.511 \)). In other words, the variance explained in QAT was driven by both affected and unaffected subjects; the effect of ASD-PRS on QAT was not modified by an ASD diagnosis. Finally, in Model 4, we investigated the interaction of PRS, sex,
and diagnosis. The interaction effect was significant ($\Delta R^2 = 0.013\%, p=0.031$), a statistical demonstration of the affected female correlation of ASD-PRS with SRS, visualized in Figure 7. All modeling is outlined in detail in Table 4.

### 4.4 Discussion

Our results reveal that autism-related polygenic load, as measured via the ASD-PRS, not only significantly predicts ASD but is also related to QAT, such as those assessed using the SRS—critically, in both affected and unaffected family members. The uniqueness of this study is the application of ASD polygenic risk predictors to a multiplex family sample, enriched for inherited liability to ASD, incorporating both affected and unaffected family members who were uniformly phenotyped using a standardized measure of QAT. These findings corroborate a large body of genetic epidemiology research which has established a biological correspondence between clinically diagnosed ASD and subclinical QAT in family members of an affected individual. Although the variance explained was modest, its magnitude was in keeping with other analyses of social behavioral variation in the general population samples (Bralten et al., 2017; St. Pourcain et al., 2013; Robinson et al., 2016). Critically, there was no statistical difference in variance predicted in unaffected vs. affected subjects, suggesting that both reflect the genotype-phenotype association indexed by the ASD-PRS. It is likely that the study design resulted in individuals without a diagnosis being enriched for PRS load relative to true controls. Further, it was notable not only that there was an interaction effect of sex; but more specifically, that the association of ASD-PRS with QAT was the strongest in affected females, indicating that ASD-PRS is acting differently on QAT in males vs. females.

Our results emphasize the importance of considering and incorporating trait variation in (a) the general population and (b) unaffected members of ASD-affected families to improve
understanding of the genetic architecture of ASD and risk prediction. Given the mounting evidence that clinically diagnosed ASD is the extreme expression of one or more heritable quantitative traits (at least as it pertains to common variant genetic risk), a failure to capture subclinical variation in QAT and related phenotypes in individuals classified as controls may confound efforts to identify genetic variants associated with ASD. Moreover, greater specification of polygenic risk among individuals with rare genetic disorders may improve efforts to predict variation in expression of deleterious variants and thereby disease severity in affected individuals (Finucane et al., 2016)

In light of these complexities, moving away from categorical classification of disease and employing quantitative approaches to understanding the causal origins of ASD stands to reveal a more complete picture of the mechanisms of inheritance in ASD. Uncovering the specific polygenic contributors to ASD is already beginning to reveal a molecular basis for cross-disorder aggregation in families. Tracing the effects of polygenic risk to specific early neurodevelopmental precursors of ASD may identify profiles of allelic variation that map to disparate developmental liabilities, for which specific combinations or permutations give rise to autism (Constantino, 2018), and according to which an individual might be typed for personalized approaches to prevention or therapy. Furthermore, the inclusion of SNPs which are primarily associated with other disorders and traits can actually improve the predictive ability of the ASD-PRS (e.g., SNPs implicated in schizophrenia, depression, IQ) (Grove et al., 2019). Therefore, characterizing patients and controls with respect to inherited traits that index genetic liability to ASD—whether or not those traits are specific to ASD—is expected to encompass more of the variance of the condition than can be identified by simply contrasting the genetic profiles of individuals with and without the diagnosis of ASD. Finally, quantitative approaches
may also provide a means by which to classify more homogeneous subgroups of patients reflecting underlying biology, thereby allowing the pursuit of underlying neural mechanisms and pathways to be conducted with higher precision.

An important future direction suggested by these results is to determine whether the predictive power of polygenic risk signals for autism might be enhanced by including cohorts of unaffected individuals quantitatively phenotyped for variation in QAT in GWAS discovery sets.

A recently-identified nuance of quantitative phenotyping in autism is that the measurements may be significantly less heritable as the threshold for clinical-level affectation is approached: quantitative trait scores in the general population exhibit heritability estimates on the order of 0.85 compared to 0.25 for quantitative measurements of symptom severity (across the wide range of affectation) among clinically-diagnosed individuals (Castelbaum et al., 2019). A possible implication, then, would be a paradoxical improvement in statistical power for genotype-phenotype association in quantitative trait analyses when sampling from populations unaffected by clinical-level aggregation for the trait of interest.

Limitations
As with all complex diseases and traits, a limitation of this study is the proportion of variance explained by the discovery GWAS, likely due at least in part to the small sample size. The largest ASD GWAS (iPSYCH-PGC) is still relatively small in comparison to GWAS of other traits, diseases, and psychiatric disorders; we intentionally limited the total available discovery set by exclusion of the PGC sample due to its inclusion of some AGRE subjects and the consequent inability to specify with certainty a PGC subset devoid of overlap with the families in our AGRE replication cohort. As ASD GWAS sample sizes increase, the predictive power of PRSs is expected to improve, and the estimation of common variant effect sizes to become more
precise. Future studies using a PRS based on weights from a larger GWAS sample is likely to capture even more of the variance in QAT and related, overlapping traits.

**Conclusions**
In conclusion, we demonstrated for the first time that a PRS for clinical ASD is associated with variation in QAT above and below the threshold for diagnosis, using a quantitative phenotype that is known to specifically index familial liability for autism. This was observed in a sample enriched for the inheritance of clinical autistic syndromes. Our study provides a critical replication of an observed polygenic risk signal for ASD in an independent sample, and supports the contribution of common polygenic risk to the aggregation of clinical and sub-clinical autistic traits in the population. These findings underscore the importance of employing quantitative approaches in future genetic studies of ASD and other neuropsychiatric disorders, specifically incorporating the measurement of quantitative variation in behavioral traits that exhibit genetic overlap with autism, not only among probands, but among unaffected relatives and controls.

4.5 Methods
4.5.1 Study Cohorts
**Discovery GWAS**
This study’s discovery dataset is a previously-published GWAS of ASD (iPSYCH) (Gandal et al., 2018) based on a Danish nationwide population-based cohort (Pedersen et al., 2018) including individuals born in Denmark between 1981 and 2005 and diagnosed with ASD prior to 2014. Cases comprised subjects with ASD as the only ascertained diagnosis (classified in accordance with the International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10)) (N=8,605 of 12,371 possible ASD cases; ICD codes F84.0, F84.1, F84.5, F84.8 and/or F84.9), and controls (N=19,526) were a randomly-selected subset of the cohort without diagnosis (ICD F00-F99). All individuals were of European ancestry.
iPSYCH was chosen as the source GWAS over the combined iPSYCH/Psychiatric Genomics Consortium (PGC) meta-analysis GWAS to avoid overlap with the AGRE target data which is included in the PGC dataset.

**AGRE Cohort**
This study’s target dataset was drawn from Autism Speaks’ Autism Genetic Resource Exchange (AGRE), the largest private repository of genetic and phenotype data of families with ASD (Lajonchere and Consortium, 2010). The sampling bias of the collection is inclusion of multiplex families. Individuals with known chromosomal or neurogenetic abnormalities (e.g., DiGeorge, Fragile X), as well as non-verbal subjects were excluded. The AGRE dataset comprised 1491 individuals (206 of whom were parents) in 663 families (ages ranged from 1.2 to 79 years); 1459 subjects contributed to the analysis of ASD diagnosis (we did not have data regarding diagnosis for 32 subjects). Section 4.5.3 provides further detail on the process by which we ensured ancestral genetic homogeneity to avoid population stratification, restricting individuals in this target dataset to those of European ancestry (which is that of our discovery dataset).

**4.5.2 Phenotypes**

**Social Responsiveness Scale (SRS)**
The quantitative trait measure utilized is the Social Responsiveness Scale (SRS) (Constantino and Gruber, 2012). The SRS is a 65-item measure of reciprocal social behavior, deficits in which are characterized as quantitative autistic traits (QAT). The SRS capitalizes on observations of individuals in naturalistic social contexts, by parent and/or teacher report. Its internal consistency is very high ($\alpha = .95$), and it distinguishes ASD-affected individuals from controls with a Cohen’s $d$ effect size of ~2.7 and from individuals with other psychiatric conditions with an effect size of $d = 1$ (Constantino et al., 2000; Constantino and Gruber, 2012). It characterizes variation in the two Diagnostic and Statistical Manual of Mental disorders, 5th edition (DSM-5)
core domains of social communication and interaction and restrictive interests and repetitive behaviors. Reciprocal social behavior, as measured by the SRS, is continuously distributed in the general population and highly heritable throughout the range observed from unaffected to sub-clinically affected to fully ASD-affected individuals.

Individuals’ scores on the SRS were obtained from the Internet System for Assessing Autistic Children (ISAAC) database (https://www.autismtools.org/). When longitudinal scores were available, the earliest assessment was employed. Total raw teacher report scores were prioritized, with parental report scores used when teacher scores were unavailable, as has been done previously (Lowe et al., 2015). Typical norming and adjustment by sex was not done in order to stratify by sex and test interactions with sex in downstream correlation analysis and regression models.

**ASD Diagnosis**
Diagnoses of subjects with ASD were established with both the Autism Diagnostic Interview-Revised (ADI-R) (Rutter et al., 2003) and Autism Diagnostic Observation Schedule (ADOS) (Lord et al., 2001), which are the current gold standard instruments for diagnosing individuals with ASD. Parents not in the AGRE pedigree dataset were counted as unaffected.

**4.5.3 Genotyping, Quality Control, and Imputation**
Genotyping, quality control, and imputation for the AGRE dataset have been described elsewhere. (Lowe et al., 2015) The original dataset comprised 1) 4444 subjects of European ancestry in 943 families (Wang et al., 2009) which were genotyped on the Illumina HumanHap550 BeadChip; and 2) 2108 additional subjects from the University of California Los Angeles (UCLA) Neuroscience Genomics Core genotyped on the Illumina Omni-1 Quad and Omni-2.5-8 platforms. Quality control exclusion filters for the first cohort were the following: subject call rate < 0.95; SNP call rate < 0.95; minor allele frequency (MAF) < 0.05; HWE<
0.001; pedigree discrepancies resolved; monozygotic (MZ) twins removed; principal components based analysis for European ancestry. Quality Control filters for the UCLA cohort were the following: subject call rate < 0.95; SNP call rate < 0.95; MAF < 0.01; Hardy Weinberg Equilibrium (HWE)< 1E-7; pedigree discrepancies resolved; MZ twins removed. Genotype imputation was performed separately on each dataset using IMPUTE2 and a cosmopolitan reference panel from the 1,000 Genomes Project. Filters for exclusion were the following: IMPUTE2 info score <0.5; SNP call rate<0.95; MAF<0.01; and HWE<1E-7. The final imputed dataset consisted of a multi-ancestry cohort of 6552 subjects (4007 males, 2545 females) with 5,814,564 variants.

**Genetic Ancestry Principal Components Analysis**
In order to create an ancestrally homogenous sample and reduce false positives generated by population stratification, we performed an ancestral principal components analysis (PCA) using Eigensoft software (Price et al., 2006). We employed 1000 Genomes Phase 3 (1KG) European (EUR) reference panel, projected the resulting eigenvectors onto AGRE subjects with both SRS and GWAS data (N=2303), resulting in factor scores for the first three principal components (PC1 – PC3 ). Using standard Eigensoft protocol, this analysis yielded a sample of 1491 subjects. PC1 – PC3 were included in all regression models to control for bias due to subpopulations within EUR ancestry.

### 4.5.3 Quantification and Statistical Analysis

**Calculation of PRS**
To calculate the ASD-PRS, we used the SNP effect sizes estimated for common variants from the previously-published iPSYCH GWAS of ASD in 28,131 individuals (Pedersen et al., 2018). We then took the intersection of variants from the ASD GWAS dataset that overlapped with the imputed variants from our AGRE dataset. Next, we reduced the list of intersecting variants to an
independent set by performing LD-clumping ($r^2 < 0.15$ within 500kb from the most significant variant in each locus) using PLINK (Purcell et al., 2007) and a 1000 Genomes Phase 3 cosmopolitan reference panel (1000 Genomes, 2015), resulting in N=172,291 SNPs. To determine what set of SNPs to include in a manner to avoid overfitting, we used the $p$-value of 0.1 as our a priori threshold based on our independent source dataset, the iPSYCH GWAS, as implemented in PRSice 2.1.6 software (Euesden et al., 2015) The PRS was calculated for each individual as the number of the effect alleles at each SNP in the AGRE dataset weighted by the betas from the discovery iPSYCH GWAS at the same SNP. The PRS was standardized to mean=0, standard deviation (SD)=1 in order to interpret the regression coefficient as the change in SRS score for one SD increase in PRS.

**Computation of Odds Ratios**
We grouped subjects into quartiles of PRS scores and modeled ASD diagnosis in a logistic regression model. We estimated odds ratios (OR) for ASD in Quartiles 2 through 4, conducting three separate tests of each quartile, as compared to the first (lowest) quartile (<25% PRS). Using generalized estimating equations (GEE) (as implemented in SAS Proc Genmod), the model was adjusted for PC1 – PC3, sex, and family variance. We also computed a Cochran-Armitage test for trend, a chi-squared test for a linear trend across the four quartiles to test for dose response.

**Correlation Analyses**
As a preliminary step and for purposes of data visualization, unadjusted Pearson correlations between SRS and PRS were computed in the total sample and stratified by sex and affected status. Graphs were produced with R version 3.4.1 Copyright (C) 2017 (R Computing, 2017).

**Linear Mixed Model Regression**
We employed a multiple linear mixed model (LMM) regression analysis because our sample comprised multi-level data in which subjects existed in groups (i.e., families), where the SRS phenotype is expected to correlate within families as well as between families. First, to account
for relatedness in the AGRE family sample, we calculated a genetic relatedness matrix (GRM) using Plink 1.90 (Purcell et al., 2007). The GRM was then entered into all models as a random effect, implemented in SAS Proc Mixed (SAS Inc., 2015), to adjust the residual variance for relatedness. Age was not included in our modeling as a covariate because SRS scores are known to be invariant to age effects (Wagner 2019). For our primary model estimating the variance of QAT explained by PRS (Model 1), the fixed effects comprised PC1 – PC3, sex, and PRS. We computed three other models in which we tested the interactions of PRS*sex (Model 2), PRS*ASD (Model 3), and PRS*sex*ASD (Model 4). As these are technically gene x environment (G x E) interactions, to properly control for confounders, we included GRM, PC1 – PC3, and sex as above, but also included covariate x environment and covariate x gene interaction terms into the models (Keller 2014) (see Table 4 for a breakdown of all interactions included in the modeling).

To estimate the variance explained, we fit a series of full and reduced models, as outlined in Table 4, and from these multi-level models, we then calculated a conditional pseudo-R² using SAS mixed fit, as traditional R² statistics are not available for LMM due to the random effect for GRM (Singer and Willett, 2003). We reduced the full model by eliminating the fixed effect or interaction term of interest and measured the difference in pseudo-R² between the full model and this reduced model. This change in R² (ΔR²) serves as our estimate of the variance explained by the fixed effect (PRS) or interaction term.

### 4.6 Acknowledgements

Supported by ACE Network grants R01 MH081754 and R01 MH100027 and ACE Center grants P50 HD055784 and R01 MH071425 to Dr. Geschwind; by grant R01 HD042541 and IDDRC Center grant P30 HD062171 to Dr. Constantino; and by NIMH grant F31 MH093086 to Dr.
Werling. The Autism Genetic Resource Exchange is a program of Autism Speaks and was supported by NIMH grant U24 MH081810.

We gratefully acknowledge the parents and families participating in AGRE for their participation in this research effort. Without their willingness to participate, this research would not have been possible, and we greatly appreciate their dedication and contribution to deepening our understanding of ASD.

We also acknowledge the assistance with data acquisition provided by Yi Zhang and Joshua Page.

This study was approved by the institutional review boards of the Autism Genetic Resource Exchange (AGRE), the University of California, Los Angeles (UCLA), and Washington University in St. Louis. All subjects provided informed consent for their data to be employed in this research.
4.7 Figures

**Figure 1: Distribution of relative frequency of PRS scores, stratified by sex.** A density plot of the relative frequency of PRS scores, stratified by sex.
Figure 2: Distribution of relative frequency of SRS scores, stratified by sex. A density plot of the relative frequency of SRS scores, stratified by sex.
Figure 3: Distribution of relative frequency of PRS scores, stratified by affected status. A density plot of the relative frequency of PRS scores, stratified by affectation status; based on n=1459 subjects with a determination of affectation status.
Figure 4: Distribution of relative frequency of SRS scores, stratified by affected status. A density plot of the relative frequency of SRS scores, stratified by affection status; based on n=1459 subjects with a determination of affected status.
Figure 5: Correlation of PRS with SRS scores, stratified by sex.
Figure 6: Correlation of PRS with SRS scores, stratified by affected status. Based on n=1459 subjects with a determination of affected status.
Figure 7: Correlation of PRS with SRS scores, stratified by sex and affected status. Based on n=1459 subjects with a determination of affected status.
Table 1: AGRE sample characteristics

<table>
<thead>
<tr>
<th>Group</th>
<th>N*</th>
<th>Mean Age (95% CI)</th>
<th>N Sibs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1491</td>
<td>14.41 (13.80, 15.02)</td>
<td>1285 (86.18)</td>
</tr>
<tr>
<td>Females</td>
<td>521</td>
<td>16.69 (15.56, 17.83)</td>
<td>415 (79.65)</td>
</tr>
<tr>
<td>Males</td>
<td>970</td>
<td>13.19 (12.49, 13.88)</td>
<td>870 (89.69)</td>
</tr>
<tr>
<td>Unaffected</td>
<td>538</td>
<td>22.80 (21.43, 24.17)</td>
<td>332 (61.71)</td>
</tr>
<tr>
<td>ASD</td>
<td>921</td>
<td>9.74 (9.49, 9.99)</td>
<td>921 (100.00)</td>
</tr>
<tr>
<td>Females: Unaffected</td>
<td>299</td>
<td>21.84 (20.12, 23.57)</td>
<td>193 (64.55)</td>
</tr>
<tr>
<td>Females: ASD</td>
<td>215</td>
<td>9.80 (9.27, 10.34)</td>
<td>215 (100.00)</td>
</tr>
<tr>
<td>Males: Unaffected</td>
<td>239</td>
<td>24.00 (21.79, 26.21)</td>
<td>139 (58.16)</td>
</tr>
<tr>
<td>Males: ASD</td>
<td>706</td>
<td>9.72 (9.44, 10.00)</td>
<td>706 (100.00)</td>
</tr>
</tbody>
</table>

Sample characteristics, broken down by sex and ASD status.
Table 2: Mean PRS scores

<table>
<thead>
<tr>
<th>Group</th>
<th>N*</th>
<th>Mean PRS (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1491</td>
<td>0.00 (-0.05, 0.05)</td>
</tr>
<tr>
<td>Females</td>
<td>522</td>
<td>-0.03 (-0.11, 0.06)</td>
</tr>
<tr>
<td>Males</td>
<td>970</td>
<td>0.01 (-0.05, 0.08)</td>
</tr>
<tr>
<td>Unaffected</td>
<td>538</td>
<td>-0.08 (-0.16, 0.01)</td>
</tr>
<tr>
<td>ASD</td>
<td>921</td>
<td>0.05 (-0.02, 0.12)</td>
</tr>
<tr>
<td>Females: Unaffected</td>
<td>299</td>
<td>-0.07 (-0.19, 0.04)</td>
</tr>
<tr>
<td>Females: ASD</td>
<td>215</td>
<td>0.05 (-0.08, 0.19)</td>
</tr>
<tr>
<td>Males: Unaffected</td>
<td>239</td>
<td>-0.08 (-0.20, 0.05)</td>
</tr>
<tr>
<td>Males: ASD</td>
<td>706</td>
<td>0.05 (-0.02, 0.12)</td>
</tr>
</tbody>
</table>

Unadjusted mean PRS scores, stratified by sex and ASD; *N=1491 subjects contributed to sex strata; n=1459 subjects contributed to strata with ASD.
Table 3: Odds ratios of ASD across quartiles of polygenic load

<table>
<thead>
<tr>
<th>Odds Ratio (95% CI), p-value</th>
<th>&lt;25th Percentile</th>
<th>25th-49th Percentile</th>
<th>50th-75th Percentile</th>
<th>&gt;75th Percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0 (Reference)</td>
<td>1.195 (0.886, 1.611); p=0.243</td>
<td>1.269 (0.931, 1.730); p=0.132</td>
<td>1.450 (1.084, 1.941); p=0.012</td>
<td></td>
</tr>
</tbody>
</table>

Table 3 outlines the OR, 95% CI, and p-value for affected individuals across quartiles of PRS load.
Table 4: Linear mixed models of fixed and random effects

**FULL MODELS**

<table>
<thead>
<tr>
<th>Model</th>
<th>Effects in Full Model</th>
<th>$R^2$</th>
<th>Full Model $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>PC1 – PC3</td>
<td>0.0015</td>
<td>0.511</td>
</tr>
<tr>
<td>B</td>
<td>PC1 – PC3 + sex</td>
<td>0.0989</td>
<td>$&lt;1E-15^*$</td>
</tr>
<tr>
<td>C</td>
<td>PC1 – PC3 + sex + PRS</td>
<td>0.1023</td>
<td>$&lt;1E-15^*$</td>
</tr>
<tr>
<td>D</td>
<td>PC1 – PC3 + sex + ASD + PRS + ASD<em>sex + PRS</em>(PC1 – PC3) + PRS*ASD</td>
<td>0.5782</td>
<td>$&lt;1E-15^*$</td>
</tr>
<tr>
<td>E</td>
<td>PC1 – PC3 + sex + ASD + PRS + ASD<em>sex + PRS</em>(PC1 – PC3) + PRS*sex</td>
<td>0.5795</td>
<td>$&lt;1E-15^*$</td>
</tr>
<tr>
<td>F</td>
<td>PC1 – PC3 + sex + ASD + PRS + ASD*(PC1 – PC3) + ASD<em>sex + PRS</em>(PC1 – PC3) + PRS<em>sex + PRS</em>ASD</td>
<td>0.5796</td>
<td>$&lt;1E-15^*$</td>
</tr>
<tr>
<td>G</td>
<td>PC1 – PC3 + sex + ASD + PRS + ASD*(PC1 – PC3) + ASD<em>sex + PRS</em>(PC1 – PC3) + PRS<em>sex + PRS</em>ASD + PRS<em>sex</em>ASD</td>
<td>0.5809</td>
<td>$&lt;1E-15^*$</td>
</tr>
</tbody>
</table>

**REDUCED MODELS**

<table>
<thead>
<tr>
<th>Model</th>
<th>Effect</th>
<th>Effect tested in $\Delta R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$ $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>PRS</td>
<td>C vs. B (full model reduced by PRS)</td>
<td>0.0034</td>
<td>0.017*</td>
</tr>
<tr>
<td>Model 2</td>
<td>PRS * Sex</td>
<td>F vs. D (full model reduced by PRS*Sex)</td>
<td>0.0014</td>
<td>0.027*</td>
</tr>
<tr>
<td>Model 3</td>
<td>PRS * ASD</td>
<td>F vs. E (full model reduced by PRS*ASD)</td>
<td>0.0001</td>
<td>0.511</td>
</tr>
<tr>
<td>Model 4</td>
<td>PRS * Sex * ASD</td>
<td>G vs. F (full model reduced by PRS<em>Sex</em>ASD)</td>
<td>0.0013</td>
<td>0.031*</td>
</tr>
</tbody>
</table>

The “Full Models” portion of this table details the variance explained (pseudo-$R^2$) for each full model and its accompanying $p$-value. The “Reduced Models” portion outlines what comprises each model and provides the delta $R^2$ of the effect which was isolated by the model reduction. An asterisk (*) indicates a statistically significant $p$-value at $p = <0.05$; $<1E-15$ indicates smaller value than machine precision.
4.8 References


Chapter 5: Conclusions

5.1 Summary
In this dissertation, I have used three approaches to attempt a refinement of our understanding of causal factors in the development of psychopathology. First, in my examination of environmental influences on neuropsychiatric outcomes, I showed that presumed toxic stressors (i.e., causal risk factors for psychopathology) did not correlate with one another in a sample of impoverished children, nor did they predict child adaptive or behavioral outcomes, as measured by the CBCL. Further, despite the high prevalence of official-report child maltreatment in our sample, we observed relatively low rates of internalizing and externalizing behavior. Also, children in families for whom there existed official documentation of child maltreatment did not demonstrate greater levels of psychopathology than those in families without official reports of maltreatment. It is important to note that this study capitalized on a unique sample of both urban and rural impoverished families enrolled in EHS, one enriched for what are considered to be some of the most deleterious toxic stressors, including poverty and child maltreatment. The study demonstrated the challenges of measuring both environmental risk factors and early childhood psychopathology and emphasized the need for future genetically-informative studies to more fully characterize risk and protective influences on child behavioral and adaptive outcomes.

In the second and third studies, I shifted my attention to a neuropsychiatric disorder whose origin is nearly entirely genetic: ASD, a more scientifically tractable disorder. Since clinically diagnosed ASD is the pathological tail of a continuous distribution of deficits in reciprocal social behavior, I believed that using a quantitative measure of AVR was important in the long-term goal of predicting risk for this disorder in early childhood. Yet our measure of
AVR, the SRS (the only quantitative measure of AVR which encompasses the entire range of variation in AVR), had not yet had its longitudinal stability assessed—a critical first step if the SRS were to be employed in risk prediction efforts. Therefore, in the second study, I analyzed the stability of the SRS in a longitudinal dataset of subjects with and without ASD, as well as individuals with and without other psychiatric conditions. The stability of the SRS throughout its entire distribution was very high from early childhood to adulthood. Furthermore, the course of individuals’ scores from early childhood to young adulthood reflected distinct separations in symptom burden for deficiency in AVR, overwhelmingly differentiating controls from individuals with ASD.

Having established the stability of the SRS and therefore its utility as a tool in the prediction of ASD risk, in the third study I examined if variance in QAT, as measured by the SRS, could be predicted by ASD-PRS. We were able to significantly explain a modest amount of variance in QAT, critically in both affected and unaffected subjects. Interestingly, there was an interaction effect of PRS*sex*ASD, indicating that the PRS is operating differently in affected females. Further work will need to be conducted to explore this relationship, as it is of particular importance, given the sex disparity in ASD.

5.2 Future Directions
The work outlined in the first portion of this dissertation suggests that current characterizations of environmental risk factors as well as measurements of early manifestations of child psychopathology likely need refinement. Ascertainment of objective measures of risk and outcome will be critical for efforts going forward that seek to identify causal environmental factors for the development of psychopathology. Further, due to gene-environment correlation of environmental variables with behavioral and adaptive outcomes of interest, genetically
informative designs will be essential to identifying modifiable risk factors for psychopathology. More refined specification of salient deleterious exposures—as well as indices of inherited sensitivity and resilience—along with the development of more sensitive measures, will improve opportunities for strategic targeting of preventive interventions to reduce enduring impairment in behavioral adaptation among infants and children at highest risk.

In the second part of this dissertation, we established the validity of using ASD-PRS to predict AVR with current available data (a small discovery GWAS and a small target sample), and thus have laid the groundwork to conduct a GWAS in a larger dataset on the quantitative phenotype itself. Future research by our group and others will examine if a GWAS of the SRS will provide greater SNP-heritability estimates than GWAS based on categorical diagnosis, and ultimately greater predictive power for autism risk. The long-term goal will be to employ the results of GWAS of the SRS along with those of other related phenotypes and endophenotypes (e.g., intelligence / educational attainment, executive function, ADHD, schizophrenia, depression) to create personalized risk profiles for individuals.

5.3 Significance
In this dissertation, I explored causal factors in the development of neuropsychiatric conditions. Specifically, in Chapter 2, I examined putative environmental risk factors for internalizing and externalizing behavior, as broadly measured by the CBCL. In Chapters 3 and 4, I shifted my attention to autism, focused on determining the stability of the SRS and then its genetic predictors. All three of these studies have important implications for rehabilitation efforts aimed at preventing or intervening in neurodevelopmental conditions.
5.3.1 The criticality of approaches to measuring environmental risk and outcomes
As demonstrated in Chapter 2, current approaches to measuring environmental risk and outcome variables for psychopathology would benefit from greater precision, more objective measurement, and genetically-informative study designs. These findings suggest that caution should be exercised in efforts to ameliorate presumed environmental risk factors. Variables assumed to be causal which are not (or which account for a very small percentage of the variance in the phenotype) will not be predictive of the development of neuropsychiatric conditions. Thus, rehabilitation scientists should incorporate a genetic perspective in their determinations of what intervention approaches to employ. Further, this study raised questions about the measurement of early childhood behavior and adaptive functioning, and rehabilitation scientists seeking to determine if their interventions are effective may consider developing more sensitive measures of these constructs.

5.3.2 Why longitudinal stability matters
ASD, characterized primarily by early-onset deficits in reciprocal social behavior, can be one of the most severe and enduring of all neuropsychiatric syndromes, with most ASD diagnoses being lifelong. However, it should be emphasized that ASD is not always severe (as the word spectrum in the disorder’s name denotes) or accompanied by intellectual disability in many autistic individuals; in fact, many value their ASD, seeing it as an integral part of their identity and a difference to be celebrated (Kapp, Gillespie-Lynch, Sherman, and Hutman 2013). However, many others with ASD (including individuals who would often be considered or self-describe as “high-functioning”) express desire for a cure, amelioration of impairments, and reduction of symptoms (Bagatelle 2010). These differing perspectives in the ASD community are legitimate, and illustrate the need for a multiplicity of research agendas and theoretical approaches in response.
Yet regardless of one’s personal perspective on ASD or the many lenses through which the diversity and disability can be legitimately viewed, reciprocal social behavior is a fundamental axis of human social development, an evolutionarily programmed, biological, and highly heritable characteristic which functions as a core determinant of human social variation—and thus warrants our investigation and understanding of its development. Because successful social engagement is critical for their survival, infants’ brains are iteratively shaped towards socialization, with reciprocal social behavior functioning as the foundation on which an individual’s social and communicative development rests. Throughout life, variations in reciprocal social behavior play an integral role in typical human social development, predicting peer relationships and inclusion in social networks. More specifically, and of relevance to the role of added genetic influences in inherited ASD susceptibility, across cultures there is evidence of strong preferential mating for variation in reciprocal social behavior. Further, when present, autism-related variation in reciprocal social behavior exacerbates the severity of nearly any psychiatric condition with which it co-occurs, which has clinical implications for enduring behavioral patterns in individuals with other diagnoses. If the longitudinal trajectories of autism-related variation in reciprocal social behavior are indeed stable throughout the life course, it is conceivable that measuring it in early childhood could make strong predictions over the course of an individual’s life, particularly in the areas of social interactions and other neurodevelopmental conditions. Therefore, the work I did in Chapter 3 to establish the longitudinal stability of an ecologically-valid, naturalistic quantitative measure of social functioning, the SRS, has important implications for rehabilitation efforts. Since the SRS captures real-world outcomes essential for individuals’ social and occupational participation, having this early childhood measure of enduring traits allows for the identification of young
children at risk who would benefit from early intervention efforts to alter the life course of their disability.

### 5.3.3 The importance of identifying genetic risk

The significance of the findings of Chapter 4 resides in the study’s employment of genetic data to ask questions of direct relevance to rehabilitation. PRS can be employed for the probabilistic estimation of disorder risk, identification of disease trajectories, generation of mechanistic insight into disorders and related phenotypes, stratification for clinical trials, and selection of individualized treatments targeted at specific underlying pathophysiology and developmental disease trajectories—all priorities of NIH’s Research Plan for Rehabilitation and critical for improving the effectiveness of rehabilitation prevention and intervention efforts. Using genetic data for the purpose of capturing real-world outcomes apart from diagnosis alone is an emerging area which already has a preliminary basis of support: recent studies have found that individuals without a diagnosis who had CNVs associated with NDDs have diminished educational and occupational attainment, a reduced ability to earn an income, a decrease in performance IQ, and higher social deprivation in middle age (Steffansson et al. 2014; Kendall et al., 2017; Mannik et al. 2015; Kendall et al. 2019). The ability to use PRS to predict quantitative variation in these real-world functional outcomes—critically, in individuals without a diagnosis and who would otherwise be missed by current approaches—is a significant advance upon which rehabilitation-driven science can capitalize in an innovative way. The genetic architecture of psychiatric disorders is not merely an academic exercise: future study design, disease classification, therapeutic approach, and early intervention based on pathophysiology all rest on an accurate understanding of the disease architecture.
5.3.4 Implications for rehabilitation science

The National Institutes of Health (NIH) Rehabilitation Research Plan is a guide for scientists in the field of rehabilitation science and communicates not only funding priorities, but also what the NIH believes to be the most important foci for the field of rehabilitation science. According to this plan, “NIH sponsors multiple programs to advance understanding of the fundamental biological…mechanisms that underlie disease. For the purpose of rehabilitation research, this understanding would encompass genomic . . . contributions” (Eunice 2016). One method in particular of developing our understanding of genomic contributors to disease is through a precision medicine approach. The research plan explains that “precision medicine is an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle” (Eunice 2016). The document goes on to specify that, “although researchers have made some advances in precision medicine, the practice is not in use for most diseases or conditions, nor has it been widely applied for use in rehabilitation medicine” (Eunice 2016). This gap in rehabilitation science presented itself as an ideal target for my research, particularly given the highly genetic etiology of ASD.

This research also aligns with current emphases of the World Health Organization (WHO) and positional statements in the field of occupational therapy. Disorder prevention and early intervention are two themes of particular relevance to ASD that are outlined in WHO’s Mental Health Action Plan (World 2013), American Occupational Therapy Association’s Occupational Therapy Practice Framework: Domain and Process (American 2008) statements on the role of occupational therapy, and in editorials in the field’s flagship journal (Hildenbrand & Lamb 2013). Because WHO has identified mental illness as the top cause of disability worldwide, its latest Mental Health Action Plan has as one of its primary four objectives the “implementation of strategies for . . . prevention” of mental health disorders (World 2013). These
emphases on prevention and early intervention are a direct complement to the NIH Rehabilitation Research plan’s objectives to prioritize precision medicine. Although early diagnosis and intervention strategies resulting from genetic science may be only in their nascent stages, we nonetheless have a responsibility to employ genetic science to this end. Currently, we can already employ genetic risk as a control variable when testing an intervention; in non-psychiatric conditions (e.g., cardiovascular), genetic information is employed as screening biomarkers to identify individuals at highest risk and target them to receive the greatest resources in prevention and intervention—ideally, within the next decade, that will become a reality for at least some individuals with certain psychiatric disorders. Finally, a core goal of rehabilitation science is to understand the biological mechanisms of disorder. For genetic disorders, such as ASD, starting with genetics is the foundation upon which we can build an understanding of biological mechanisms underlying the disorder, with the goal of employing this knowledge in the development of rational therapeutics.
5.4 References


Eunice Kennedy Shriver National Institute of Child Health and Human Development and the NIH Medical Rehabilitation Coordinating Committee. (2016). NIH Research Plan on Rehabilitation.


