



Published in final edited form as:

Dev Psychopathol. 2016 May ; 28(2): 415–431. doi:10.1017/S095457941500053X.

Neighborhood × Serotonin Transporter Linked Polymorphic Region (*5-HTTLPR*) interactions for substance use from ages 10 to 24 years using a harmonized data set of African American children

MICHAEL WINDLE^a, STEVEN M. KOGAN^b, SUNBOK LEE^b, YI-FU CHEN^b, KARLO MANKIT LEI^b, GENE H. BRODY^b, STEVEN R. H. BEACH^b, and TIANYI YU^b

^aEmory University

^bUniversity of Georgia

Abstract

This study investigated the influences of neighborhood factors (residential stability and neighborhood disadvantage) and variants of the serotonin transporter linked polymorphic region (*5-HTTLPR*) genotype on the development of substance use among African American children aged 10–24 years. To accomplish this, a harmonized data set of five longitudinal studies was created via pooling overlapping age cohorts to establish a database with 2,689 children and 12,474 data points to span ages 10–24 years. A description of steps used in the development of the harmonized data set is provided, including how issues such as the measurement equivalence of constructs were addressed. A sequence of multilevel models was specified to evaluate Gene × Environment effects on growth of substance use across time. Findings indicated that residential instability was associated with higher levels and a steeper gradient of growth in substance use across time. The inclusion of the *5-HTTLPR* genotype provided greater precision to the relationships in that higher residential instability, in conjunction with the risk variant of *5-HTTLPR* (i.e., the short allele), was associated with the highest level and steepest gradient of growth in substance use across ages 10–24 years. The findings demonstrated how the creation of a harmonized data set increased statistical power to test Gene × Environment interactions for an under studied sample.

With some notable exceptions (Tarter & Vanyukov, 1994; Zucker, 2006), most etiological models of substance use during adolescence have been based on a governing assumption that that the substantial and statistically significant associations between a wide range of environmental risk factors and various risk behavior outcomes represent environmentally mediated effects. Mounting evidence over the past few decades has demonstrated that statistical associations of some environmental risk factors with substance use may be partially genetically mediated or conditioned by genetic differences (Brody, Beach, et al., 2013). These genetically influenced effects may arise because individuals' behaviors shape and select their environments and influence others' responses to them; this process is termed

gene–environment correlation. Another form of gene–environment interplay, Gene \times Environment interaction ($G \times E$), occurs when genetic variation alters an individual's sensitivity to specific environmental events (e.g., maltreatment in the development of depression) or when environmental events (e.g., exposure to life stress) exert differential control over genetic influences (Kendler & Eaves, 1986). As noted by Shanahan and Hofer (2005) and others, $G \times E$ effects imply that genetic variation brings about differences in individuals' resilience or vulnerability to the environmental causes of many disorders, including drug abuse.

Although studies are increasingly considering genetic factors in the development of adolescent and young adult substance use there are critical limitations in the extant literature. There has been a dearth of research studies on African American youths' genetic characteristics and substance use trajectories that are adequately powered and include longitudinal data (Espeland et al., 2006; Green et al., 2006; McQuillan, Pan, & Porter, 2006; McQuillan, Porter, Agelli, & Kington, 2003). This is a problem for developmental and public health scientists because some risk mechanisms (e.g., poverty and community disadvantage) disproportionately affect African Americans (Schuster et al., 2012). In addition, across ethnic groups, there is substantial genetic variation at loci known to be important in moderating responses to environmental risk mechanisms. The majority of studies that consider the $G \times E$ interplay involve cross-sectional or short-term longitudinal research designs. Few studies with African Americans have considered substance use trajectories across relatively long intervals (e.g., ages 10–24 years). This is problematic as $G \times E$ effects may be present at one development phase (e.g., adolescence) but not another (middle adulthood), as youth experience changing environmental risk and protective processes across development. The lack of data from preadolescence through the transition to young adulthood is particularly important in investigating African Americans because epidemiological studies have identified important variations in the patterns of drug use across time in African Americans as compared other racial/ethnic groups (Chen & Jacobson, 2012; French, Finkbiner, & Duhamel, 2002; Wallace et al., 2002).

There have been some cross-sectional studies that have examined relationships between the serotonin transporter linked polymorphic region (*5-HTTLPR*) and psychiatric outcomes with African American adults. For example, Odgerel, Talati, Hamilton, Levinson, and Weissman (2013) reported that African Americans had a lower frequency than European Americans of the short allele (0.25:0.43) and the short–short genotype (0.06:0.19), but there was a lack of a significant association for either racial/ethnic group between *5-HTTLPR* and the disorders of major depression or anxiety, or the trait of neuroticism. By contrast, Patkar et al. (2001) reported no significant differences between African Americans and European Americans in the distribution of the *5-HTTLPR* genotype or allele frequencies for a sample of 197 cocaine dependent subjects and 100 control subjects. Xie et al. (2009) examined the interactions among *5-HTTLPR*, childhood adversity, and adult traumatic events on posttraumatic stress disorder for a community sample of 582 European Americans and 670 African Americans who had reported experiencing childhood adversity and/or adult traumatic events. The *5-HTTLPR* genotype significantly interacted with childhood adversity and adult traumatic events to increase risk for posttraumatic stress disorder, especially among those with both types of trauma exposure. The odds ratio was somewhat higher for European Americans

(2.86, 95% confidence interval [CI] = 1.50–5.45, $p = .002$) than for African Americans (1.88, 95% CI = 1.04–3.40, $p = .04$), but both were statistically significant and the increased risk for both racial/ethnic groups was for one or two copies of the short allele compared with long allele homozygotes. Although these studies are mixed with regard to the possibility of some minor differences in genotype and allele frequencies for *5-HTTLPR*, the relationships with psychiatric disorders appear more similar than different across racial/ethnic groups for these studies. Perhaps of more importance to this study and its potential contributions to the literature, none of these studies measured children/adolescents, none measured neighborhood characteristics, none was longitudinal, and none focused in intraindividual growth in substance use across time.

Recent advances in integrated data analysis (IDA; Hussong, Curran, & Bauer, 2013) and modern missing data techniques (Schafer, 1997) provide a unique opportunity to address the challenges related to the longitudinal study of $G \times E$ influences on adolescent substance use. The goals of the present study are twofold. For the first goal, we describe the development of a unique harmonized data set of genetic and environmental variables that was developed by pooling data from five distinct longitudinal studies of African American youth and young adults. This data resource, labeled the Genes Environments and Development Integrated Data set (GEDID), leveraged recent developments in IDA, item response theory (IRT), and missing data estimation to provide a resource for testing $G \times E$ interplay hypotheses. The second goal involved using the GEDID to test a longitudinal hypothesis regarding the moderating influence of the *5-HTTLPR* genotype on the link between exposure to environmental disadvantages and trajectories of substance use from preadolescence to young adulthood.

Drug Use Trends and Consequences Among African American Youth

The African American youth who are the focus of this study reside in small towns and communities in which poverty rates are among the highest in the nation and unemployment rates are above the national average (Dalaker, 2001). Many rural African American families live with severe, chronic economic stress that takes a toll on children, youth, and young adults and increases their risk for substance use during adolescence and young adulthood. Recent data indicate that rural African American youth use substances at rates equal to or exceeding those of African Americans residing in densely populated inner cities (Kogan, Berkel, Chen, Brody, & Murry, 2006; Levine & Coupey, 2003).

Patterns of substance use among African Americans demonstrate considerable departures from other racial/ethnic groups. US national data for adolescents indicate a higher prevalence of alcohol use, heavy episodic drinking, and cigarette use for European Americans and Hispanics relative African Americans and roughly comparable rates of marijuana use across these three racial/ethnic groups by 12th grade (Johnston, O'Malley, Bachman, & Schulenberg, 2013; Kann et al., 2014). Using prospective data from the National Longitudinal Survey of Adolescent Health, Chen and Jacobson (2012) reported that initial rates and rates of growth of substance use for Hispanics and European Americans from early adolescence to early young adulthood were higher than for African Americans. Wallace et al. (2002) have also reported that rural African Americans initiate drug use later

than do European Americans. However, across the transition to young adulthood they experience rapidly escalating rates that ultimately equal and, for several substances, surpass European Americans' rates (French et al., 2002; Galea & Rudenstine, 2005). For example, marijuana use rates at age 15 are lower for African Americans than for other racial/ethnic groups; by age 20, African Americans are more likely than members of other groups to be dependent on marijuana (Reardon & Buka, 2002). Thus, evidence is accumulating that drug use trajectories differ by race; although African Americans start using drugs later than European American adolescents and tend to use less frequently, this difference diminishes by the mid-20s. In terms of problematic use, the difference appears to reverse in the 20s; that is, among those who use substances, the percentage reporting abuse is higher for African Americans. This phenomenon is referred to as the "racial crossover" (Kandel, 1995; Kandel, Schaffran, Hu, & Thomas, 2011).

Developmental Orientation to Adolescent Substance Use

Our research group, along with a large number of child and adolescent substance use investigators, have adopted a developmental psychopathology perspective to investigate both changes in substance use across time and predictors of these changes (Chassin, Sher, Hussong, & Curran, 2013; Schulenberg & Maggs, 2002; Windle, 2010; Windle & Davies, 1999; Zucker, 2006). This perspective was used to guide the hypotheses in the current study. A common feature of this perspective for substance use includes the importance of identifying individual differences in intraindividual change and a recognition of multilevel influences, including genes and individual, social, neighborhood, and cultural factors that may interact with one another and vary in their influence across the life course.

In recent years, the most common data analytic models to investigate intraindividual change and predictors of such changes in the development of child and adolescent substance use have been either growth models using either structural equation modeling or multilevel modeling or latent growth mixture models. For example, Duncan and Duncan (1996) used a second-order multivariate extension of the latent growth model to support a higher order substance use construct that could be modeled as a linear growth model of alcohol, cigarette, and marijuana use among adolescents. Several variables predicted initial levels of substance use (e.g., family status, parent-child conflict, peer encouragement for substance use, parent substance use, age) and several predicted changes (i.e., rate of increase) in substance use across time (e.g., peer encouragement, change in peer encouragement, change in parent-child conflict, age, gender). Chen and Jacobson (2012) used data from the National Adolescent Health Study and reported developmental trajectories of substance use and variations in the intercept and linear rates of growth across gender and racial/ethnic groups.

There is an extensive developmental literature that has used growth in substance use and predictors of substance use that has served to identify which predictors are significant for whom and when (see extensive listing of studies in Sher, Jackson, & Steinley, 2011). These findings have contributed to the larger field of prevention science and have impacted the development, implementation, and evaluation of preventive interventions, for example, with regard to multilevel interventions (Monahan, Egan, Van Horn, Arthur, & Hawkins, 2011; Pentz et al., 1989), with regard to focusing on middle childhood and the early adolescent

periods of the life span prior to the onset and escalation of substance use (Ellickson, Bell, & McGuigan, 1993; Windle et al., 2008), and with regard to considering age-specific booster sessions (e.g., in middle adolescence) to an initial intervention applied in early adolescence (Hawkins, Catalano, & Arthur, 2002).

In the current study we used this developmental framework to pursue the study of intraindividual growth in substance use among African American youth aged 10–24 years. We also embedded this study of intraindividual change in substance use in a multilevel level model that included neighborhood level influences (neighborhood disadvantage, residential stability) as measured via census tract indicators and included an individual level genetic marker: *5-HTLPPR*. We hypothesized a cross-level interaction such that the short allele of *5-HTLPPR* would significantly interact with neighborhood characteristics to predict variation in the acceleration rate of substance use for individuals across time.

Neighborhood Factors and Substance Use

For African Americans, residence in socioeconomically disadvantaged neighborhoods is a robust predictor of substance use. Research conducted on neighborhood factors has generally indicated impaired functioning and poorer health outcomes associated with characteristics such as neighborhood disadvantage and residential instability (Hawkins, Catalano, & Miller, 1992; Hurd, Stoddard, & Zimmerman, 2013; Kawachi & Berkman, 2003; Sampson, Raudenbush, & Earls, 1997). For example, with an adolescent sample Mennis and Mason (2012) reported higher levels of substance use among adolescents in neighborhoods with concentrated disadvantage (i.e., higher rates of unemployment, lower educational attainment, and higher rates of public assistance income). Tucker, Pollard, de la Haye, Kennedy, and Green (2013) reported that higher neighborhood unemployment rates were associated with an earlier onset of marijuana use among a national sample of adolescents. Mechanisms proposed for these relationships between neighborhood factors and substance use are guided by a range of theories, such as social disorganization theory (Shaw & McKay, 1942), which posits that neighborhood features such as concentrated disadvantage and residential instability serve to undermine critical social processes related to neighborhood cohesion and support. Low levels or the absence of these critical neighborhood resources (i.e., cohesion and support) may undermine a range of social control mechanisms such as engaged constructive parenting, deviant peer affiliation processes, academic performance, and self-regulation among youth (Brody et al., 2001; Simons, Simons, Burt, Brody, & Cutrona, 2005).

In discussing the investigation of the *5-HTTLPR* (G) by neighborhood (E) interaction on changes in substance use across time, it is informative to describe two different overarching conceptual models that have been promulgated (Belsky & Pluess, 2009). The first is the *diathesis–stress model* that posits that individual variation in vulnerability characteristics (e.g., temperament, specific genes) interact with levels of stressful exposure to predict psychopathology. For example, in highly vulnerable individuals, lower levels of stress exposure are required to precipitate a disorder; whereas, higher levels of stress exposure are required to precipitate a disorder among individuals who are low in vulnerability. The second is the *differential susceptibility model* that posits that some individual characteristics

(e.g., specific genes) are associated with *plasticity* rather than vulnerability. That is, the expression of specific susceptibility genes (typically referred to as risk genes in the diathesis–stress model) are influenced by the surrounding environment whether it is highly positive or highly negative. Such susceptibility genes may interact with impoverished environments to yield negative outcomes, but they may also interact with enriched environments to yield positive outcomes. Thus, susceptibility genes are plastic and their expression is integrally related to features of the environment. There have been a number of studies supportive of the differential susceptibility model (Beach, Brody, Lei, & Philibert, 2011; Belsky & Pluess, 2009; Brody, Chen, & Beach, 2013) and Roisman et al. (2012) have provided a statistical approach to evaluate both diathesis–stress and differential susceptibility models. Of interest in this study will be the ability to expand the study of diathesis–stress and differential susceptibility models to social disorganization theory (discussed subsequently) and other theories about processes (e.g., collective socialization) associated with neighborhood disadvantage (Brody et al., 2001; Cutrona et al., 2005). We address how *5-HTTLPR* may interact with characteristics of neighborhoods to support one or the other of these models.

Serotonin (*5-HTTLPR*) and Substance Use

A number of recent studies suggest that genetic polymorphisms in key neurotransmitter systems increase youths' susceptibility to substance use by affecting their responses to challenging environments (Brody, Beach, et al., 2013). A well-studied polymorphism that modulates serotonin activity, *5-HTTLPR* is one such candidate (Caspi, Hariri, Holmes, Uher, & Moffitt, 2010). A key regulator of serotonergic neurotransmission, *5-HTTLPR* is localized to 17p13 and consists of 14 exons and a single promoter. The common polymorphism in the promoter region results in two variants, a short and a long allele, with the short allele resulting in lower serotonin transporter availability. The short variant contains 12 copies of a 22bp repeat element and the long variant has 14 copies of the repeat element. The short variant has been reported to be associated with lower basal and transcriptional efficiency of solute carrier level C6, member 4 (*SLC6A4*), resulting in lower serotonin uptake activity compared with the long variant (Heils et al., 1996).

Allelic variation in genes encoding serotonin functioning have been associated with impulsive behavior, novelty seeking, and poor performance on delay of gratification tasks (Auerbach, Faroy, Ebstein, Kahana, & Levine, 2001; Kreek, Nielsen, Butelman, & LaForge, 2005). Furthermore, a broad range of studies using different methodologies (e.g., longitudinal observational, experimental neuroscience, genetic knockouts) across multiple species have been conducted and indicate that genetic variation in the serotonin transporter (*5-HTTLPR*) is integrally involved in emotional regulation and differential susceptibility to stress reactivity (Caspi et al., 2010). Neurobiological findings have supported the role of the amygdala and modifications in the hypothalamic–pituitary–adrenal axis responses to stress as mechanisms involved in linking genetic variation in the serotonin transporter to stress sensitivity and emotional regulation (El Hage, Powell, & Surguladze, 2009; Hariri, 2011). Canli et al. (2006) reported a $G \times E$ interaction in that short allele carriers of *5-HTTLPR* with higher stress levels had higher amygdala resting-state cerebral blood flow than those who were long allele carriers or those with lower levels of stress. Robinson et al. (2013)

have also provided findings supporting how serotonin may modulate the dorsal medial prefrontal amygdala “aversive amplification” circuit to impact negative affective biases such as rumination and preferential bias toward threat-related stimuli (see also Hariri, 2011). These findings are consistent with the notion that short allele carriers, relative to long allele carriers, are more likely to be more hypervigilant, have more affective biases, and be more stress reactive to chronic stress exposures such as residential instability and neighborhood/community disadvantage. Hence, in this study the genetic susceptibility of short allele carriers is posited to interact with environmental neighborhood exposure factors to impact higher rates of growth in substance use across time as children attempt cope with and emotionally regulate behaviors within high demand contexts.

The serotonin transporter *5-HTTLPR* has been studied much more extensively in relation to mood disorders (Hariri, 2011) and to some extent alcohol dependence than to substance use (Feinn, Nellissery, & Krnazler, 2005), although a number of studies have been conducted on *5-HTTLPR* in relation to adolescent and young adult drinking, with mixed findings (Young-Wolff, Enoch, & Prescott, 2011). For example, five-wave prospective findings by van der Zwaluw et al. (2010) indicated that the short allele of *5-HTTLPR* was significantly associated with increases in the rate of growth in adolescent alcohol use. Earlier onset alcohol use was associated with the short allele (Buchmann et al., 2009; Kaufman et al., 2007), as was binge drinking in young adult women (Herman et al., 2005). Contrary to these findings, others have not found a significant association between *5-HTTLPR* and alcohol use (Hopfer et al., 2005) or found that *5-HTTLPR* was significantly associated with alcohol use in young adulthood but not in adolescence (Guo, Wilhelmsen, & Hamilton, 2007). Among the limitations of drawing firm conclusions about the relationship between *5-HTTLPR* and adolescent substance use is that much of this research has adopted a “main effects” versus G × E approach, statistical power has often been limited, and substance use often has been limited to alcohol phenotypes.

Data Harmonization and IDA

The concept of data harmonization has entered prominently in many fields of study, including molecular genetics, cross-national economics, hospital record systems, and more recently, human psychosocial behaviors (Bennett et al., 2011; Fortier et al., 2010; Granda & Blascyk, 2010). Commonalities regarding the need for data harmonization are similar across these fields of study in that there is a need to pool data from different studies or settings to address broader questions such as the generalizability of findings; however, data sets may be heterogeneous with regard to a number of characteristics, such as sampling design, region of the country, populations, and measures that present methodological challenges.

Hussong et al. (2013) refer to the attempt to pool data across studies as IDA, which they describe as a methodological framework to pool studies rather than as a particular data analytic or statistical modeling approach. Some of the major potential advantages of IDA are that pooled data sets may (a) substantially increase statistical power to assess lower frequency dependent variables or to test statistical interactions, (b) extend the age spans for hypothesis testing if pooling multiple longitudinal data sets may be structured within a longitudinal sequential or accelerated research design, and (c) enable the investigation of

study features of different studies (e.g., sampling design, methods of assessment for common constructs) in ways not possible with individual studies. For example, data integration or harmonization for a range of important child and family variables across multiple longitudinal studies may facilitate the testing of more complex research questions across a longer developmental age span about interactions that are sufficiently powered statistically; it is conceivable that none of the individual longitudinal studies may be sufficiently powered to adequately test research questions about substantively interesting statistical interactions, but an integrated, harmonized data set may be well suited for this purpose.

Although the advantages of data integration are relatively straightforward, there are challenges to establishing a database of pooled data sets due to a number of sources of heterogeneity across studies (e.g., sampling procedures and sample differences, variation in measures) that, if not addressed, may limit the findings and therefore the advantages of a harmonized data set (Curran & Hussong, 2009; Hussong et al., 2013). For example, if the different individual studies to be pooled all measured a common construct (e.g., depressive symptoms) but did so with distinct measures of depression, it may be difficult to pool studies to investigate depression. However, if different measures of depression contained some overlapping but not identical items, measurement models may be used to provide a method of developing commensurate measures to facilitate data pooling and subsequent IDA (Curran et al., 2008; Hussong, Flora, Curran, Chassin, & Zucker, 2008). Other issues and examples of potential threats to internal validity for the pooling of individual studies are expanded on in several methodologically oriented articles (Bauer & Hussong, 2009; Curran & Hussong, 2009; Hussong et al., 2013).

Current Study Using the Integrated (Harmonized) Data Set

The functional consequences of *5-HTTLPR* as well as other susceptibility genes and the $G \times E$ interplay processes as they relate to substance use and abuse from preadolescence to young adulthood are not well understood. Genes and environments may have time-dependent influences on substance use and abuse, as rural African American youth take more active roles in constructing their environments and selecting friends and romantic partners (Rowe, 1994; Scarr, 1992; Scarr & McCartney, 1983). To date, genetically informed studies have had a limited representation of African American children/adolescents, have been based on cross-sectional research designs, and have not simultaneously included in the same longitudinal study neighborhood factors, the *5-HTTLPR* genotype, and substance use. Moreover, rarely are prospective $G \times E$ interplay studies adequately powered statistically to detect $G \times E$ interactions (Duncan & Keller, 2011; Duncan, Pollastri, & Smoller, 2014; Young-Wolff et al., 2011). In this study we addressed these limitations by testing our hypotheses using a harmonized data set of over 2,689 African American children, adopting an accelerated, longitudinal research design, and including measures of neighborhood factors, the serotonin transporter polymorphism *5-HTTLPR*, and substance use. By measuring all three of these factors with the prospective research design, we were able to investigate three-way interrelationships (or interactions) between neighborhood factors, *5-HTTLPR*, and time (or age) on the trajectory of substance use from ages 10 to 24.

We should also note that while future studies using this harmonized data set will need to address issues related to controlling alpha for multiple testing of single nucleotide polymorphisms, this is the first study using this data set, and *5-HTTLPR* was selected because of prior research (Buchmann et al., 2009; Kaufman et al., 2007; van der Zwaluw et al., 2010) and substantive associations with stress sensitivity and emotional regulation (El Hage et al., 2009; Hariri, 2011). The development of this data set represents a substantive advance in $G \times E$ research by utilizing methodological advances in several areas. Based on the limited literature (Mennis & Mason, 2012; van der Zwaluw, 2010), in this study we hypothesized that the short allele of *5-HTTLPR* would significantly interact with selected neighborhood factors across time to influence trajectories of substance use that would be more elevated (i.e., have a steeper upward slope) than for those with the long allele of *5-HTTLPR*. We used multilevel modeling with appropriate control factors to investigate these cross-level interactions.

Method

Development of the GEDID

To test our $G \times E$ hypotheses about interactions between *5-HTTLPR* and neighborhood factors on rates of growth of substance use in African Americans aged 10–24 years, we used data from the GEDID. Our goal in developing an integrated, harmonized database of these five longitudinal studies was to extend the capabilities of any one of the individual studies in three interrelated ways. First, we sought to increase the sample size to facilitate the testing of more complex longitudinal models that included both genetic and environmental variables and would be sufficiently powered to evaluate two-way interactions, including $G \times E$ interactions, as well as some three-way interactions that included time (or age levels) as one of the variables. Second, we wanted to expand the age span to test developmental hypotheses about change in phenotypes such as substance use and how other factors (e.g., neighborhood factors) may interact with genes to impact longer term trajectories (e.g., 10–24 years). Third, it was important to provide a large longitudinal database to investigate $G \times E$ issues for substance use for an understudied population of rural African American children.

Studies included in GEDID

The five longitudinal data sets that were used in the development of the GEDID are briefly characterized in Table 1 (listed as S1 to S5 for the purposes of confidentiality and fulfilling the Institutional Review Board requirements). The vast majority of data collection for these studies occurred with families in the rural South. Highly similar procedures were used in the data collection in that family members were interviewed individually in their homes by locally trained interviewers, and similar measures and construct domains were assessed. As summarized in Table 1, there was limited variation across studies in the number of waves of measurement (four studies have four annual waves of measurement and one study has six annual waves of measurement). There was variation with respect to sample size that ranged from 163 participants to 889 participants. The age ranges varied across studies, although there was substantial overlap that enabled the organization of the data into an accelerated longitudinal research design framework. Note that across all five samples, there were small differences across the individual level socioeconomic indicators as well as the census tract

indicators. Note also that when comparing individual level per capita income with census tract per capita income, the samples across studies were selected from the lower income households of the census tract. Table 2 provides a summary of the number of participants per age level across the five studies. The column totals indicate the maximum number of observations available at each age level and the row totals indicate the maximum total number of data points available from each study (i.e., the sum of the number of individual data points across waves for participants in each study). Hence, there are a maximum of 12,474 data points available for analyses across ages 10–24 years. This extends considerably the age range and statistical power to test $G \times E$ hypotheses beyond any of the individual studies.

In the GEDID, we used a model-based (vs. design-based) approach to IDA in which we adopted a statistical model orientation that linked substantive theory to the sample data via approximating the mechanism by which the dependent variable was generated. This model-based approach stems from Fisher (1922) and contrasts with a design-based approach advocated by Neyman (1934) that utilized randomness via a set of known selection probabilities to weight samples in accordance with the characteristics of the population. As noted by Hussong et al. (2013), most behavioral science studies that attempt to pool data sets adopt a model-based approach because selection probabilities for each of the samples to be pooled are unknown, though it is possible to do so with probability samples. We also used the fixed, rather than random, effects IDA modeling approach because we did not select our five studies from a broader universe or sampling frame of all available studies on African American youth but rather on studies generated in a specific research setting. The fixed-effects approach offered some advantages with respect to common measures used across studies, common data collection procedures, and relatively homogeneous samples. Nevertheless, a limitation of the fixed-effects approach is that generalizability is restricted to the sample characteristics of the studies included in the harmonized data set (i.e., the studies were not randomly selected to represent a broader universe of studies on African American rural youth).

Potential sources of between-group heterogeneity

There are many potential sources of between-group heterogeneity that may be considered in evaluating harmonized data sets (Curran & Hussong, 2009; Hofer & Piccinin, 2009; Hussong et al., 2013). Although the central hypothesis of this study was not primarily focused on the analysis of between-study heterogeneity, such heterogeneity may bias model-specified parameter estimates. With regard to study heterogeneity, there are several features to note. As presented in Table 1, the samples appear highly similar with regard to a number of socioeconomic-based characteristics (e.g., primary caregivers' education, percentage of households below the poverty level) at both the individual and neighborhood (census tract) level. The recruitment procedure for one study (Study 3) primarily targeted single mothers, and hence the prevalence of single-family status was somewhat higher in this study relative to other studies, but the prevalence of single mothers was above 60% across all studies. Thus, socioeconomic status does not appear to be a major source of between-group heterogeneity. All studies included samples recruited from the same geographic region, and hence geographic region does not contribute to study heterogeneity. All samples consisted of

African Americans, and thus race/ethnicity does not contribute to study heterogeneity. Methods of data collection (in-home assessments) were also highly similar across studies, as were the majority of measurement instruments. A characteristic that may contribute to heterogeneity across samples is that for Sample 1, the individual level annual per capita income was somewhat higher than that for the other samples, as was the standard deviation (i.e., greater income variation than for Studies 1–4). This source of variation was not evident at the neighborhood level, where annual per capita income and the standard deviation were similar across all five studies. This suggests that the sample in Study 1 included a subset of individuals with higher levels of income than those in the other four studies. This subset contributed both to an overall higher mean level and to variation in individual-level per capita income. While certainly not exhaustive of all possible between-study differences, these findings suggest a number of between-studies' characteristics that do not distinguish the samples, with the exception of a somewhat higher level of, and greater variation in, individual-level per capita income for Sample 1. However, even this difference does not undermine IDA but rather augments the variability in the samples (Hussong, Huang, Curran, Chassin, & Zucker, 2010).

IRT modeling

As is characteristic of the IDA approach, there were some challenges with regard to establishing commensurate measures across studies (Hussong et al. 2008, 2013). One challenge pertained to a few items across studies that were designed to tap the same construct but had different response options. For example, some substance use items in a given study had more response options than those in another study. By collapsing across-response options for some survey items in some studies we were able to rescale items to a common scale across studies. A second challenge pertained to items across studies that were designed to measure the same construct but contained only a subset of identical or near-identical items across studies. This challenge was not an issue in the current study, but it did arise in constructing the larger GEDID. To address this issue we used Samejima's two-parameter IRT model (Samejima, 1997) to estimate and calibrate items and scales to a common metric to facilitate comparisons across studies (Bauer & Hussong, 2009). We used Samejima's graded-response model for this purpose because many of the survey response options were ordinal with respect to measurement.

Missing value data estimation

Another challenge that we confronted pertained to the treatment of missing values for these five longitudinal studies. To address this challenge, we used multiple imputation methods to estimate missing values (Schafer & Graham, 2002). We used the AMELIA II program (Honaker, King, & Blackwell, 2012) because it was able to accommodate missing values for both continuous and categorical variables. AMELIA II is a multiple-imputation program that uses the expectation maximization bootstrapping (B) algorithm that combines the expectation maximization algorithm that is commonly used in missing value estimation procedures with a bootstrapping approach that operates on the posterior estimates derived from expectation maximization to obtain estimates for the multiple imputed data sets. This estimation procedure, similar to other approaches to estimating missing values, assumes multivariate normality and a missing at random model. Although the distribution of

observed data often deviates from the assumption of multivariate normality, studies have indicated that the multiple imputation approach used by AMELIA II yields findings similar to those of more complicated models even with categorical or mixed data (Schafer, 1997; Schafer & Olsen, 1998).

Variables were initially identified for missing value estimation by age within study. Missing values were not imputed for “planned missingness” cells (i.e., for samples in which no data were collected for a given age level). Furthermore, within sample and age level, missing values were not estimated if more than 40% of the data were missing (Graham, 2009). Missing data were estimated by age level across samples for those variables that were not excluded by the previous criteria (i.e., planned missingness and less than 40% missing data). All missing data estimates were for Level 1 variables; no data were missing for our Level 2 (census tract) variables.

Measures/constructs in GEDID

A multilevel assessment approach was used across all five longitudinal studies in the GEDID and ranged from the collection of data on genes to neighborhood factors. A brief description of measures/constructs from each of these levels is provided now. It is also important to note that for many of the constructs there are multiple reporters (e.g., child, mother, teacher). With regard to genes, the GEDID includes data on three variable number tandem repeats (*5-HTTLPR*, dopamine receptor D4, and monoamine oxidase A) and 95 single nucleotide polymorphisms that have been identified in the literature via significant associations with substance use and mental health phenotypes. A broad range of sociodemographic variables were collected pertaining to such items as household income, number of children in the home, marital status, educational level, and employment status of parents. Behavioral (e.g., externalizing problems), emotional (e.g., anxiety and depressive symptoms), and substance use variables were measured to facilitate an evaluation of a range of relevant health outcome variables. Sexual behaviors were also assessed to determine causes and consequences of risky sexual behavior across time. Negative life events reported by parents and by children were measured to facilitate the testing of $G \times E$ (stressful life events) relationships across time on outcome variables. Parenting characteristics and associations with deviant and substance using peers were collected, and census tract data were collected to permit the inclusion of neighborhood factors. In sum, the GEDID contains a wealth of information across various levels of influence to facilitate the testing of theoretically driven research hypotheses for a relatively large sample of African American children ages 10–24.

Measures/variables used in the current study

Substance use—A composite variable of substance use was created by summing the responses to three items that assessed past year alcohol use, binge drinking, and marijuana use. Item response options for each of these three items were recoded to *yes* (1) and *no* (0). Consistent with efforts to accommodate age differences in size and weight (Donovan, 2009), binge drinking was measured at younger ages (10–14) by the stem question “3 or more drinks at one time” and at older ages (15+) both by “3 or more drinks at one time” and “4 or more drinks at one time.” These items were dichotomized as “0” or “1” in the derivation of the substance use score. Factor analyses of the substance use index across each age group

and across the full sample yielded one dominant factor that accounted for between 46% and 63% of the variance. Reliability estimates ranged across groups from 0.40 to 0.71 with a mean of 0.56. Although this level of reliability is lower than preferred, the derived score is viewed as a summated index (i.e., the sum of three scores) rather than a latent trait in which the underlying trait is the source (cause) of the manifest indicators. This composite measure has been used in prior studies to assess substance use (Brody, Chen, & Kogan, 2010; Wills et al., 2007).

Neighborhood—The neighborhood-level data were created using the 2000 Census Summary Tape File 3 that was geocoded with participants' residential addresses at the initial assessment (Wave 1) of each study. Census tracts are stable subdivisions within each county that include between 1,500 and 8,000 residents. Local authorities typically assist in delineating the census tract boundaries, which typically follow existing physical (e.g., roads, rivers, and railroads) and demographic (e.g., ethnic composition) divisions in the community. A total of 232 census tracts were identified across studies for the pooled data set. Median household income in these census tracts ranged from \$17,399 to 83,293 with a median of \$36,161.

Using the 2000 census tract data and guided by prior studies (Leventhal & Brooks-Gunn, 2000; Sampson et al., 1997; Simons et al., 2005), we created two neighborhood level variables. *Concentrated disadvantage* was assessed by five items: average per capita income, percentage of unemployment, percentage of residents below the poverty threshold, percentage of female-headed households, and percentage of those receiving public assistance. To provide equal weight for each item, per capita income was reverse coded, and we used factor scores obtained through principal components analyses to form the composite scale score. Neighborhood *residential stability* was measured by the percentage of housing tenure (percentage living in the same house since at least 1995) and the percentage of housing occupied by owners.

Genotyping—Participants' DNA was obtained using identical data collection protocols across studies and was collected via Oragene DNA kits (Genetek; Calgary, Alberta, Canada). Participants rinsed their mouths with tap water then deposited 4 ml of saliva in the Oragene sample vial. The vial was sealed, inverted, and shipped via courier to a central laboratory in Iowa City, where samples were prepared according to the manufacturer's specifications. For this study, genotype at the *5-HTTLPR* was determined for each sample as described previously (Bradley, Dodelzon, Sandhu, & Philibert, 2005) using the primers F-GGCG TTGCCGCTCTGAATGC and R-GAGGGACTGAGCTGGACAACCAC, standard *Taq* polymerase and buffer, standard dNTPs with the addition of 100 μ m M7-deaza GTP, and 10% DMSO. The resulting polymerase chain reaction products were electrophoresed on a 6% nondenaturing polyacrylamide gel, and products were visualized using silver staining. Genotype was then called by two individuals blind to the study hypotheses and other information about the participants. Across samples, 6.0% were homozygous for the short allele (short-short), 36.1% were heterozygous (short-long), and 57.9% were homozygous for the long allele. None of the alleles deviated from Hardy-Weinberg equilibrium ($p=.61$, ns). Consistent with prior research (van der Zwaluw et al., 2010; Hariri et al., 2005),

genotyping results were used to form two groups of participants: those homozygous for the long allele (coded as 0, 57.9%) and those with either one or two copies of the short allele (coded as 1, 42.1%).

Statistical analyses

Growth curve modeling (GCM) using multilevel modeling (Bryk & Raudenbush, 1992) was used to test the study hypotheses. GCM estimates the mean and variance for the slope and intercept of substance use, treating the substance use intercept and slope as random effects. The slope corresponding to change in substance use varies depending on the mathematical form being tested (e.g., linear, quadratic, and so forth). In this study, we centered age at 10 years and evaluated both a linear model and a quadratic model of change of drug use across time. The overall model fit statistics for the quadratic model did not indicate improvement in model fit over the linear model, and the key parameter estimates (e.g., in two- and three-way interactions) for the age-squared term were not statistically significant; therefore, the findings were reported for the linear model specification. A linear model specification was consistent with prior research of substance use across this age range and adequately represented the substance use data (Duncan et al., 2014; Chen & Jacobson, 2012), though others have also reported the value of piecewise linear modeling of growth and growth mixture models to approximate complex nonlinear functions (Bollen & Curran, 2006).

Testing the hypotheses involved several sequential model-building steps. Note that we could have simply tested one model (the final model in the sequence), but it is standard practice using this multilevel modeling approach to provide a sequence to decompose how the model changes when moderators are added to the equations (Bryk & Raudenbush, 1992), analogous to how sequential multiple regression model findings are commonly presented; these should not be viewed as independent statistical models with alpha adjustments associated with multiple hypothesis testing but rather as a sequence of related models to facilitate insight into the underlying processes as variables are added to the model. The first model was the specification of an unconditional growth curve analysis to evaluate the significance of the (within-person) slope factor across time (ages 10–24) for substance use. The second model included the within-person slope factor plus the between-individuals' variables of *5-HTTLPR*, primary caregiver's education level, single-family status, family per capita income, sex of the child, and four dummy variables corresponding to the studies in the database.

For the cross-level interaction hypotheses to be tested, about *5-HTLPR* × Neighborhood interactions on changes in substance use across time, the between-individuals' sociodemographic variables (primary caregiver's education level, single family status, family per capita income, sex of the child) were included so as to more clearly interpret the cross-level interactions by controlling for individual level confounders. For example, if we excluded these sociodemographic variables and directly tested the multilevel $G \times E$ interactions, the interpretation of a resulting significant interaction would be ambiguous because we would not know if the relationships were indicated because of a gene by neighborhood effect or by individual level factors that may be associated with, but are not equivalent to, the neighborhood level factors. That is, two individuals may be the same with

respect to sociodemographic variables but differ with respect to the census tracts (neighborhoods) that they reside in, and it is the variation associated with the neighborhood level factor that interacts with the genetic factor to impact the growth of substance use across time. Thus, the inclusion of the individual level sociodemographic variables was to control for (analogous to a covariate set) individual level variation in socioeconomic status based characteristics to facilitate the (unconfounded) testing of the multilevel interaction between neighborhood factors and *5-HTTLPR*. The four dummy variables for study were also included as covariates to facilitate the testing of time trends in the pooled data (Hussong et al., 2013).

The third model included those variables specified in the second model plus two between-neighborhood level factors (based on geographic information system data) of residential stability and neighborhood disadvantage, and two two-way interactions of Age \times Residential Stability and Age \times Neighborhood Disadvantage. These latter two interactions facilitated the testing if growth in substance use across time (or age) was differentially associated with lower versus higher residential stability and lower versus higher neighborhood disadvantage. The fourth model included those variables specified in the third model plus an additional three two-way interactions and two three-way interactions. The additional two-way interactions facilitated the testing of whether the relationship between the *5-HTTLPR* genotype and time (or age) differentially predicted growth in substance use and whether similar conditional relationships between *5-HTTLPR* and each of the two neighborhood factors of residential stability and neighborhood disadvantage predicted growth in substance use. The two three-way interactions facilitated the testing of whether the *5-HTTLPR* genotype significantly varied across time in relation to each of the two neighborhood factors of residential stability and neighborhood disadvantage. For the models specified and estimated, the predictor variables were mean centered and the demographic variables of primary caregiver education, single-family status, family per capita income, child gender, and the four “study” dummy variables were controlled when predicting the means and slopes for substance use. Note that it is possible that there may be studies in which there is substantive interest in hypotheses about individual-level predictors and higher level factors (Hussong et al., 2008) and that there also may be methodological studies that focus on variation in study characteristics (samples, region of country, measures). As is the case in with all data analyses, the specified model needs to be justified in terms of hypotheses and the role(s) that various covariates/predictors assume in the model.

Results

The findings from the GCMs are summarized in Table 3. The unconditional growth model (Model 1) indicated that there was statistically significant variation both for the estimated intercept and slope (age/time) parameters. Model 2 findings indicated that growth in substance use from ages 10 to 24 years was significantly predicted at conventional levels of significance by higher levels of primary caregiver education, being a male, and two of the study dummy variables (S1 and S2) that were associated with a slower rate of acceleration of substance use. In Model 3, the neighborhood factor of residential stability emerged as significant as did the Age (time) \times Residential Stability two-way interaction. Consistent with the recommendations for plotting interactions (Aiken & West, 1991), a plot of this two-way

interaction is provided in Figure 1 and illustrates that for youth in neighborhoods with lower residential stability there is a higher and steeper slope of growth in substance use across time than for youth in neighborhoods with higher residential stability; hence, residential stability moderated the rapidity, or escalation, by which participants increased their substance use from ages 10 to 24 years such that lower residential stability predicted greater increases in substance use across time.

The findings in Model 4 provide another layer of complexity in specifying and evaluating the GCM and the multilevel hypotheses by including the two, three-way interactions of Age (time) \times *5-HTTLPR* \times Neighborhood Factors (residential stability and neighborhood disadvantage). The three-way interaction was significant for residential stability at conventional levels of statistical significance and approached significance for ND. A plot of the significant three-way interaction of Age \times *5-HTTLPR* \times Residential Stability is provided in the Figure 2 and indicates that lower residential stability in combination with the short allele of *5-HTTLPR* yielded the steepest growth in substance use across the age interval of 10–24 years relative to other combinations of residential stability and *5-HTTLPR*. Youth who lived in more residentially stable neighborhoods, regardless of allele status for *5-HTTLPR*, had slower growth trajectories of substance use than those with lower residential stability. Although not different statistically at conventional levels of significance, the three-way interaction of Age \times *5-HTTLPR* \times Neighborhood Disadvantage approached significance ($p < .10$), and plots demonstrated a pattern similar to that provided for residential stability, with the highest and steepest slope for those youth with the short risk allele and higher neighborhood disadvantage.

Overall model fit statistics for the multilevel model presented in Table 3 were based on the likelihood ratio test on deviance statistics; the higher the deviance statistic, the poorer the model fit (Byrk & Raudenbush, 1992). In addition, because our models are nested, the chi-square difference test may be used to determine between model statistical significance. The Model 1 versus Model 2 comparison indicated that adding the individual level predictors to the time-only model significantly improved model fit (χ^2 diff test with 9 $df = 236.529$, $p < .001$). Adding the neighborhood level predictors and the Neighborhood \times Age interactions (Model 3) significantly improved model fit relative to Model 2 (χ^2 diff test with 4 $df = 25.332$, $p < .001$). Finally, adding the two- and three-way interactions that include *5-HTTLPR* significantly improved model fit relative to Model 3 (χ^2 test with 5 $df = 11.621$, $p < .05$). Hence, overall model fit based on the deviance statistics was best for Model 4 that included the three-way interaction among neighborhood, *5-HTTLPR*, and Age (time). The estimated R^2 values for the first- and second-level models are also provided in Table 3 and indicate more potent effects for the first level model (equation).

Test for differential susceptibility

For the statistically significant *5-HTTLPR* \times Residential Stability interaction on growth in substance use, we also conducted a test of the differential susceptibility model versus the diathesis–stress model using the procedures described by Roisman et al. (2012). This procedure identifies regions of significance to evaluate if the association between a moderator (*5-HTTLPR*) and outcome (slope of substance use) at all values of X (residential

stability) fall within the range of interest. If the association between the moderator and the outcome is significant at both low and high ends of the distribution, the differential susceptibility is supported; if the association is only significant at the lower end of the distribution, the diathesis–stress model is supported. We used $\pm 2 SD$ to evaluate the regions of significance and the findings are portrayed in Figure 3. The association between the moderator and the outcome was only significant at the lower end of distribution, thereby supporting a diathesis–stress interpretation of the findings.

Discussion

To our knowledge this is the largest longitudinal study of substance use among African American children and youth with annual assessments spanning the age range of 10–24 years. This was facilitated by the creation of the harmonized database of five longitudinal studies with overlapping age samples to increase sample size and the length of the developmental window. Identical or near-identical survey/questionnaire measures were used across these five longitudinal data sets to assist with the establishment of constructs on a common measurement scale. The five samples were also relatively homogeneous on a several other characteristics (e.g., socioeconomic status, race/ethnicity, geographic region) that can contribute to cross-study heterogeneity (Curran & Hussong, 2009; Hofer & Piccinin, 2009; Hussong et al., 2013). Multiple imputation methods were used to estimate missing values under the model assumptions of missing at random. Individually none of these five studies would have been sufficiently enriched with regard to subjects, measures, or time-points to pursue the research questions of interest in this study. Furthermore, the larger sample size of the five-study harmonized data set enabled the evaluation of $G \times E$ interactions with more adequate power (Duncan & Keller, 2011; Duncan et al., 2014) for a minority sample (African American youth) that has been underrepresented in $G \times E$ studies.

The findings of this study were consistent with those in the literature in supporting significant associations between neighborhood factors and substance use among youth (Mennis & Mason, 2012; Tucker et al., 2013). However, in our study, the neighborhood factor of residential stability was the most prominent neighborhood factor rather than neighborhood disadvantage as results indicated that greater residential instability was associated with higher levels of substance use across ages 10–24 years, as well as a steeper growth (or more rapid increase) in substance use across time. Neighborhood disadvantage and residential instability are highly associated with each other, but in this study two specific characteristics may have been influential in impacting our findings. First, the majority of our samples reside in poor rural communities, and there may be range restrictions on some of the indicators of neighborhood disadvantage (e.g., most communities had a high percentage of residents below the poverty line and receiving public assistance); such range restrictions may have limited the strength of relationships between neighborhood disadvantage and substance use across time. Second, residential stability, which is assessed by indicators of living in the same home for at least 5 years and percentage of housing occupied by owner, may be a proxy that taps other community dimensions (e.g., commitment, cohesion) that are more closely associated with community institutions (e.g., schools, churches) that may jointly provide greater social controls with respect to substance use among youth. This is consistent with notions of how broader neighborhood factors, such as residential stability

may influence collective socialization processes and reduce social disorganization (Brody et al., 2001; Sampson et al., 1997; Shaw & McKay, 1942). Hence, there may be greater variability for the residential stability dimension largely independent of the neighborhood disadvantage dimension that is more strongly associated with lower levels and a lower acceleration of youth substance use across time.

The association of residential stability with changes in substance use across time was moderated by variants of the serotonin transporter *5-HTTLPR*; the short allele of *5-HTTLPR* interacted with residential instability and age (or time) to significantly predict the highest levels of substance use as well as the most rapid acceleration of substance use across age 10–24 years. Thus, the primary hypothesis of this study that the short allele of *5-HTTLPR* would significantly interact with neighborhood factors across time to influence elevated trajectories of substance use was supported by the data. The findings were also suggestive of the possible buffering or protective effects of neighborhood stability on levels and rate of increase of substance use across this age interval. Higher residential stability was associated with lower levels and a lower rate of increase of youth substance use.

In interpreting the findings of this study, it is important to appreciate the challenges encountered by rural African American youth such as those in our sample. These youth and their predecessors have historically lived in circumstances of very high poverty and unemployment and limited educational facilities and job opportunities, and they have confronted a host of structural (e.g., limited healthcare coverage and access) and financial barriers. These conditions, and well as other challenges (e.g., discrimination, mistrust of others), collectively contribute to significantly higher stress levels and poorer health outcomes and fewer prevention and treatment options as well as restrict the prospects of upward mobility within their own communities or of having adequate income or resources to move to new communities. As the current study demonstrates, this does not imply that all rural African American youth will succumb to drug use and abuse. Rather, even within these relatively financially impoverished rural areas, there are individual differences in genes and neighborhood factors that contribute to variation in drug use across time. Our study suggests that residential stability, that is the percentage of houses occupied by owners and families living in the same house for a long period of time, was associated with lower levels of substance use and a slower rate of growth of use from ages 10 to 24 years. These findings are likely due to a greater personal investment and more social cohesion in neighborhoods where individuals own their homes and/or have resided in their homes for a longer period of time (Hawkins et al., 1992; Hurd et al., 2013; Kawachi & Berkman, 2003; Sampson et al., 1997).

In addition, lower residential stability, in conjunction with the risk allele of *5-HTTLPR*, was associated with higher levels and a faster rate of growth of substance use from ages 10 to 24 years. Hence, variation both in a genetic factor and neighborhood conditions were of importance in identifying substance use levels and across-time trajectories of drug use for these rural African American youth.

This study has focused on multilevel developmental relations between individual characteristics (e.g., sociodemographic variables, *5-HTTLPR*) and neighborhood factors

(residential stability, neighborhood disadvantage) that influence increases in substance use from ages 10 to 24 years. Although this two-level approach provides insight into the across-time, developmental nature of the interrelationships of these two contributors, human development occurs across multiple levels of analyses, and as noted by many investigators (Chassin et al., 2013; Schulenberg, Maggs, & O'Malley, 2003; Windle, 2010; Zucker, 2006), substance use and substance disorders are impacted by developmental changes across these levels. That is, the interval spanning the age-course in this study (10–24 years) includes major developmental changes in biological (e.g., onset of and growth in puberty, increases in physical size, brain alterations), cognitive (e.g., personal identity), affective (e.g., romantic involvements), and social domains (e.g., stronger peer relationships, autonomy from and renegotiated relationships with parents), as well as physical settings (e.g., changes from elementary to middle and high schools, and then to either college or the workplace). There is vast literature that has linked these changes in development to changes in substance use (for reviews, see Brown et al., 2008; Chassin et al., 2013; Windle et al., 2008; Zucker, 2006). Furthermore, the NIH currently has a Request for applications to pursue a large-scale longitudinal consortium study entitled Adolescent Brain Cognitive Development (<http://grants.nih.gov/grants/guide/rfa-files/RFA-DA-14.html>) to investigate how exposure to various substances (e.g., alcohol, marijuana) may influence normative and atypical brain development across an interval from ages 10 to 21 years. Hence, the investigation of associations between features of human development (e.g., puberty, changes in family and peer relationships) and substance use is a prominent topic in the literature, with much to be learned about the timing and intersection of substance use and development across the multiple domains previously identified.

The findings for the *5-HTTLPR* polymorphism and substance use among adolescents and young adults have been mixed (Guo et al., 2007; Hopfer et al., 2005; van der Zwaluw et al., 2010). Our findings suggest that one possible contributor to the variation in these findings may be that the serotonin transporter is most effectively studied within a $G \times E$ framework rather than a “main effects” framework. That is, the *5-HTTLPR* genotype, and possibly other candidate genes, may best be viewed in more of a developmental contextual model where genes interact with other features of the environment (e.g., parents, peers, neighborhood factors) to impact risk or protectiveness across time (Brody, Beach, et al., 2013; Windle, 2010; Zucker, 2006). Such a model seems more consistent with broader genetic findings of common gene variants having smaller and more subtle effects (Goldstein, 2009). Based on the analytic approach of Roisman et al. (2013), the diathesis–stress model was more strongly supported than the differential susceptibility model (Belsky & Pluess, 2009). However, even though five samples were integrated for this article, it is important to consider that the environmental range was relatively restricted toward the poorer end of the spectrum of poor to enriched environments, and the differential susceptibility model may have been supported if the range had been broader. Such an extended range would also have been more beneficial to study of neighborhood dimensions related to social control, collective socialization, and disorganization as described in the theoretical frameworks of other investigators (Brody et al., 2001; Sampson et al., 1997; Shaw & McKay, 1942). In addition, these samples represented only African American youth and hence cross-racial/ethnic group comparisons for the $G \times E$ relationships could not be investigated.

In summary, our findings supported the study hypothesis that the short allele of the 5-HTTLPR genotype moderated the relationship between neighborhood factors (specifically residential instability) and the rate of increase in substance use among African American youth from ages 10 to 24 years. Residential instability in conjunction with the short allele of 5-HTTLPR was associated with the highest levels and a faster rate of change in substance use across time. These findings assist in suggesting more subtle relationships between genetic and environmental factors, and much research remains to be completed to explicate the causal linkages across levels of analyses to identify and more fully comprehend the specific distal and proximal mechanisms that account for the $G \times E$ relationships and their implications for interventions and public health practice (Lesch, 2005; Moffitt, Caspi, & Rutter, 2006). Furthermore, in our study the samples were African American and primarily drawn of southern, rural regions; variation both with respect to gene frequencies and to local environmental factors (e.g., patterns of use and availability of substances) may have contributed to our findings. Replication and extension of our findings supporting this $G \times E$ interaction await further study by other research groups, though as reported by Duncan et al. (2014), large-scale genomic and $G \times E$ studies of African American samples has been quite limited, and longitudinal studies of African American youth even more limited. Smaller sample sizes present a serious challenge to the investigation of $G \times E$ interactions, and this is compounded in studies of African Americans as it imposes difficulties on replication and cross-validation of findings. Nevertheless, the findings of this study provided an application of a methodological approach (creating a harmonized data set) to increase statistical power to test $G \times E$ interactions and to expand the age interval to address an important issue in the literature pertaining to gene–neighborhood relationships and their impact on trajectories of substance use among African American children from ages 10 to 24 years. In future research, we plan to use the GEDID to examine other $G \times E$ hypotheses that incorporate multilocus (Brody, Chen, et al., 2013) and polygenic indexes (Purcell et al., 2009) of genetic influences and other variables (e.g., family, peer) in the data set to address dynamic $G \times E$ change processes for these samples.

Acknowledgments

This research was supported by Award 1P30DA027827 from the National Institute on Drug Abuse (all authors) and Award K05AA021143 from the National Institute on Alcoholism and Alcohol Abuse (to M.W.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Drug Abuse, the National Institute on Alcoholism and Alcohol Abuse, or the NIH.

References

- Aiken, L.S.; West, S.W. Multiple regression: Testing and interpreting interactions. Newbury Park, CA: Sage; 1991.
- Auerbach JG, Faroy M, Ebstein R, Kahana M, Levine J. The association of the dopamine D4 receptor gene (DRD4) and the serotonin transporter promoter gene (5-HTTLPR) with temperament in 12-month-old infants. *Journal of Child Psychology and Psychiatry*. 2001; 42:777–783. [PubMed: 11583250]
- Bauer DJ, Hussong AM. Psychometric approaches for developing commensurate measures across independent studies: Traditional and new models. *Psychological Methods*. 2009; 14:101–125. [PubMed: 19485624]

- Beach SRH, Brody GH, Lei MK, Philibert RA. Differential susceptibility to parenting among African American youths: Testing the DRD4 hypothesis. *Journal of Family Psychology*. 2010; 24:513–521. [PubMed: 20954761]
- Belsky J, Pluess M. Beyond diathesis stress: Differential susceptibility to environmental influences. *Psychological Bulletin*. 2009; 135:885–908. [PubMed: 19883141]
- Bennett SN, Caporaso N, Fitzpatrick AL, Agrawal A, Barnes K, Boyd HA, et al. Phenotype harmonization and cross-study collaboration in GWAS Consortia: The GENEVA experience. *Genetic Epidemiology*. 2011; 35:159–173. [PubMed: 21284036]
- Bollen, KA.; Curran, PJ. *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley; 2006.
- Bradley SL, Dodelzon K, Sandhu HK, Philibert RA. Relationship of serotonin transporter gene polymorphisms and haplotypes to mRNA transcription. *American Journal of Medical Genetics*. 2005; 136B:58–61. [PubMed: 15858822]
- Brody GH, Beach SRH, Hill KG, Howe GW, Prado G, Fullerton SM. Using genetically informed, randomized prevention trials to test etiological hypotheses about child and adolescent drug use and psychopathology. *American Journal of Public Health*. 2013; 103(Suppl. 1):S19–S24. [PubMed: 23927515]
- Brody GH, Chen YF, Beach SRH. Differential susceptibility to prevention: GABAergic, dopaminergic, and multilocus effects. *Journal of Child Psychology and Psychiatry*. 2013; 54:863–871. [PubMed: 23294086]
- Brody GH, Chen YF, Kogan SM. A cascade model connecting life stress to risk behavior among rural African American emerging adults. *Development and Psychopathology*. 2010; 22:667–678. [PubMed: 20576186]
- Brody GH, Ge X, Conger R, Gibbons FX, McBride MV, Gerrard M, et al. The influence of neighborhood disadvantage, collective socialization, and parenting on African American children's affiliation with deviant peers. *Child Development*. 2001; 72:1231–1246. [PubMed: 11480944]
- Bryk, AS.; Raudenbush, SW. *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage; 1992.
- Buchmann AF, Schmid B, Blomeyer D, Becker K, Treutlein J, Zimmermann US, et al. Impact of age at first drink on vulnerability to alcohol-related problems: Testing the marker hypothesis in a prospective study of young adults. *Journal of Psychiatric Research*. 2009; 43:1205–1212. [PubMed: 19332346]
- Canli T, Qiu M, Omura K, Congdon E, Haas BW, Amin Z. Neural correlates of epigenesis. *Proceedings of the National Academy of Sciences*. 2006; 103:16033–16038.
- Caspi A, Hariri AR, Holmes A, Uher R, Moffitt TE. Genetic sensitivity to the environment: The case of the serotonin transporter gene and its implications for studying complex diseases and traits. *American Journal of Psychiatry*. 2010; 167:509–527. [PubMed: 20231323]
- Chassin L, Sher KJ, Hussong A, Curran P. The developmental psychopathology of alcohol use and alcohol disorders: Research achievements and future directions. *Development and Psychopathology*. 2013; 25:1567–1584. [PubMed: 24342856]
- Chen P, Jacobson KC. Developmental trajectories of substance use from early adolescence to young adulthood: Gender and racial/ethnic differences. *Journal of Adolescent Health*. 2012; 50:154–163. [PubMed: 22265111]
- Curran PJ, Hussong AM. Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*. 2009; 14:81–100. [PubMed: 19485623]
- Curran PJ, Hussong AM, Cai L, Huang W, Chassin L, Sher KJ, et al. Pooling data from multiple longitudinal studies: The role of item response theory in integrative data analysis. *Developmental Psychology*. 2008; 44:365–380. [PubMed: 18331129]
- Cutrona CE, Russell DW, Brown PA, Hessling RM, Clark LA, Gardner KA. Neighborhood context, personality, and stressful life events as predictors of depression among African American women. *Journal of Abnormal Psychology*. 2005; 114:3–15. [PubMed: 15709807]
- Dalaker, J. *Poverty in the United States, 2000*. Washington, DC: US Census Bureau; 2001.
- Donovan JE. Estimated blood alcohol concentrations for child and adolescent drinking and their implications for screening instruments. *Pediatrics*. 2009; 123:e975–e981. [PubMed: 19482748]

- Duncan LE, Keller MC. A critical review of the first 10 years of candidate gene-by-environment interaction research in psychiatry. *American Journal of Psychiatry*. 2011; 168:1041–1049. [PubMed: 21890791]
- Duncan LE, Pollastri AR, Smoller JW. Mind the gap: Why many geneticists and psychological scientists have discrepant views about gene-environment interaction (GxE) research. *American Psychologists*. 2014; 69:249–268.
- Duncan SC, Duncan TE. A multivariate latent growth curve analysis of adolescent substance use. *Structural Equation Modeling*. 1996; 3:323–347.
- El Hage W, Powell JF, Surguladze SA. Vulnerability to depression: What is the role of stress genes in gene x environment interaction? *Psychological Medicine*. 2009; 39:1407–1411. [PubMed: 19215634]
- Ellickson PL, Bell RM, McGuigan K. Preventing adolescent drug use: Long-term results of a junior high program. *American Journal of Public Health*. 1993; 83:856–861. [PubMed: 8498624]
- Espeland MA, Dotson K, Jaramillo SA, Kahn SE, Harrison B, Montez M, et al. Consent for genetics studies among clinical trial participants: Findings from Action for Health in Diabetes (Look AHEAD). *Clinical Trials*. 2006; 3:443–456. [PubMed: 17060218]
- Feinn R, Nellissery M, Kranzler HR. Meta-analysis of the association of a functional serotonin transporter promoter polymorphism with alcohol dependence. *American Journal of Medical Genetics*. 2005; 133B:79–84. [PubMed: 15635638]
- Fisher RA. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society (London)*. 1922; 222A:309–368.
- Fortier I, Burton PR, Robson PJ, Ferretti V, Little J, L'Heureux F, et al. Quality, quantity and harmony: The DataSHaPER approach to integrating data across bioclinical studies. *International Journal of Epidemiology*. 2010; 39:1383–1393. [PubMed: 20813861]
- French, K.; Finkbiner, R.; Duhamel, L. Patterns of substance use among minority youth and adults in the United States: An overview and synthesis of national survey findings. Fairfax, VA: US Department of Health and Human Services; 2002.
- Galea S, Rudenstine S. Challenges in understanding disparities in drug use and its consequences. *Journal of Urban Health*. 2005; 82(2, Suppl. 3):iii, 5–12.
- Goldstein DB. Common genetic variation and human traits. *New England Journal of Medicine*. 2009; 360:1696–1698. [PubMed: 19369660]
- Graham JW. Missing data analysis: Making it work in the real world. *Annual Review of Psychology*. 2009; 60:549–576.
- Granda, P.; Blasczyk, E. Data harmonization: Guidelines for best practice in cross-cultural surveys. Ann Arbor, MI: University of Michigan, Institute for Social Research, Survey Research Centre; 2010.
- Green D, Cushman M, Dermond N, Johnson EA, Castro C, Arnett D, et al. Obtaining informed consent for genetic studies: The multiethnic study of atherosclerosis. *American Journal of Epidemiology*. 2006; 164:845–851. [PubMed: 16928727]
- Guo G, Wilhelmsen K, Hamilton N. Gene-life course interaction for alcohol consumption in adolescence and young adulthood: Five monoamine genes. *American Journal of Medical Genetics*. 2007; 144B:417–423. [PubMed: 17440951]
- Hariri, AR. Neurobiological mechanisms supporting gene–environment interaction effects. In: Dodge, K.; Rutter, M., editors. *Gene–environment interactions in developmental psychopathology*. New York: Guilford Press; 2011. p. 59-69.
- Hariri AR, Drabant EM, Munoz KE, Kolachana BS, Mattay VS, Egan MF, et al. A susceptibility gene for affective disorders and the response of the human amygdala. *Archives of General Psychiatry*. 2005; 62:146–152. [PubMed: 15699291]
- Hawkins JD, Catalano RF, Arthur MW. Promoting sciencebased prevention in communities. *Addictive Behaviors*. 2002; 27:951–976. [PubMed: 12369478]
- Hawkins JD, Catalano RF, Miller JY. Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance use prevention. *Psychological Bulletin*. 1992; 112:64–105. [PubMed: 1529040]

- Heils A, Teufel A, Petri S, Stober G, Riederer P, Bengel D, et al. Allelic variation of human serotonin transporter gene expression. *Journal of Neurochemistry*. 1996; 66:2621–2624. [PubMed: 8632190]
- Herman AI, Kaiss KM, Ma R, Philbeck JW, Hasan A, Dasti H, et al. Serotonin transporter promoter polymorphism and monoamine oxidase Type A VNTR allelic variants together influence alcohol binge drinking risk in young women. *American Journal of Medical Genetics*. 2005; 133B:74–78. [PubMed: 15635592]
- Hofer SM, Piccinin AM. Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological Methods*. 2009; 14:150–164. [PubMed: 19485626]
- Honaker, J.; King, G.; Blackwell, M. AMELIA II: A program for missing data[Computer software]. 2012. Retrieved from <http://gking.harvard.edu/amelia>
- Hopfer CJ, Timberlake D, Haberstick B, Lessem JM, Ehringer MA, Smolen A, et al. Genetic influence on quantity of alcohol consumed by adolescents and young adults. *Drug and Alcohol Dependence*. 2005; 78:187–193. [PubMed: 15845322]
- Hurd NM, Stoddard SA, Zimmerman MA. Neighborhoods, social support, and African American adolescents' mental health outcomes: A multilevel path analysis. *Child Development*. 2013; 84:858–874. [PubMed: 23199188]
- Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology*. 2013; 9:61–89.
- Hussong AM, Flora DB, Curran PJ, Chassin LA, Zucker RA. Defining risk heterogeneity for internalizing symptoms among children of alcoholic parents. *Development and Psychopathology*. 2008; 20:165–193.
- Hussong AM, Huang W, Curran PJ, Chassin L, Zucker RA. Parent alcoholism impacts the severity and timing of children's externalizing symptoms. *Journal of Abnormal Child Psychology*. 2010; 38:367–380. [PubMed: 20084453]
- Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Demographic subgroup trends among adolescents for fifty-one classes of licit and illicit drugs. Ann Arbor, MI: University of Michigan, Institute for Social Research; 2013. p. 1975-2012. Monitoring the Future Occasional Paper No. 79 Retrieved from <http://www.monitoringthefuture.org/pubs/occpapers/mtf-occ79.pdf>
- Kandel, DB. Ethnic differences in drug use: Patterns and paradoxes. In: Botvin, GJ.; Schinke, S.; Orlandi, MA., editors. *Drug abuse prevention with multiethnic youth*. Thousand Oaks, CA: Sage; 1995. p. 81-104.
- Kandel D, Schaffran C, Hu MC, Thomas Y. Age-related differences in cigarette smoking among Whites and African-Americans: Evidence for the crossover hypothesis. *Drug and Alcohol Dependence*. 2011; 118:280–287. [PubMed: 21561724]
- Kann L, Kinchen S, Shanklin S, Flint KH, Hawkins J, Harris WA, et al. Youth risk behavior surveillance—United States, 2013. *Mortality and Morbidity Weekly Report*. 2014; 63:1–168.
- Kaufman J, Yang B, Douglas-Palumber H, Crouse-Artus M, Lipschitz D, Krystal J, et al. Genetic and environmental predictors of early alcohol use. *Biological Psychiatry*. 2007; 61:1228–1234. [PubMed: 17123474]
- Kawachi, I.; Berkman, LF., editors. *Neighborhoods and health*. Oxford: Oxford University Press; 2003.
- Kendler KS, Eaves L. Models for the joint effect of genotype and environment on liability to psychiatric illness. *American Journal of Psychiatry*. 1986; 143:279–289. [PubMed: 3953861]
- Kogan SM, Berkel C, Chen YF, Brody GH, Murry VM. Metro status and African-American adolescents' risk for substance use. *Journal of Adolescent Health*. 2006; 38:454–457. [PubMed: 16549312]
- Kreek MJ, Nielsen DA, Butelman E, LaForge S. Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction. *Nature Neuroscience*. 2005; 8:1450–1457. [PubMed: 16251987]
- Lesch KP. Alcohol dependence and Gene x Environment interaction in emotion regulation: Is serotonin the link? *European Journal of Pharmacology*. 2005; 526:113–124. [PubMed: 16288736]
- Leventhal T, Brooks-Gunn J. The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes. *Psychological Bulletin*. 2000; 126:309–337. [PubMed: 10748645]

- Levine SB, Coupey SM. Adolescent substance use, sexual behavior, and metropolitan status: Is “urban” a risk factor? *Journal of Adolescent Health*. 2003; 32:350–355. [PubMed: 12729984]
- McQuillan GM, Pan Q, Porter KS. Consent for genetic research in a general population: An update on the National Health and Nutrition Examination Survey experience. *Genetics in Medicine*. 2006; 8:354–360. [PubMed: 16778597]
- McQuillan GM, Porter KS, Agelli M, Kington R. Consent for genetic research in a general population: The NHANES experience. *Genetics in Medicine*. 2003; 5:35–42. [PubMed: 12544474]
- Mennis J, Mason MJ. Social and geographic contexts of adolescent substance use: The moderating effects of age and gender. *Social Networks*. 2012; 34:150–157.
- Moffitt TE, Caspi A, Rutter ML. Measured gene–environment interactions in psychopathology: Concepts, research strategies, and implications for research, intervention, and public understanding of genetics. *Perspectives on Psychological Science*. 2006; 1:5–27. [PubMed: 26151183]
- Monahan K, Egan EA, Van Horn ML, Arthur M, Hawkins D. Community-level effects of individual and peer risk and protective factors on adolescent substance use. *Journal of Community Psychology*. 2011; 39:478–498. [PubMed: 23976800]
- Neyman J. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*. 1934; 97:558–606.
- Odgerel Z, Talati A, Hamilton SP, Levinson DF, Weissman MM. Genotyping serotonin transporter polymorphisms 5-HTTLPR and rs25531 in European- and African-American subjects from the National Institute of Mental Health’s Collaborative Center for Genomic Studies. *Translational Psychiatry*. 2013; 3:e307. [PubMed: 24064711]
- Patkar A, Berrettini WH, Hoehe M, Hill KP, Sterling RC, Gotthel E, et al. Serotonin transporter (5-HTT) gene polymorphisms and susceptibility to cocaine dependence among African-American individuals. *Addiction Biology*. 2001; 6:337–345. [PubMed: 11900612]
- Pentz MA, Dwyer JH, MacKinnon DP, Flay BR, Hansen WB, Wang EY, et al. A multicommunity trial for primary prevention of adolescent drug abuse: Effects on drug use prevalence. *Journal of the American Medical Association*. 1989; 261:3259–3266. [PubMed: 2785610]
- Purcell SM, Wray NR, Stone JL, Visscher PM, O’Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009; 460:748–752. [PubMed: 19571811]
- Reardon SF, Buka SL. Differences in onset and persistence of substance abuse and dependence among Whites, Blacks, and Hispanics. *Public Health Reports*. 2002; 117(Suppl. 1):S51–S59. [PubMed: 12435827]
- Robinson OJ, Overstreet C, Allen PS, Letkiewicz A, Vytal K, Pine DS, et al. The role of serotonin in the neurocircuitry of negative affective bias: Serotonergic modulation of the dorsal medial prefrontal-amygdala “aversive amplification” circuit. *NeuroImage*. 2013; 78:217–223. [PubMed: 23583742]
- Roisman GI, Newman DA, Fraley C, Haltigan JD, Groh AM, Haydon KC. Distinguishing differential susceptibility from diathesis stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*. 2012; 24:389–409. [PubMed: 22559121]
- Rowe, D. *The limits of family influence*. New York: Guilford Press; 1994.
- Samejima, F. Graded response model. In: van der Linden, WJ.; Hambleton, RK., editors. *Handbook of modern item response theory*. New York: Springer-Verlag; 1997. p. 85-100.
- Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science*. 1997; 277:5328, 918–924.
- Scarr S. Developmental theories for the 1990s: Development and individual differences. *Child Development*. 1992; 63:1–19. [PubMed: 1343618]
- Scarr S, McCartney K. How people make their own environments: A theory of genotype → environment effects. *Child Development*. 1983; 54:424–435. [PubMed: 6683622]
- Schafer, JL. *Analysis of incomplete multivariate data*. London: Chapman & Hall; 1997.
- Schafer JL, Graham JW. Missing data: Our view of the state of the art. *Psychological Methods*. 2002; 7:147–177. [PubMed: 12090408]

- Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: A data analyst's perspective. *Multivariate Behavioral Research*. 1998; 33:545–571. [PubMed: 26753828]
- Schulenberg JE, Maggs JL. A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. *Journal of Studies on Alcohol*. 2002; 14(Suppl):54–70. [PubMed: 12022730]
- Schulenberg, JE.; Maggs, JM.; O'Malley, PM. How and why the understanding of developmental continuity and discontinuity is important: The sample case of long-term consequences of adolescent substance use. In: Mortimer, JT.; Shanahan, MJ., editors. *Handbook of the life course*. New York: Plenum Press; 2003. p. 413-436.
- Schuster MA, Elliott MN, Kanouse DE, Wallander JL, Tortolero SR, Ratner JA, et al. Racial and ethnic health disparities among fifth-graders in three cities. *New England Journal of Medicine*. 2012; 367:735–745. [PubMed: 22913683]
- Shanahan MJ, Hofer SM. Social context in gene–environment interactions: Retrospect and prospect. *Journals of Gerontology*. 2005; 60B:65–76. [PubMed: 15863711]
- Shaw, CR.; McKay, H. *Juvenile delinquency in urban areas*. Chicago: University of Chicago Press; 1942.
- Sher KJ, Jackson KM, Steinley D. Alcohol use trajectories and the ubiquitous cat's cradle: Cause for concern? *Journal of Abnormal Psychology*. 2011; 120:322–335. [PubMed: 21319874]
- Simons RL, Simons LG, Burt CH, Brody GH, Cutrona C. Collective efficacy, authoritative parenting and delinquency: A longitudinal test of a model integrating community- and family level processes. *Criminology*. 2005; 43:989–1029.
- Tarter RE, Vanyukov M. Alcoholism: A developmental disorder. *Journal of Consulting and Clinical Psychology*. 1994; 62:1096–1107. [PubMed: 7860808]
- Tucker JS, Pollard MS, de la Haye K, Kennedy DP, Green HD Jr. Neighborhood characteristics and the initiation of marijuana use and binge drinking. *Drug and Alcohol Dependence*. 2013; 128:83–89. [PubMed: 22938829]
- van der Zwaluw CS, Engels RCME, Vermulst AA, Rose RJ, Verkes RJ, Buitelaar J, et al. A serotonin transporter polymorphism (5-HTTLPR) predicts the development of adolescent alcohol use. *Drug and Alcohol Dependence*. 2010; 112:134–139. [PubMed: 20598814]
- Wallace JM Jr, Bachman JG, O'Malley PM, Johnston LD, Schulenberg JE, Cooper SM. Tobacco, alcohol, and illicit drug use: Racial and ethnic differences among U.S. high school seniors, 1976–2000. *Public Health Reports*. 2002; 117(Suppl. 1):S67–S75. [PubMed: 12435829]
- Wills TA, Murry VM, Brody GH, Gibbons FX, Gerrard M, Walker C, et al. Ethnic pride and self-control related to protective and risk factors: Test of the theoretical model for the Strong African American Families program. *Health Psychology*. 2007; 26:50–59. [PubMed: 17209697]
- Windle M. A multilevel developmental contextual approach to substance use and addiction. *BioSocieties*. 2010; 5:124–136. [PubMed: 22754585]
- Windle, M.; Davies, P. Developmental theory and research. In: Leonard, KE.; Blane, HT., editors. *Psychological theories of drinking and alcoholism*. 2nd. New York: Guilford Press; 1999. p. 164-202.
- Windle M, Spear LP, Fuligni AJ, Angold A, Brown JD, Pine D, et al. Transitions into underage and problem drinking: Developmental processes and mechanisms between ages 10 and 15. *Pediatrics*. 2008; 121(Suppl. 4):S273–S289. [PubMed: 18381494]
- Xie P, Kranzler HR, Poling J, Stein MB, Anton RF, Brady K, et al. Interactive effect of stressful life events and the serotonin transporter 5-HTTLPR genotype on posttraumatic stress disorder diagnosis in 2 independent populations. *Archives of General Psychiatry*. 2009; 66:1201–1209. [PubMed: 19884608]
- Young-Wolff KC, Enoch MA, Prescott CA. The influence of gene–environment interactions on alcohol consumption and alcohol use disorders: A comprehensive review. *Clinical Psychology Review*. 2011; 31:800–816. [PubMed: 21530476]
- Zucker, RA. Alcohol use and the alcohol use disorders: A developmental–biopsychosocial systems formulation covering the life course. In: Cicchetti, D.; Cohen, DJ., editors. *Developmental psychopathology: Vol. 3. Risk, disorder, and adaptation*. 2nd. Hoboken, NJ: Wiley; 2006. p. 620-656.

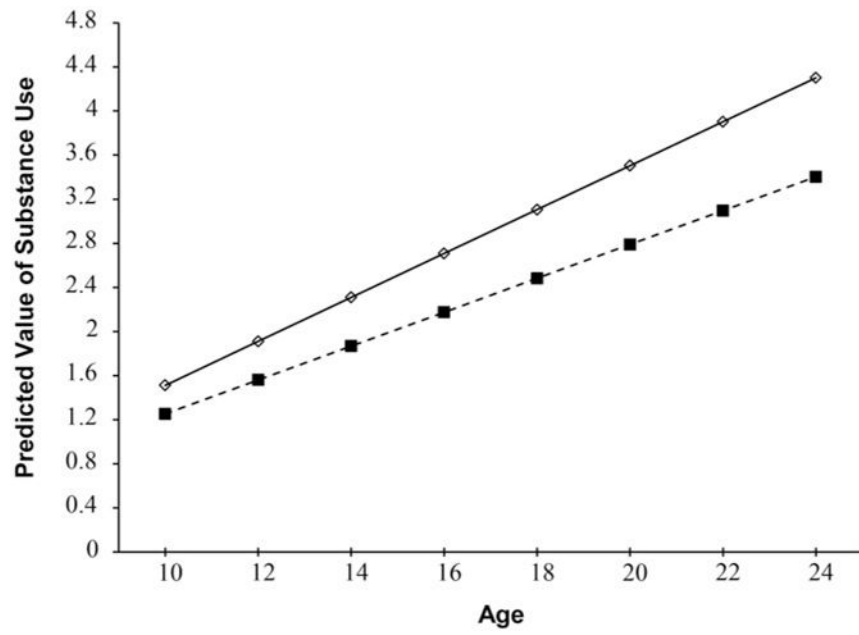


Figure 1. Growth in substance use by residential stability across age levels. Low stress = 1 *SD* below the mean; high stress = 1 *SD* above the mean.

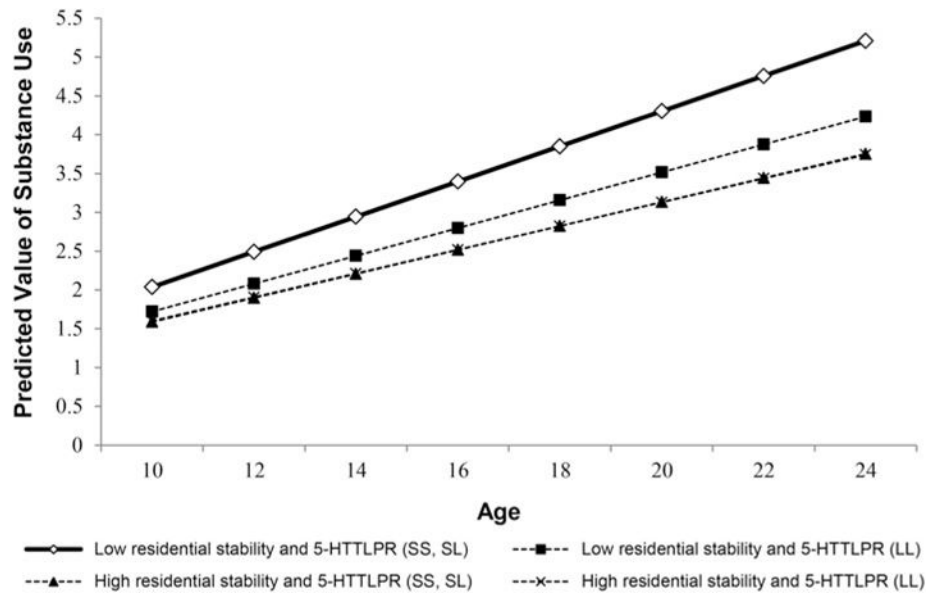


Figure 2. Growth in substance use by *5-HTTLPR* status, residential stability, and *5-HTTLPR* by residential stability interaction. Low residential stability = 1 *SD* below the mean; high residential stability = 1 *SD* above the mean.

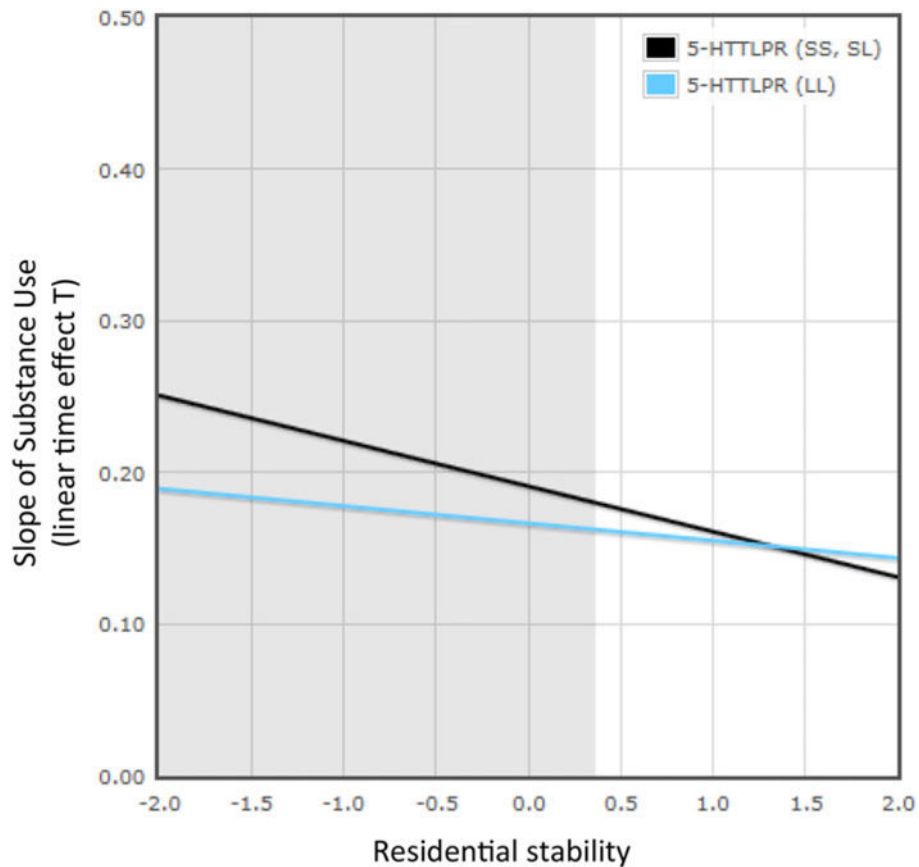


Figure 3.

(Color online) The effects of neighborhood residential stability (range = <2 SD to >2 SD from mean) on the predicted growth parameter of substance use by *5-HTTLPR*, and the effects of neighborhood residential stability (range <2 SD to >2 SD from mean) on the predicted growth parameter of substance use by *5-HTTLPR* using a multilevel growth model. The lines represent the regression lines for short–short (SS) and short–long (SL) allele carriers (simple slope test: $b = -0.03$, $p = .000$) and long (LL) allele carriers (simple slope test: $b = -0.01$, $p = .021$). The gray areas are significant confidence regions (0.365). The proportion of the interaction is $.04$.

Table 1

Data sets used in harmonization

	<u>Wave 4 S1</u>	<u>Wave 6 S2</u>	<u>Wave 4 S3</u>	<u>Wave 4 S4</u>	<u>Wave 4 S5</u>
	<i>M (SD), %</i>	<i>M (SD), %</i>	<i>M (SD), %</i>	<i>M (SD), %</i>	<i>M (SD), %</i>
<i>N</i>	468 (17.4%)	667 (24.8%)	163 (6.1%)	502 (18.7%)	889 (33.1%)
Age range	16-22	11-19	15-20	13-19	10-24
<i>5-HTTLPR SS/SL</i>	43.8%	40.1%	43.0%	41.3%	43.1%
<i>5-HTTLPR LL</i>	56.2%	59.9%	57.0%	58.7%	56.9%
Primary caregivers' education	12 years	12 years	12 years	12 years	12 years
	229 (48.9%)	346 (51.9%)	113 (69.3%)	250 (49.8%)	504 (56.7%)
Single family status	321 (68.6%)	423 (63.4%)	151 (92.6%)	335 (66.7%)	571 (64.2%)
Annual per capita income at Wave 1 (US\$)	9,940 (13,925)	6,230 (4,662)	4,806 (4,140)	5,658 (5,641)	6,732 (6,049)
Residential Stability Census Tract Indicators					
Living in same house since at least 1995	62.0%	60.6%	58.7%	61.4%	50.7%
Owner occupied housing units	72.6%	71.7%	70.0%	69.4%	59.0%
Neighborhood Disadvantage Census Tract Indicators					
Single-mom families	15.5%	16.5%	14.6%	19.0%	19.2%
Annual per capita income (US\$)	15,505 (4,235)	16,858 (5,220)	16,882 (3,773)	14,669 (2,989)	15,316 (5,075)
Households on public assistance	4.7%	4.7%	3.8%	6.0%	5.9%
Households below poverty level	19.0%	19.7%	17.3%	24.5%	19.1%
Total unemployed	6.7%	7.7%	6.0%	9.3%	9.0%

Table 2

Frequency of observations for samples by age (years) for the five studies

Study	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total
S1						55	237	425	467	413	231	43	1			1872
S2	249	664	418	252	445	568	628	211	3							4002
S3			11	74	134	162	153	92	26							652
S4					98	422	487	410	80	9						1506
S5	165	501	360	425	253	200	380	277	205	338	313	299	367	288	71	4442
Total	165	750	1024	854	579	877	1587	1717	1760	1122	738	530	410	289	71	12474

Table 3

Neighborhood residential stability and 5-HTTLPR as predictors of change in substance use over time

Fixed Effect	Model 1	Model 2	Model 3	Model 4
	Unstand. b	Unstand. b	Unstand. b	Unstand. b
Within-individual level				
Age (time)	0.148** (0.008)	0.177** (0.008)	0.176** (0.008)	0.167** (0.008)
Between-individuals level				
5-HTTLPR (1 = short allele)		0.007 (0.024)	0.007 (0.024)	-0.090* (0.041)
PCs' education (1 = college)		0.057* (0.022)	0.056* (0.022)	0.056* (0.022)
Single family status		0.152 (0.023)	0.018 (0.024)	0.018 (0.024)
Family per capita income		-0.001 (0.001)	-0.001 (0.001)	-0.001 (0.001)
Male		0.047* (0.021)	0.047* (0.021)	0.048* (0.021)
S1		-0.760** (0.059)	-0.728** (0.065)	-0.729** (0.064)
S2		0.038 (0.025)	0.021 (0.030)	0.021 (0.030)
S3		-0.160 [†] (0.090)	-0.163 [†] (0.091)	-0.160 [†] (0.091)
S4		-0.149* (0.061)	-0.131* (0.065)	-0.129* (0.065)
Between-neighborhoods level				
RS			0.082** (0.020)	0.059** (0.023)
ND			-0.002 (0.022)	-0.029 (0.025)
Cross-level interaction				
Age × RS			-0.019** (0.005)	-0.011* (0.005)
Age × ND			-0.002 (0.005)	0.005 (0.006)
Age × 5-HTTLPR				0.024* (0.010)
5-HTTLPR × RS				0.054 [†] (0.031)
5-HTTLPR × ND				0.064* (0.030)
Age × 5-HTTLPR × RS				-0.019* (0.008)
Age × 5-HTTLPR × ND				-0.015 [†] (0.009)
Intercept	-0.269** (0.035)	-0.391** (0.037)	-0.379** (0.040)	-0.340** (0.042)

	Model 1	Model 2	Model 3	Model 4
Fixed Effect	Unstand. b	Unstand. b	Unstand. b	Unstand. b
Random effect				
E	0.635	0.635	0.634	0.634
e (intercept)	0.085	0.074	0.071	0.070
e (slope)	0.018	0.016	0.015	0.015
Γ	0.003	0.0001	0.0001	0.0001
R^2 e		.1310	.1668	.2212
R^2 E		.0000	.0001	.0003
Deviance	26701.763	26465.234	26439.902	26428.271
No. est. parameters	7	16	20	25

Note: Unstandardized coefficient with robust standard errors in parentheses. Neighborhood measures are grand centered; S5 was the reference group; N persons = 2088 and N neighborhoods = 232. RS, Residential stability; ND, neighborhood disadvantage.

\dagger $p < .10$.

* $p < .05$.

** $p < .01$. Two-tailed tests.