

## Gene-Environment Correlation and Interaction in Peer Effects on Adolescent Alcohol and Tobacco Use

K. Paige Harden · Jennifer E. Hill ·  
Eric Turkheimer · Robert E. Emery

Received: 2 November 2007 / Accepted: 4 March 2008 / Published online: 27 March 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** Peer relationships are commonly thought to be critical for adolescent socialization, including the development of negative health behaviors such as alcohol and tobacco use. The interplay between genetic liability and peer influences on the development of adolescent alcohol and tobacco use was examined using a nationally-representative sample of adolescent sibling pairs and their best friends. Genetic factors, some of them related to an adolescent's own substance use and some of them independent of use, were associated with increased exposure to best friends with heavy substance use—a *gene-environment correlation*. Moreover, adolescents who were genetically liable to substance use were more vulnerable to the adverse influences of their best friends—a *gene-environment interaction*.

**Keywords** Gene-environment interaction · Gene-environment correlation · Peer influence · Alcohol use · Adolescence · Substance use

Peer relationships are important contexts for adolescent development and socialization (Hartup 1986), including the development of alcohol and tobacco use, two health risk behaviors that are highly comorbid in adolescence (Istvan and Matarazzo 1984) and may share a common genetic etiology (Young et al. 2006). Affiliation with alcohol or tobacco using peers is one of the strongest correlates of adolescents' own use. This observation has led to

conclusions about the adverse effects of “peer pressure.” There are, however, major difficulties in concluding that resemblance between adolescent peers is indicative of a straightforward causal mechanism (Bauman and Ennett 1996; Kandel 1996).

First, adolescents do not randomly choose their friends. Rather, certain factors will increase the likelihood that an adolescent will choose a best friend who uses alcohol and tobacco. Both Cleveland et al. (2005) and Fowler et al. (2007b) found that adolescents' *genes* influenced their exposure to peers who used alcohol and cigarettes. This process, whereby genetic predispositions affect the likelihood of being exposed to environmental risks, is known as gene-environment correlation, or rGE (Rutter and Silberg 2002). Moreover, these same genetic factors may also influence adolescents' own use, thus accounting for the resemblance between adolescent peers. This kind of gene-environment correlation has been found to account for the similarity of best friends for smoking and drinking and for serious delinquency in the few extant genetically-informed studies (Hill et al. *in press*; Rowe et al. 1984).

Second, any causal effect of peers may be moderated by characteristics of the adolescent, such that some adolescents are more vulnerable or resilient to peer influence. Self-esteem, insecure attachment style, and previous drinking behavior have been shown to predict individual differences in the magnitude of peer effects (Allen et al. 2006; Duncan et al. 2005; Urberg et al. 2003). One unexplored possibility is that genetic factors related to alcohol and tobacco use may also influence vulnerability to peer influence. This process, whereby genetic predispositions affect one's vulnerability to environmental risks, is known as gene-environment interaction, or  $G \times E$  (Plomin et al. 1977).

In the current report, we present analyses of a genetically informative sample that simultaneously considered

---

Edited by Stacey Cherny.

---

K. P. Harden (✉) · J. E. Hill · E. Turkheimer · R. E. Emery  
Department of Psychology, University of Virginia,  
P.O. Box 400400, Charlottesville, VA 22904-4400, USA  
e-mail: paigeharden@virginia.edu

the role of peer influence, gene-environment correlation, and gene-environment interaction in the development of adolescent alcohol and tobacco use (hereafter, referred to as *substance use*). We fit a series of models that subsumed the following four hypotheses:

1. Genetic factors will have significant main effects on adolescents' substance use.
2. Genetic factors will influence exposure to peer substance use (rGE).
3. Even after controlling for genetic factors, exposure to peer substance use will predict adolescents' own substance use.
4. Genetic factors will influence adolescents' vulnerability to their best friends' substance use ( $G \times E$ ).

## Method

### Participants

Data were drawn from In-School survey collected for the National Longitudinal Study of Adolescent Health (AddHealth), a nationally-representative study of adolescent health and risk behaviors collected during the 1994–1995 academic year. Sampling for Add Health began with identification of all high schools in the United States that had at least 30 enrollees ( $N = 26,666$ ). Schools were stratified according to geographic region, urbanicity, school size or type, and racial composition. From these strata, a random sample of schools was selected. If the school did not include seventh or eighth grade, the study recruited students from the feeder middle school sending students to that high school. Overall, 79% of the schools selected agreed to participate (final sample  $N = 134$  schools). School population ranged from under 100 students to over 3,000 students. Ninety-six percent of the participating schools ( $N = 129$ ) agreed to have students ( $N = 90,118$ ) complete a confidential In-School survey during the 1994–1995 academic year.

The present study used an AddHealth sub-sample composed of same-sex sibling pairs (*Targets*) and their closest same-sex friend (*Best Friends*). Information concerning whether adolescents reported living with another adolescent between 11 and 20 years of age was gathered from school rosters. The final sub-sample of *Target* adolescents comprised 1,636 sibling pairs [241 monozygotic (MZ) twin pairs, 215 dizygotic (DZ) twin pairs, 681 full-sibling (FS) pairs, 217 half-sibling (HS) pairs, 100 cousin (CO) pairs, and 182 non-biologically related (NR) pairs]. All sibling pairs, regardless of biological relationship, were same-sex (833 female–female pairs, 803 male–male pairs). Degree of genetic relatedness varies by sibling pair type:

MZ twins share 100% of their genes, DZ twins and full siblings share 50%, half siblings share 25%, cousins share 12.5%, and non-related pairs share 0%. Twin zygosity was determined by matching 11 molecular genetic markers (Jacobson and Rowe 1999; Smolen and Hewitt 2003) and on the basis of self-report and responses to four questionnaire items concerning similarity of appearance and frequency of being confused for one's twin. Similar questionnaires have been utilized widely in twin research and have been repeatedly cross-validated with zygosity determinations based on DNA (Loehlin and Nichols 1976). Congruence between siblings' reports of physical similarity was necessary to definitively assign zygosity.

Because the current analysis used a sub-sample of sibling pairs, we did not use the sampling weights available with the AddHealth data, which are only appropriate for analyses designed to characterize the behavior of individual adolescents. In addition, there was considerable lack of overlap between the original probability sample, collected to make nationally representative estimates, and the sample collected specifically for genetic analyses. There is no available information to compute sampling weights for pairs that were not included in the probability sample (Chantala 2001), which includes 35.83% of all sibling pairs and 77.19% of sibling pairs who are not biological related. Consequently, sampling weights were not used in the current analysis. Regarding the representativeness of the sibling sample, Jacobson and Rowe (1999) compared the sociodemographic composition of the sibling pairs sub-sample to the full AddHealth sample and found negligible differences with regard to age, ethnicity, or maternal education.

All AddHealth participants were asked to nominate up to five same-sex and five opposite-sex friends, ranked by closeness of friendship. Data from the closest, same-sex friend were used in the current analysis, because previous research suggests that best friend characteristics are the most robust predictors of behavior (Jaccard et al. 2005). Data from the student who was nominated as the target's best friend were used regardless of whether this nomination was reciprocated or not. Adolescents were specifically directed not to nominate their siblings as best friends, thus the best friend relationship may be thought of as the adolescent's closest *extra-familial same-sex* social relationship. Of course, it is possible that some adolescents experienced their sibling, particularly their twin, or another adolescent of opposite-gender, as their most intimate relationships. Also, best friends were not necessarily the same age as the target; however, 51% of targets nominated best friends who were less than 1 year apart in age, and 98% of targets nominated best friends who were less than 3 years apart in age. The age difference between targets and best friends (range =  $-6$  to 5 years) was not

predictive of either targets' ( $r = 0.04$ ,  $P = 0.14$ ) or best friends' ( $r = 0.01$ ,  $P = 0.51$ ) substance use.

Of the 1857 same-sex sibling pairs, 450 pairs (24%) had a valid best friend nomination for both siblings, and 657 pairs (35%) had a valid best friend nomination for one sibling. The remaining 750 pairs (40%) had missing best friend data for both siblings in the pair. The most common reason for missing best friend data was nominating a friend who did not attend a high school included in the AddHealth study (60% of invalid nominations in the full AddHealth sample). Targets' substance use did not significantly predict whether the target had a valid peer nomination. Controlling for targets' own substance use, the likelihood of missing best friend data significantly increased with age in years (OR = 1.137; 95% CI = 1.056, 1.224), and was significantly lower in females than males (OR = 0.784; 95% CI = 0.618, 0.995). Of the 750 pairs with completely missing best friend data, 221 pairs were excluded from all further analyses, because they were also missing data on both siblings' substance use. The remaining pairs with incomplete best friend data were included in analyses, despite being uninformative about the relation between best friend and target use, because they were informative about the magnitude of genetic variance in target substance use. In the final sample of 1,636 sibling pairs, age ranged from 11.9 to 21.3 years ( $M = 16.1$ ,  $SD = 1.67$ , 25–75% = 14.9–17.5).

## Measures

During the In-School survey, targets and best friends reported how often in the last 12 months (on a 6-point Likert scale: *Never, Once or twice, Once a month or less, Two or three days a month, Once or twice a week, Three to five days a week, Nearly every day*) they had drunk beer, wine, or liquor; had gotten drunk; and had smoked cigarettes. The availability of direct peer report is the key advantage of the In-School AddHealth data; relying on adolescents' reports of their peers' behavior has been found to inflate estimates of similarity between adolescents and their peers (Kandel 1996). We considered the best friend report to be a characteristic of the target adolescent; that is, how exposed is the target adolescent to peer drinking and

smoking? The modal response for both targets and best friends for all items was *Never*; approximately 50% of adolescents reported never drinking alcohol and over 70% reported never smoking cigarettes or getting drunk. Previous exploratory and confirmatory factor analyses indicated that these three items were strongly indicative of a single standardized latent factor (Hill et al. *in press*). Factor scores combining these three items were estimated using the software program *Mplus* (Muthén and Muthén 1998–2004). Analysis of categorical variables in *Mplus* presumes that a set of thresholds imposed on a normal "latent response distribution" produces observed discrete values (Muthén 1983). Thus, the large proportion of people with scores in the *Never* category was modeled with relatively high estimates for the first thresholds of the substance use items. Flora and Curran (2004) showed that estimation of CFA models with categorical variables is robust to violations of the normality assumption for the latent response distribution. Target substance use factor scores had a mean near zero ( $M = -0.043$ ;  $SD = 0.860$ ; range =  $-0.726, 3.380$ ), as did best friend substance use factor scores ( $M = 0.006$ ;  $SD = 0.878$ ; range =  $-0.726, 3.410$ ). Correlations among targets, targets' best friends, siblings, and siblings' best friends, by sibling pair type, are shown in Table 1.

## Preliminary models

As preliminary analyses, we fit a model in *Mplus* that decomposed variation in target substance use into three parts: additive genetic ( $A$ ), shared environmental ( $C$ ), and non-shared environmental ( $E$ ). Target age and gender were also included in the model as statistical covariates of target substance use, to prevent bias in the estimation of genetic and shared environmental variance due to twin pairs, but not other sibling pair types, being necessarily the same age. (All sibling pair types were same-sex.) For a complete description of the classical twin model (or  $ACE$  model), see Neale and Cardon (1992). Parameter estimates for the full model are summarized in the left-hand columns of Table 2. The full model was compared to two nested models, in which the variance of either the additive genetic or the shared environmental component

**Table 1** Correlations among target, target's best friend, sibling, and sibling's best friend by sibling pair type

Relationship	Correlations by pair type					
	MZ	DZ	FS	HS	CO	NR
Target—Sibling	0.600	0.510	0.285	0.279	0.052	0.449
Target's BF—Sibling's BF	0.437	0.316	0.211	0.142	0.082	0.321
Target—Sibling's BF/Sibling—Target's BF	0.346/0.318	0.414/0.376	0.136/0.175	0.128/0.016	-0.213/0.368	0.370/0.509
Target—Target's BF/Sibling—Sibling's BF	0.461/0.382	0.457/0.317	0.304/0.382	0.644/0.244	-0.074/0.052	0.560/0.386

**Table 2** ACE models of target’s and best friend’s substance use

Parameter	Target substance use			Best friend substance use		
	Full	CE	AE	Full	CE	AE
A	0.36 (0.22, 0.49)	[0]	0.46 (0.40, 0.53)	0.37 (0.12, 0.61)	[0]	0.41 (0.30, 0.53)
C	0.08 (−0.01, 0.17)	0.25 (0.21, 0.30)	[0]	0.03 (−0.11, 0.16)	0.19 (0.12, 0.25)	[0]
E	0.28 (0.22, 0.35)	0.46 (0.42, 0.50)	0.25 (0.21, 0.30)	0.34 (0.21, 0.47)	0.54 (0.48, 0.61)	0.33 (0.23, 0.43)
$b_{sex}$	−0.11 (−0.19, −0.04)	−0.11 (−0.19, −0.04)	−0.12 (−0.19, −0.04)	−0.13 (−0.23, 0.04)	−0.13 (−0.22, −0.04)	−0.13 (−0.23, −0.04)
$b_{age}$	0.09 (0.07, 0.11)	0.09 (0.07, 0.11)	0.09 (0.07, 0.11)	0.11 (0.09, 0.14)	0.12 (0.09, 0.14)	0.11 (0.09, 0.14)
Model fit						
$\chi^2$	100.11	120.25	103.61	85.26	91.33	85.42
$\Delta\chi^2 (P)$	–	20.14 (<0.01)*	3.50 (0.06)	–	6.07 (0.01)*	0.18 (0.67)

Note: All models estimated using maximum likelihood in Mplus. For CE model, the variance of A was fixed to zero. For AE model, the variance of C was fixed to zero

\* Model fit is significantly worse than full model

was fixed to zero. Models were compared using differences in chi-square. The CE model, without additive genetic variance, fit significantly worse than the full model ( $\Delta\chi^2 = 20.14, P < 0.001$ ). In contrast, the AE model, without shared environmental variance, did not fit significantly worse than the ACE model ( $\Delta\chi^2 = 3.50, P = 0.06$ ).

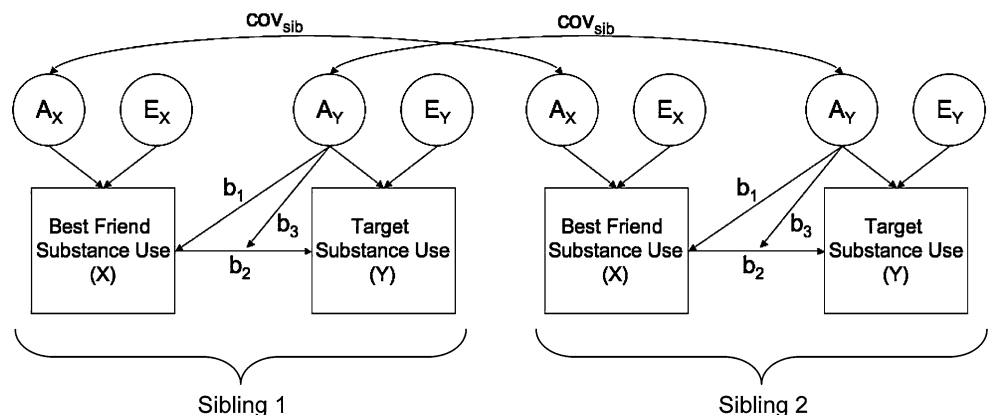
Next, we fit the same series of ACE models for best friend’s substance use, with best friend age and gender as statistical covariates. Results from these models are summarized in the right-hand columns of Table 2. Again, the CE model fit significantly worse than the full model ( $\Delta\chi^2 = 6.07, P = 0.01$ ), but the AE model did not ( $\Delta\chi^2 = 0.18, P = 0.67$ ). Overall, results from preliminary models replicate previous analyses of these data by Hill et al. (in press) and Cleveland et al. (2005). Results suggest that environmental factors shared by siblings in the same family are not a source of variation in adolescents’ substance use, as defined by the current study, or their affiliation with substance using best friends, thus the shared environmental component was dropped for subsequent models.

Interactive models: specification

We estimated an interactive model of relations between best friend’s and target’s substance use, as illustrated in Fig. 1. Although not shown, target and best friend gender, target age, and best friend age were also included in model as statistical controls. This model was designed to evaluate our four research questions:

1. Do genetic factors influence targets’ substance use? To the extent that genetic factors influence adolescents’ substance use, similarity between siblings should increase with genetic relatedness. MZ twins should be more similar than DZ twins and full siblings, and even more similar than half siblings, cousins, and non-related pairs. Variation in targets’ substance use into a component shared by siblings due to genetic factors ( $A_y$ ) and an environmental component unique to each individual sibling ( $E_y$ ) (Neale and Cardon 1992). Because of the results of preliminary analyses, a shared environmental component was not estimated.
2. Do genetic factors influence exposure to best friend substance use (rGE)? It is important to note that the data are silent regarding the genes of the best friends. All

**Fig. 1** Model of Relations between best friend and target substance use. Note: The covariance between genetic components ( $cov_{sib}$ ) was fixed such that the correlation equaled 1.0 in MZ pairs, 0.5 in DZ and FS pairs, 0.25 in HS pairs, 0.125 in CO pairs, and 0 in NR pairs. Target/best friend gender, target age, and best friend age were included as covariates but are not illustrated



references to genes or genetic factors refer to the genes of the *target* adolescent. Our model represents the extent to which target genetic factors influence exposure to best friend substance use in two ways:

(a) *Genetic factors related to target substance use.*

Imagine a sensation-seeking adolescent, who finds an alcohol “buzz” to be especially pleasurable, and who avoids other adolescents whom he or she perceives as boring. In this hypothetical example, the same genes (genes “for” sensation-seeking) are influencing the target’s substance use *and* influencing his or her selection of substance-using peers. The first representation of rGE in the model estimates this type of association, in which the genes related to target substance use also influence exposure to best friend substance use. Specifically, best friend substance use was regressed onto the genetic component of targets’ substance use ( $A_y$ ), as represented with the path,  $b_1$ .

(b) *Genetic factors not related to target substance use.*

Alternatively, there may be genetic influences on exposure to best friend substance use that do *not* overlap with genetic influences on target substance use. For example, imagine if a target’s genetic predisposition for lean body type increased the likelihood of her studying ballet very seriously. Because the adolescent was always in ballet class, this would decrease her exposure to best friends who smoked heavily. Consequently, she would have genes (“for” lean body type) that influence exposure to best friend substance use (always being around ballerinas who neither smoke nor drink), but that do not directly influence her own substance use (skinny people do not necessarily drink or smoke less). The second representation of rGE in the model estimates this type of association using the residual variance in best friend substance use. Residual variance in best friends’ substance use was divided into a component shared by siblings due to genetic factors ( $A_x$ ), and a component unique to each individual sibling ( $E_x$ ). Again, a shared environment component was not included, because of the results from preliminary models.

3. *After controlling for genetic factors, does best friend substance use predict target substance use?* Next, we included a regression of target substance use on best friend substance use,  $b_2$ . This represents the main effect of peer substance use, but the interpretation of this parameter is affected by the other parameters in the model. Because we have included gene-environment correlation in the model, the regression coefficient  $b_2$  estimates the main effect of best friend substance use *within pairs of siblings*. That is, if the best friend of sibling A drinks and smokes more heavily than the best friend of sibling B, does sibling A also smoke and drink more than sibling B? This within-sibling pair

association controls for all genetic factors shared by siblings, but also for all other environmental variables shared by siblings, including family structure, socioeconomic status, neighborhood, school environment, ethnicity, religious affiliation, parental alcohol and substance use, and genetic factors (Dick et al. 2000). It is, however, confounded by environmental variables that vary within siblings, thus we describe the within-sibling pair association as “quasi-causal.” Despite this limitation, the within-sibling pair association can provide stronger evidence of a causal relation than a comparison of unrelated persons.

4. *Do genetic factors influence adolescents’ vulnerability to their best friends’ substance use ( $G \times E$ )* Finally, the  $b_3$  path represents an interaction between target genetic factors ( $A_y$ ) and best friend substance use. Like any interaction, a significant  $b_3$  path would indicate that the effect of one covariate depends on the other, in this case, that the effect of best friend substance use depends on the target’s genetic liabilities. Also, a significant interaction effect would indicate that the magnitude of genetic variance in substance use is modified by exposure to best friends who also use.

#### Model estimation

This model was estimated using the Markov Chain Monte Carlo (MCMC) method via the software program WinBUGS (MRC Biostatistics Unit, University of Cambridge). WinBUGS implements the Gibbs sampling algorithm (Geman and Geman 1984) to iteratively simulate values for model parameters, given a specified prior distribution and an initial value for each parameter. The output of the Gibbs sampler constitutes a Markov chain. Under a wide set of conditions, the distribution of the Markov chain converges on the posterior distribution of parameters, i.e., on the distribution of parameters given the data (Gelman et al. 2003). The primary advantage of applying MCMC to the analysis of interactive behavior genetic models is that it makes the estimation of such a model computationally feasible; using traditional maximum likelihood methods, the models are almost always intractable (Eaves and Erkanli 2003).

Code and initial values for the full model are available from the first author upon request. WinBUGS was used to simulate a chain of 30,000 updates of the Gibbs sampler for each model. The first 20,000 iterations were discarded as overly influenced by initial values, and the remaining 10,000 were used to characterize the posterior distribution of the parameters of interest. The posterior distribution mean is reported as the point estimate for a given parameter; the 2.5%-ile and 97.5%-ile of distributions (95% *credible interval*) are reported to describe uncertainty about parameter values. Whether the distribution of the Markov

chain was stationary (i.e., whether the model had “converged”) was evaluated using basic diagnostic plots (traceplots, autocorrelation plots, and density plots). These diagnostic plots are also available from the first author upon request.

## Results

Results from the full model were consistent with all four of our hypotheses. First, there were significant genetic main effects on target substance use [ $A_Y = 0.422$  (95% bounds = 0.371, 0.479)], as well as significant non-shared environmental main effects [ $E_Y = 0.141$  (95% bounds = 0.114, 0.172)]. Second, target genetic factors—both related to targets’ own substance use and independent of targets’ own substance use—were correlated with exposure to best friend substance use. Specifically, the effect of genetic factors related to target substance use on exposure to best friend substance use ( $b_1$ ) was estimated as 0.205 (95% bounds = 0.104, 0.313), and there was significant genetic variance in best friend substance use due to target genes that were not related to the adolescent’s own drinking and smoking [ $A_X = 0.407$  (95% bounds = 0.284, 0.514)]. The remainder of the variance in exposure to best friend substance use was non-shared environmental [ $E_X = 0.376$  (95% bounds = 0.288, 0.497)]. Third, even after controlling for rGE, best friend use predicted target adolescent use ( $b_2 = 0.263$ ; 95% bounds = 0.210, 0.316), consistent with the quasi-causal hypothesis. Fourth, the effect of best friend substance use on target adolescents was stronger for adolescents with higher genetic liabilities ( $b_3 = 0.513$ ; 95% bounds = 0.459, 0.565).

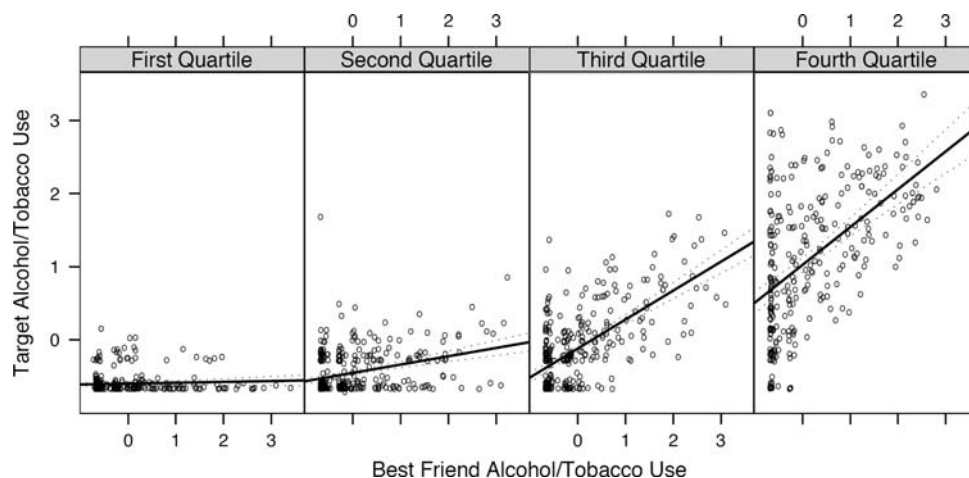
Regarding additional model parameters not directly relevant to the study hypotheses, the intercept for target substance use was estimated to be nearly zero ( $\mu_X = -0.053$ ; 95% bounds =  $-0.101$ , 0.004), as was the

intercept for best friend substance use ( $\mu_Y = 0.159$ ; 95% bounds =  $-0.060$ , 0.004). Gender did not reliably predict target use ( $b_{\text{sex}-y} = -0.053$ ; 95% bounds =  $-0.115$ , 0.006); however, female best friends had reliably lower substance use ( $b_{\text{sex}-x} = -0.115$ ; 95% bounds =  $-0.204$ ,  $-0.025$ ). Target age reliably predicted target use ( $b_{\text{age}-y} = 0.049$ ; 95% bounds = 0.032, 0.065). Likewise, best friend age reliably predicted best friend use ( $b_{\text{age}-x} = 0.111$ ; 95% bounds = 0.086, 0.136).

To further assess the contribution of  $G \times E$  to adolescent substance use, we fit a reduced model without  $G \times E$  ( $b_3 = 0$ ) and compared model fit using the Deviance Information Criterion, or DIC (Speigalhalter et al. 2002). Lower values of the DIC indicate better model fit, with differences greater than 10 DIC ruling out the model with higher DIC. Comparing Model 2 with Model 1, failure to include the effect of  $G \times E$  dramatically worsened model fit ( $\Delta\text{DIC} = 1,141$ ). Moreover, dropping  $G \times E$  biased other parameter estimates. Most notably, the main environmental effect of best friend substance use on target substance use was no longer evident ( $b_2 = -0.065$ ; 95% CI =  $-0.208$ , 0.087). Also, the effect of genetic liabilities related to target substance use on best friend substance use was overestimated ( $b_1 = 0.572$ ; 95% CI = 0.393, 0.700). It appears that, when not considering  $G \times E$ , the effect of best friend substance use on target substance use was obscured by gene-environment correlation.

To illustrate these results, we capitalized on WinBUGS’ capacity to estimate values for *all* unknowns, including genetic latent variables. The final (30,000th) update of the Gibbs sampler for the genetic latent variable ( $A_Y$ ) was used as an estimated genetic risk score. Estimated genetic risk scores were approximately normal, although positively skewed, with a mean near zero, and they were uncorrelated with target age ( $r = 0.02$ ). The estimated genetic risk scores were used to divide the sample into quartiles of genetic risk. Figure 2 shows scatterplots of best friend

**Fig. 2** Relation between best friend and target substance use by quartile of genetic risk. *Note:* Solid line represents regression line of best fit; broken lines represent 95% confidence intervals around regression line



substance use versus target substance use, separately for each quartile of genetic risk. For each plot, a regression line of best fit is drawn with a solid line. The differences among the regression lines across quartiles of genetic risk illustrate genotype-environment interaction. Adolescents with the highest genetic liability have the highest levels of substance use, even when their best friends report minimal substance use. Moreover, these high-risk adolescents demonstrate a sharp increase in substance use with increasing best friend substance use. The low-risk adolescents, in contrast, show minimal substance use, regardless of their best friends' behavior.

As mentioned previously, target and best friend age were statistically controlled in all models, and genetic risk estimates were independent of target age. Nevertheless, twin siblings and their best friends were necessarily matched for age, whereas there were within-pair age differences for all other sibling types. Consequently, we fit the full interactive model to a sub-sample comprised of only twin pairs ( $N$  pairs = 456), as a post-hoc test of whether age differences between siblings had biased the results described. Results from the twins-only model (listed in Appendix) remained consistent with all four research hypotheses, including a strong  $G \times E$  effect. In fact, the magnitude of the  $G \times E$  effect was estimated to be larger in the twin sample ( $b_3 = 0.677$ ), with a 95% credible interval that overlapped with the interval estimated in the sibling pairs sample by less than 0.1. Additionally, the estimate for genetic variance in target substance use was larger (0.422) in the twin sample compared to the sibling pairs sample (0.3). Differences in parameters between the twin and the sibling pairs samples were not formally tested, but we can speculate that analyses of non-twin siblings, who differ in the extent to which they have passed through the period of risk for substance use initiation, may underestimate both genetic main effects and gene-environment interaction effects.

## Discussion

The current study contributes to the understanding of peer effects on adolescent substance use in several respects. First, little research has considered the determinants of adolescent vulnerability to negative peer influences. We have demonstrated that vulnerability to peer influence, as measured by how well a best friend's substance use predicts an adolescent's use, is affected by some of the same genetic factors that contribute directly to substance use. Second, the current study reconciles apparently contradictory lines of research. Intervention research has repeatedly warned about the iatrogenic effects of aggregating adolescents who have histories of socially problematic

behavior in treatment groups (Poulin et al. 2001). However, genetically-informed studies that have controlled for genetic selection effects have failed to find any evidence for a causal effect of peers (Hill et al. *in press*). Results from our model that includes both  $G \times E$  and rGE indicate that the effects of best friends' tobacco/alcohol use are strongest for adolescents with the highest genetic liability—precisely the adolescents that are most likely to be involved in treatment groups. Treatment groups may be composed of adolescents who are most vulnerable to peer influence, whereas peer influence may be more innocuous in the general population. In addition, a model that failed to include  $G \times E$  yielded biased parameter estimates that underestimated the main effect of peers, whereas the effect of genetic selection was overestimated. Thus, prior genetically-informative research may have overlooked causal effects of peer association by collapsing across all levels of genetic risk.

The current study has three important limitations. First, peer influence is modeled as a unidirectional effect of 'best friends' on 'targets,' but adolescents were arbitrarily designated as targets of peer influence only because information on their siblings was available, not because we believe them to be passive receptacles for social experiences. Target adolescents, of course, are just as likely to be influencing their best friends as vice versa, as has been demonstrated with longitudinal research (Curran et al. 1997). The recognition of bidirectional peer effects introduces an alternative explanation for the similarity evident between best friends and targets at the highest level of genetic risk: Adolescents with genetic liabilities for substance use are more effective at influencing their best friends. Although substance use is considered deviant by adult authority figures (or, perhaps, *because* substance use is considered deviant), experience with smoking and drinking may be a "coveted social asset" (Moffitt 1993) and is associated with adolescent popularity (Allen et al. 2005). The adolescents at genetic risk for substance use, then, may wield considerable social influence. One methodology useful for discriminating between these hypotheses is to operationalize 'susceptibility' to peer influence independently from the cross-sectional similarity between best friends, such as the extent to which an adolescent changes his or her opinions to match the opinions of a best friend in an observed laboratory-based discussion task (Allen et al. 2006). Allen et al. (2006) found that susceptible adolescents were very similar to their peers for drug and alcohol use, whereas more dominant adolescents (i.e., those who did not change their opinions in discussion with their best friends) were less similar to their peers in drug and alcohol use, *and* they had lower levels of multiple forms of problem behavior. These results are consistent with our hypothesis that the greater similarity to best

friends evident for targets at higher genetic risk reflects increased social vulnerability, rather than increased social dominance. Further research that combines longitudinal and behavior genetic research designs with direct observation of adolescents' social interactions will be necessary to fully characterize the processes by which adolescents become similar to each other over time.

Second, adolescents' smoking and drinking were modeled with a single latent factor, because of the high observed covariation among substance use items. This measurement model is consistent with problem behavior theory (Jessor and Jessor 1977), which posits that diverse adolescent behaviors, including smoking, drinking, drug use, risky sexual behavior, and delinquency, are all reflections of a single underlying dimension of "unconventionality," or willingness to transgress societal norms. There are, however, etiological mechanisms that are specific to alcohol versus cigarette use. Kendler et al. (2007) demonstrated that alcohol and nicotine dependence in adults are influenced by both common and substance-specific genetic factors; 63% of the genetic variance in nicotine dependence was unique. Additionally, Dick and her colleagues found evidence for a gene-environment interaction between parental monitoring and genetic influences on adolescent smoking (Dick et al. 2007b), but not adolescent drinking (Dick et al. 2007a). Future research is necessary to understand the extent to which peer effects—and genetic vulnerabilities to peer influence—are substance-specific. Having a best friend who smokes, for example, may increase an adolescent's risk for smoking in particular, or for using substances in general.

Third, abstinence is modeled as existing on the lower end of a single liability dimension that extends to include heavy substance use at the other end. Previous research, however, has suggested that there may be qualitatively different risk factors for initiation versus progression of substance use, particularly for alcohol (Fowler et al. 2007a; Heath et al. 1991). Therefore, our results do not characterize etiological processes that are specific to initiation alone. Previous research has identified shared environmental and sibling effects on initiation (Heath et al. 1991; Koopmans et al. 1999; Pagan et al. 2006), and there were very small (and non-significant) shared environmental effects which were dropped from the current analyses. Given the likely existence of initiation-specific shared environmental processes, it is important not to overstate the importance of genetic processes in the overall development of adolescent substance use. This limitation, again, points to the need for longitudinal genetically-informed research, in order to characterize the etiological processes underlying various stages of substance use.

Overall, the current study begins to untangle the complex interplay between genes and the environment in the

development of adolescent substance use. Results encourage a more complex conception of the genetic basis for alcohol and tobacco use. It appears that genetic risks for adolescent substance use are, at least in part, made manifest via increased exposure to and vulnerability to negative influential environments, namely substance-using best friends. Similarly, an overly narrow focus on the environmental determinants of alcohol and tobacco use that fails to consider genetic differences in vulnerability to peers may paradoxically result in an underestimate of peer influence. Future research that rectifies some of the limitations of the current work, particularly the cross-sectional nature of the data, should continue to build a comprehensive model of how genetic liabilities and social experiences interact in adolescence.

**Acknowledgements** Preparation of this article was supported by grants from the National Institute on Drug Abuse (1F31DA023751) to Paige Harden and from the National Institute of Child Health and Human Development to Eric Turkheimer (1R01HD053550) and Robert Emery (1R01HD056354). The Add Health project was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies.

## Appendix

Parameter	Estimate	95% Interval	Credible
$A_y$	0.302	0.239, 0.379	
$E_y$	0.153	0.123, 0.187	
$A_x$	0.393	0.258, 0.533	
$E_x$	0.327	0.239, 0.436	
$b_1$	0.222	0.013, 0.431	
$b_2$	0.292	0.194, 0.391	
$b_3$	0.677	0.559, 0.881	
$b_{sex-y}$	-0.077	-0.176, 0.020	
$b_{sex-x}$	-0.044	-0.205, 0.124	
$b_{age-y}$	0.050	0.019, 0.083	
$b_{age-x}$	0.141	0.090, 0.193	

*Note:* Results from full interactive model fit to sub-sample of same-sex twin pairs only

## References

- Allen JP, Porter MR, McFarland CF (2006) Leaders and followers in adolescent close friendships: susceptibility to peer influence as a predictor of risky behavior, friendship instability, and depression. *Dev Psychopathol* 18:155–172
- Allen JP, Porter MR, McFarland CF, Marsh P, McElhaney KB (2005) The two faces of adolescents' success with peers: adolescent popularity, social adaptation, and deviant behavior. *Child Dev* 76:747–760
- Bauman KE, Ennett ST (1996) On the importance of peer influence for adolescent drug use: commonly neglected considerations. *Addiction* 91:185–198



- Chantala K (2001) Constructing weights to use in analyzing pairs of individuals from the Add Health data. Carolina Population Center, University of North Carolina, Chapel Hill. Accessed on June 24, 2007 at <http://www.cpc.unc.edu/projects/addhealth/files/pweights.pdf>
- Cleveland HH, Wiebe R, Rowe DC (2005) Genetic influences on associations with substance using peers. *J Genet Psychol* 166:153–169
- Curran PJ, Stice E, Chassin L (1997) The relation between adolescent alcohol use and peer alcohol use: a longitudinal random coefficients model. *J Consul Clin Psychol* 65:130–140
- Dick DD, Johnson JK, Viken RJ, Rose RJ (2000) Testing between-family associations in within-family comparisons. *Psychol Sci* 11:409–413
- Dick DM, Pagan JL, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ (2007a) Changing environmental influences on substance use across development. *Twin Res Hum Genet* 10:315–326
- Dick DM, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ (2007b) Parental monitoring moderates the importance of genetic and environmental influences on adolescent smoking. *J Abnorm Psychol* 116:213–218
- Duncan GJ, Boisjoly J, Kremer M, Levy DM, Eccles J (2005) Peer effects in drug use and sex among college students. *J Abnorm Child Psychol* 33:375–385
- Eaves L, Erkanli A (2003) Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and GxE interaction. *Behav Genet* 33:279–299
- Flora DB, Curran PJ (2004) An empirical evaluation of alternative models of estimation for confirmatory factor analysis with ordinal data. *Psychol Methods* 9:466–491
- Fowler T, Lifford K, Shelton K, Rice F, Thapar A, Neale MC, McBride A, van den Bree MBM (2007a) Exploring the relationship between genetic and environmental influences on initiation and progression of substance use. *Addiction* 101:413–422
- Fowler T, Shelton K, Lifford K, Rice F, McBride A, Nikolov I, Neale MC, Harold G, Thapar A, van den Bree MBM (2007b) Genetic and environmental influences on the relationship between peer alcohol use and own alcohol use in adolescents. *Addiction* 102:894–903
- Gelman A, Carlin JB, Stern HS, Rubin DB (2003) Bayesian data analysis, 2nd edn. CRC Press, London
- Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans Pattern Anal Mach Intell* 6:721–741
- Hartup WW (1986) Relationships and development. Erlbaum, Hillsdale, NJ
- Heath AC, Meyer J, Jardine R, Martin NG (1991) The inheritance of alcohol consumption patterns in a general population sample: II. Determinants of consumption frequency and quantity consumed. *J Stud Alcohol* 52:425–431
- Hill JE, Turkheimer E, Harden KP, Mendle J, Emery RE (in press) The relationship between adolescent externalizing and affiliation with externalizing peers: A behavior genetic approach. *J Abnorm Child Psychol*
- Istvan J, Matarazzo JD (1984) Tobacco, alcohol, and caffeine use: a review of their interrelationships. *Psychol Bull* 95:301–326
- Jaccard J, Blanton H, Dodge T (2005) Peer influences on risk behavior: an analysis of the effects of a close friend. *Dev Psychol* 41:135–147
- Jacobson KC, Rowe DC (1999) Genetic and environmental influences on the relationships between family connectedness, school connectedness, and adolescent depressed mood. *Dev Psychol* 35:926–939
- Jessor R, Jessor SL (1977) Problem behavior and psychosocial development: a longitudinal study of youth. Academic Press, New York
- Kandel DB (1996) The parental and peer contexts of adolescent deviance: an algebra of interpersonal influences. *J Drug Issues* 26:289–315
- Kendler KS, Myers J, Prescott CA (2007) Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry* 64:1313–1320
- Koopmans JR, Slutske WS, Heath AC, Neale MC, Boomsma DI (1999) The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behav Genet* 29:383–393
- Loehlin JC, Nichols RC (1976) Heredity, environment, and personality: a study of 850 sets of twins. University of Texas Press, Austin, TX
- Moffitt TE (1993) Adolescent-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol Rev* 100:674–701
- Muthén BO (1983) Latent variable structural equation modeling with categorical data. *J Econom* 22:48–65
- Muthén LK, Muthén BO (1998–2004) Mplus user's guide, 3rd edn. Muthén & Muthén, Los Angeles, CA
- Neale MC, Cardon LR (1992) Methodology for genetic studies of twins and families. Kluwer, Dordrecht, The Netherlands
- Pagan JL, Rose RJ, Viken RJ, Pulkkinen L, Kaprio J, Dick DM (2006) Genetic and environmental influences on stages of alcohol use across adolescence and into young adulthood. *Behav Genet* 36:483–497
- Plomin R, DeFries JC, Loehlin JC (1977) Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* 84:309–322
- Poulin F, Dishion TJ, Burraston B (2001) 3-year iatrogenic effects associated with aggregating high-risk adolescents in cognitive-behavioral preventive interventions. *Appl Dev Sci* 5:214–224
- Rowe DC, Osgood DW (1984) Heredity and sociological theories of delinquency: a reconsideration. *Am Sociol Rev* 49:526–540
- Rutter M, Silberg J (2002) Gene-environment interplay in relation to emotional and behavioral disturbance. *Annu Rev Psychol* 53:463
- Smolen A, Hewitt JK (2003) Chapter VI: DNA sampling, genotyping, and determination of zygosity. In: Biomarkers in Wave III of the AddHealth Study, Add Health User Guides, from <http://www.cpc.unc.edu/addhealth/files/biomark.pdf>
- Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A (2002) Bayesian measures of model complexity and fit. *J R Stat Soc Ser B* 64:134
- Urberg KA, Luo Q, Pilgrim C (2003) A two-stage model of peer influence in adolescent substance use: individual and relationship-specific differences in susceptibility to influence. *Addict Behav* 28:1243–1256
- Young SE, Rhee SH, Stallings MC (2006) Genetic and environmental vulnerabilities underlying adolescent substance use and problem use: General or specific? *Behav Genet* 36:603–615