Educational attainment, smoking initiation and lifetime nicotine dependence among male Vietnam-era twins

J. M. McCaffery[1]*, G. D. Papandonatos[2], M. J. Lyons[3,4,5], K. C. Koenen[6,7], M. T. Tsuang[4,5,8] and R. Niaura[9]

[1] *Centers for Behavioral and Preventive Medicine, Brown Medical School and The Miriam Hospital, Providence, RI, USA*

[2] *Center for Statistical Sciences, Brown Medical School, Providence, RI, USA*

[3] *Department of Psychology, Boston University, Boston, MA, USA*

[4] *Harvard Institute of Psychiatric Epidemiology and Genetics, Departments of Epidemiology and Psychiatry, Boston, MA, USA*

[5] *Harvard Medical School Department of Psychiatry at Massachusetts Mental Health Center, Boston, MA, USA*

[6] *Departments of Society, Human Development and Health and Epidemiology, Harvard School of Public Health, Boston, MA, USA*

[7] *Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA*

[8] *Institute of Behavioral Genomics, Department of Psychiatry, University of California, San Diego and VA San Diego Health Care System, San Diego, CA, USA*

[9] *Transdisciplinary Research Group, Butler Hospital and Brown Medical School, Providence, RI, USA*

Background. Smoking initiation and persistence are clearly associated with factors commonly thought to be environmental in origin, including socio-economic status. However, twin models that incorporate gene–environment correlation and gene × environment interaction have not been applied to elucidate the genetic or environmental role that socio-economic status plays in smoking initiation and nicotine dependence.

Method. Twin structural equation modelling was used to examine gene–environment correlation and gene × environ-environment interaction of one index of socio-economic status, educational attainment, with smoking initiation and nicotine dependence among 5119 monozygotic and 4295 dizygotic male–male Vietnam-era twins from the Vietnam Era Twin Registry, a national registry of twin pairs who served in the military during the Vietnam era.

Results. Educational attainment correlated significantly with smoking initiation ($r = -0.27$, $p < 0.001$). Additive genetic ($p = 0.011$), shared environment ($p = 0.002$) and unique environment ($p = 0.027$) components contributed to the correlation between educational attainment and smoking initiation. Educational attainment also significantly moderated the variance in smoking initiation ($p < 0.001$), suggestive of gene × environment interaction, although the interaction with the additive genetic, shared environmental and unique environmental components could not be resolved due to multi-collinearity. In contrast, educational attainment neither correlated with nor moderated nicotine dependence, once smokers had initiated.

Conclusions. Our study suggests that educational attainment is associated with smoking initiation, in part due to gene–environment correlation and gene × environment interaction. However, once smoking initiation is taken into account, there are no effects – be they gene–environment correlation or gene × environmental interaction – of educational attainment on nicotine dependence.

Introduction

Cigarette smoking is clearly associated with factors commonly thought to be environmental in origin, such

---

* Address for correspondence : J. M. McCaffery, Ph.D., Centers for Behavioral and Preventive Medicine, 196 Richmond Street, Providence, RI 02903, USA.

(Email: Jeanne_McCaffery@brown.edu)

A version of this paper was presented as an abstract at the 2005 Annual Meeting of the Society for Research on Nicotine and Tobacco in Prague, Czech Republic.

as socio-economic status (SES). For example, lifetime daily smoking is overrepresented among individuals in lower socio-economic strata (Matthews *et al.* 1989; Winkleby *et al.* 1990; Adler *et al.* 1994), as are smoking persistence and failed cessation (Kaprio & Koskenvuo, 1988; Escobedo *et al.* 1990; Pugh *et al.* 1991).

Twin studies, on the other hand, suggest that smoking initiation and nicotine dependence are each, in part, heritable (Shenassa *et al.* 2003). In the largest review of these data (Sullivan & Kendler, 1999), Sullivan and Kendler report a heritability of 56%

($a^2 = 0.56$) for smoking initiation, with environmental factors shared by twins and those unique to each individual accounting for 24% ($c^2 = 0.24$) and 20% ($e^2 = 0.20$) of the remaining variance, respectively. Further, smoking persistence, often used as a 'proxy measure of nicotine dependence', is more strongly heritable ($a^2 = 0.67$) and influenced by unique environment ($e^2 = 0.31$) with little or no contribution of shared environment ($c^2 = 0.02$). Familial transmission of nicotine dependence *per se* (Kendler *et al.* 1999; True *et al.* 1999; Lessov *et al.* 2004; Vink *et al.* 2005), as well as other proxy measures for nicotine dependence, including heavy smoking (Swan *et al.* 1997), total number of daily cigarettes (Hettema *et al.* 1999; Koopmans *et al.* 1999*a*) and failed smoking cessation (Xian *et al.* 2003), also appear to be primarily influenced by genetic and unique environmental effects, with little contribution of shared environmental factors. These differences in liability to smoking initiation and proxy measures of nicotine dependence, as well as nicotine dependence *per se*, have been confirmed in stage models in which progression of one phenotype is explicitly dependent upon the expression of an earlier phenotype with a potentially distinct liability (Heath *et al.* 1991, 2002; Heath & Martin, 1993; Koopmans *et al.* 1999*a*; Madden *et al.* 1999; Neale *et al.* 2006).

The differences in the effects of shared environment on smoking initiation *versus* nicotine dependence suggest that measured environmental factors, such as SES, may be differentially related to smoking initiation, relative to smoking persistence and progression to nicotine dependence. However, measured 'environmental' factors often show genetic influence when tested in genetically informative designs. For example, educational attainment, one of the components of individual SES most strongly associated with cigarette smoking, is often heritable when examined in twin studies (Heath & Berg, 1985). Thus, it is also possible that the association between socioeconomic indicators and smoking outcomes reflects gene–environment correlation, such that common genetic factors contribute to both SES and liability to smoking, or gene × environment interaction, as indicated by changes in the magnitude of the heritability of smoking initiation or nicotine dependence at various levels of SES.

Twin studies provide a unique opportunity to examine evidence for gene–environmental correlation and gene × environment interaction. Here, we employ twin structural equation models of gene–measured environment (educational attainment) correlation and gene × measured environment interaction to test whether: (1) educational attainment is more strongly associated with smoking initiation relative to nicotine dependence; (2) common environmental factors shared across twins predict both educational attainment and smoking initiation (shared environment–measured environment correlation); (3) common genetic factors predict both educational attainment and smoking initiation (gene–measured environment correlation); and (4) educational attainment modifies the heritability of smoking initiation and nicotine dependence (gene × education interaction). To our knowledge, this is the first study proposing to examine an indicator of SES, educational attainment, and cigarette smoking in the context of a twin design.

Method

Sample

Participants were 5119 monozygotic (MZ) and 4295 dizygotic (DZ) male–male Vietnam-era twins drawn from the Vietnam Era Twin (VET) Registry. The VET Registry is a nationally distributed cohort consisting of male–male twin pairs born between 1939 and 1957 in which both siblings served on active military duty during the Vietnam War era (Eisen *et al.* 1987). Zygosity was determined using a questionnaire and blood group typing methodology that achieved 95% accuracy (Eisen *et al.* 1987). Registry members are representative of all twins who served in the military during the Vietnam War on a variety of sociodemographic and other variables (Goldberg *et al.* 1987; Henderson *et al.* 1990). The data used in the present study were from the Survey of Health (1987), the National Heart, Lung and Blood Institute (NHLBI) survey (1990) and the Harvard Twin Study of Drug Abuse and Dependence (1991–1992). The current study was approved by the Miriam Hospital Institutional Review Board and procedures that were followed were in accordance with The Miriam Hospital Guidelines. All participants gave verbal informed consent at the time of the interviews.

Measures

Twin educational attainment (highest grade or year of school completed) was obtained from the Survey of Health (1987). Smoking initiation was collected as part of the NHLBI survey (1990). Smoking initiation reflected response to the question 'Have you smoked at least 100 cigarettes in your life?' Other demographic information was also taken from the NHLBI survey. Lifetime diagnosis of nicotine dependence was obtained using the Mental Health Diagnostic Interview Schedule Version III, revised (DIS-III-R; Robins *et al.* 1988) as part of the Harvard Drug Study (1991–1992). The DIS-III-R is a structured psychiatric interview for epidemiological research that yields clinical diagnoses
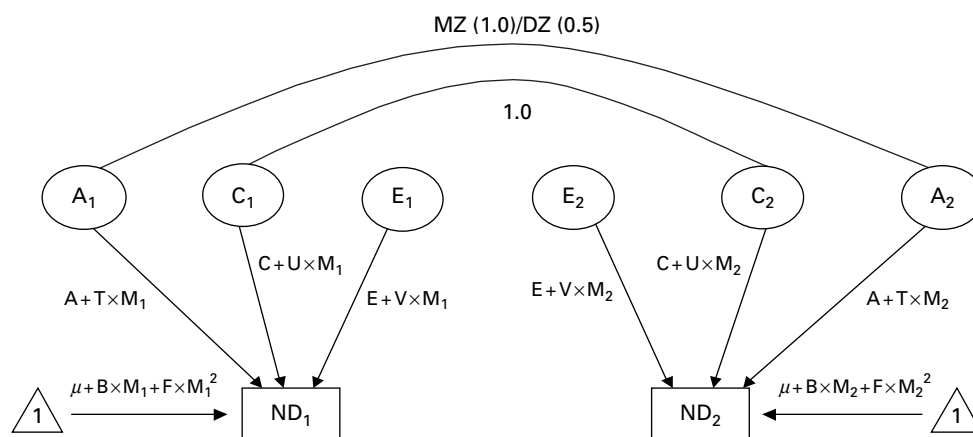
Fig. 1. Gene × environment interaction model adapted from Purcell (2002). MZ, Monozygote; DZ, dizygote; A, additive genetic effects; C, shared environmental effects; E, non-shared environmental effects; M, moderator; T, moderated component of A; U, moderated component of C; V, moderated component of E; B, F, linear and quadratic effects of moderator on mean (forced entry); ND, nicotine dependency.

based on the DSM-III-R (APA, 1987). Nicotine dependence was classified as no abuse/dependence (0–2 symptoms), mild (3–4 symptoms), moderate/severe (5–7 symptoms) based on DSM-III-R criteria. Details of the interview procedure have been reported previously (Koenen *et al.* 2002).

Statistical analyses

The primary method of analysis was twin structural modelling, which aims to explain the observed total phenotypic variation and covariation between MZ and DZ twins in terms of latent causes due to additive (A) or non-additive (D) genetic effects and shared (C) or unique (E) environmental effects. All models and maximum likelihood parameter estimates were calculated using the raw data capabilities of the Mx program (Neale *et al.* 2002). The significance of individual parameters was determined by comparing the fit of models omitting the parameters of interest with the fit of a full model, with twice the difference in the log-likelihood ratio ($\Delta - 2LL$) distributed asymptotically as a χ^2 variable with degrees of freedom (df) reflecting the difference in the number of parameters between the full and reduced models. However, for tests of the variance components, the corresponding p values were halved, since their value under the null hypothesis was on the boundary of the parameter space (Self & Liang, 1987). As there was little evidence that correlations among MZ twins substantially exceeded twice those among DZ twins, we focused primarily on ACE (*v.* ADE) models.

The polychoric correlation between the smoking phenotypes of interest and the putative environmental moderator, as well as the extent to which it is attributable to common additive genetic, shared environmental or unique environmental variance, may be estimated using bivariate twin modelling (Neale & Cardon, 1992). Using this approach, the correlation between additive genetic, shared environmental and unique environmental components of years of education and the smoking phenotypes may be quantified, the first specifically reflecting gene–measured environment correlation. Since the smoking phenotypes were ordinal in nature (smoking initiation, degree of nicotine dependence), and Mx cannot currently model polyserial correlations between ordinal and continuous data, polychoric correlations were obtained instead by discretizing years of education into five categories (6–11, 12, 13–14, 15–16, 17–20) chosen to retain information, while also ensuring adequate cell counts. Polychoric correlations between pairs of ordinal variables can be estimated based on a liability model (Falconer, 1965), which presumes an underlying, normally distributed susceptibility for the expression of the phenotype of interest with zero mean and unit variance. Categories were defined via thresholds on this underlying curve, such that the area under the curve reflects the proportion of the population in each category. To ensure convergence, the thresholds in the liability scale were assumed to be fixed across educational levels.

Gene × measured environment interaction may be detected within twin models by modelling the variance components attributable to latent genetic, shared and unique environmental effects as a function of the putative environmental moderator (Purcell, 2002). Briefly, both effects on the mean and on the variance are modelled (see Fig. 1). Modelling the effects of years of education on the mean of the liability distribution

avoids any potential confounding of gene × environment interaction by gene–environment correlation, by regressing out genetic and environmental effects common to both the smoking outcome and the moderator. Further, in order to avoid possible model mis-specification, we deliberately over-parameterized the model for the mean, by forcing both linear and quadratic effects of the moderator (e.g. $B \times M_i + F \times M_i^2$, where i reflects moderator values for individuals within a twin pair), irrespective of the statistical significance of the two regression coefficients. Variability in residual susceptibility to smoking initiation and nicotine dependence was described in terms of three latent variables, A, C and E, with the path coefficients associated with each variable expressed as linear functions of the moderator (e.g. $A + T \times M_1$, $C + U \times M_1$, $E + V \times M_1$). A significant deterioration of model fit when parameters T, U and V are fixed to zero was taken as evidence of significant moderation of the additive genetic, shared and unique environmental variance components, respectively. For example, a significant moderation of additive genetic variance alone would suggest that the magnitude of the heritability of nicotine dependence changes as the moderator increases or decreases. Variance components were only tested for significance if the respective interaction terms had been dropped from the model, e.g. A was not tested unless T was not significant, to avoid modelling interactions in the absence of main effects. Of note, it would have been possible to formulate models for testing for gene × environment interaction in the presence of gene–environment correlation (Purcell, 2002). However, we chose not to examine such models, since the joint distribution of the smoking and educational attainment liabilities is no longer bivariate normal, once educational attainment is allowed to moderate both common and specific additive genetic paths.

As noted by Heath *et al.* (2002), a full bivariate model cannot be fit to smoking initiation and progression to dependence when a binary initiation measure is employed, since the unique environment correlation is not identifiable. An ideal approach for analysing such data would have been based upon an ordinal measure of initiation with three or more categories, at least two of which had been observable among twins on which progression data were available. Categorizing smoking initiation as early *versus* late, based upon age of onset, would have served just such a purpose. Unfortunately, age of onset of smoking was available for only one-quarter of the VET Registry sample, despite the fact that about two-thirds of the sample had initiated smoking by the interview data. Therefore, in analysing nicotine dependence, we compared and contrasted analyses in both the full

sample and the subsample of initiated smokers in order to identify gene–environment correlation and gene × environment interaction effects on liability to nicotine dependence distinct from those seen for smoking initiation. Heath *et al.* (2002) showed that analysing data on initiated smokers alone captures all available information on genetic and environmental effects on nicotine dependence under an orthogonal liability model for smoking initiation and progression to dependence. Further, estimates of the variance components based on concordant-exposed twin pairs alone remain unbiased, even if a correlated liability model holds instead.

Although the full VET Registry comprises 12 436 male twins (6218 male–male twin pairs), only 5119 MZ twins and 4295 DZ twins provided information on at least one of the smoking phenotypes of interest. In bivariate models, only complete pairs contributed to the estimation of cross-twin correlations, although all available subjects were used to estimate means and variances. The sample size for the gene × measured environment analyses was smaller than that of the corresponding bivariate models, since Mx retained only twin pairs for which this subject-level covariate was measured on both members of the twin pair. Since the data on smoking initiation and nicotine dependence were collected as part of two different surveys, there were participants who had responded to one but not the other survey, resulting in participants with available smoking initiation data, but missing nicotine dependence data or vice versa. As information on smoking initiation could not be completely recovered from knowledge of nicotine dependence status (e.g. for those who did not meet criteria for nicotine dependence, we do not know whether they reached 100 cigarettes lifetime), no imputation was used.

To obtain confidence intervals (CIs) for parameter estimates, thresholds, and genetic and environmental correlations, bootstrapping methods were employed. Specifically, at each bootstrap iteration, pairs were drawn with replacement from the original sample of the same size. All runs in which Mx gave warning messages about possible lack of convergence were dropped, and the process was repeated until 1000 bootstrap iterations had converged successfully. For each parameter of interest, the estimates were ordered and endpoints of 95% bootstrap CIs were obtained from the 2.5 and 97.5 bootstrap sample percentiles. For comparison with the parameter estimates from Mx, the median of the bootstrap results was calculated as well. When there is little bias in the maximum likelihood estimates, one would expect the two to be in close agreement, and the 95% CIs to be well centred.

It should be noted that in twin structural equation modelling, it is assumed that rearing environment

Table 1. *Demographic and descriptive statistics for the VETR sample*

	Full sample (n = 9414, 5178 pairs)	Initiated smokers (n = 6949, 3929 pairs)
Age, years, mean (S.D.)	42.6 (2.8)	42.5 (2.8)
Education, years, mean (S.D.)	13.8 (2.0)	13.6 (2.0)
Race, *n* (%)		
Caucasian	8829 (93.8)	7404 (94.3)
African-American	542 (5.8)	419 (5.3)
Hispanic	5 (0.1)	5 (0.1)
Other	33 (0.3)	26 (0.3)
Married, *n* (%)	5174 (76.7)	3858 (77.0)
Smoking initiation, *n* (%)	5959 (69.1)	6949 (100.0)
Daily cigarettes, *n* (%)		
0	2993 (35.4)	330 (5.7)
1–10	924 (10.9)	924 (16.0)
11–20	2201 (26.1)	2201 (38.1)
21–30	1170 (13.9)	1170 (20.2)
31–40	771 (9.1)	771 (13.3)
41+	386 (4.6)	386 (6.7)
Nicotine dependence, *n* (%)		
None	4349 (58.2)	1355 (33.3)
Mild	1927 (25.8)	1682 (41.3)
Moderate/severe	1194 (16.0)	1037 (25.4)

VETR, Vietnam Era Twin Registry; S.D., standard deviation.

for the behaviours under study was similar for MZ and DZ twin pairs. Empirical tests of the equal environment assumption by zygosity suggest that it holds for many traits (Matheny, 1979; Scarr & Carter-Saltzman, 1979; Kendler *et al.* 1993). In addition, lack of assortative mating by phenotype is assumed.

Results

Sample and demographics

Demographic information and descriptive statistics for educational attainment, rates of smoking initiation and nicotine dependence are presented in Table 1. In the full sample, participants were on average 42.6 years of age at the NHLBI survey; 94% were Caucasian; 77% currently married. Participants had obtained an average of 13.8 years of education, consistent with slightly less than 2 years of college or technical school. Lifetime smoking measures suggested that nearly seven out of 10 participants had initiated smoking, about two-fifths showed at least mild

symptoms of nicotine dependence, and over a quarter smoked more than a pack of cigarettes per day. Initiated smokers differed little from the full sample in terms of age, education, race and marital status. As expected, they were more likely to exhibit at least mild symptoms of nicotine dependence (67% *v.* 42%) and to smoke more than a pack of cigarettes per day (40% *v.* 28%), relative to the full sample which included participants who did not initiate smoking in addition to those who did. Finally, for the gene × environment interaction analyses, the sample size was smaller due to the requirement of complete data on the covariate (educational attainment) for both twins of a pair for the individuals to enter into analyses. Missing data at the individual level was permitted for the dependent variables (the smoking phenotypes). This resulted in 8003 individuals, including 4459 MZ and 3544 DZ twins, available for gene × environment interaction analyses.

Univariate models

Standardized additive genetic, shared environmental and unique environmental components, total phenotypic variance and the prevalences within each of the ordinal categories for educational attainment, smoking initiation and nicotine dependence, respectively, are presented in Table 2. All parameter estimates were derived using Mx; CIs were derived via a bootstrap procedure. The median of the bootstrap results closely reflected the Mx parameter estimates in nearly all cases, suggesting that the maximum likelihood estimates from Mx showed little bias, and that the 95% bootstrap CIs were well centred.

In the full sample, educational attainment, smoking initiation and nicotine dependence were each significantly heritable, with the largest effects seen for smoking phenotypes. In addition, shared environment contributed significantly to educational attainment and smoking initiation. Among initiated smokers, nicotine dependence continued to show significant additive genetic effects with no evidence for shared environmental effects, while additive genetic and shared environmental effects contributed to educational attainment in this subgroup.

Bivariate models

The total polychoric correlation, and additive genetic, shared and unique environmental correlations between educational attainment and the smoking-related outcomes in both the full sample and among initiated smokers are listed in Table 3. In the full sample, there was a significant negative correlation between liability for educational attainment and

Table 2. *Univariate parameter estimates for education level, smoking initiation and nicotine dependence*[a]

	a² (95% CI)	c² (95% CI)	e² (95% CI)	Total variance (95% CI)	P1 (95% CI)	P2 (95% CI)	P3 (95% CI)	P4 (95% CI)	P5 (95% CI)
Full sample									
Education	0.29 (0.18–0.37)	0.36 (0.29–0.45)	0.35 (0.32–0.38)	2.37 (2.26–2.47)	0.05 (0.05–0.06)	0.32 (0.31–0.33)	0.36 (0.35–0.37)	0.16 (0.15–0.16)	0.11 (0.10–0.11)
Smoking initiation	0.49 (0.33–0.63)	0.29 (0.16–0.43)	0.22 (0.19–0.25)	1.47 (1.43–1.52)	0.30 (0.29–0.31)	0.70 (0.69–0.71)	–	–	–
Nicotine dependence	0.55 (0.40–0.61)	0.04 (0.00–0.17)	0.41 (0.38–0.45)	6.60 (5.71–6.86)	0.59 (0.58–0.60)	0.25 (0.24–0.26)	0.16 (0.15–0.17)	–	–
Initiated smokers									
Education	0.22 (0.06–0.40)	0.33 (0.18–0.47)	0.45 (0.39–0.50)	1.00 (0.95–1.05)	0.05 (0.05–0.06)	0.34 (0.33–0.35)	0.38 (0.36–0.39)	0.14 (0.13–0.15)	0.09 (0.08–0.09)
Nicotine dependence	0.35 (0.18–0.41)	0.00 (0.00–0.16)	0.65 (0.58–0.72)	0.80 (0.77–0.89)	0.34 (0.32–0.36)	0.41 (0.39–0.43)	0.25 (0.24–0.26)	–	–

CI, Confidence interval; a², additive genetic variance; c², shared environmental variance; e², non-shared environmental variance; P1–P5, prevalences within successive categories of education, smoking initiation and nicotine dependence (for education level, P1–P5 reflects 6–11, 12, 13–14, 15–16, 17–20 years of education; for smoking initiation, P1 and P2 reflect no initiation and initiation; for nicotine dependence, P1–P3 reflect no nicotine dependence, mild nicotine dependence and moderate or severe nicotine dependence).

[a] Parameter estimates from Mx modelling; 95% CIs from bootstrap simulation (included a minimum of 1000 runs without convergence warnings).

liability for both smoking initiation ($r = -0.27$, $p < 0.001$) and nicotine dependence ($r = -0.17$, $p < 0.001$), reflecting a greater prevalence of smoking initiation and nicotine dependence at lower levels of education. For smoking initiation and education level, all three liability components were statistically significant. Additive genetic effects alone contributed to the association between educational attainment and nicotine dependence ($r_A = -0.32$, $p < 0.001$). Among initiated smokers, the magnitude of the total polychoric correlation between educational attainment and level of nicotine dependence was greatly diminished and failed to attain statistical significance ($r = -0.05$, $p = 0.078$), despite the large sample size. None of its individual components was statistically significant either (all $p > 0.45$).

Gene × environment interaction models

In Table 4, comparative model fits are presented that test the extent to which educational attainment serves as a moderator of the smoking-related outcomes in both the full sample and among initiated smokers. A path diagram for the full model is depicted in Fig. 1. Of note, each model includes parameters B and F, representing linear and quadratic effects of educational level on the mean of the respective variables. These terms adjust the smoking outcomes for the correlation with educational level described in the bivariate models, allowing a model of gene × environment interaction for the variance components of the smoking phenotypes independent of any gene–environment correlations. The full model (ACE-TUV-BF) is first presented in each analysis, followed by an overall test of moderation of the variance components by educational attainment (T, U and V fixed to zero, resulting in an ACE model). If significant, a backwards stepwise elimination procedure is followed for the individual interaction parameters, testing the extent to which the parameter associated with the smallest change in log-likelihood ratio contributes significantly to the model. As mentioned earlier, variance components were only tested if the associated interaction term was dropped from the model.

Smoking initiation

We first examined the extent to which genetic and environmental contributions to liability to smoking initiation may be moderated by educational attainment. In comparative modelling fitting, there was evidence for significant moderation of the total variance by educational level. The backwards elimination procedure indicated that none of the moderation parameters could be dropped without a significant

Table 3. *Total polychoric correlation, and additive genetic, shared environmental and unique environmental correlations between ordinal education level, smoking initiation and nicotine dependence*

	Total correlation	r_A	r_C	r_E
Full sample				
Education and smoking initiation				
Initial model[a] (95% CI[b])	−0.27 (−0.30 to −0.24)	−0.30 (−0.57 to −0.07)	−0.39 (−0.65 to −0.17)	−0.10 (−0.19 to −0.01)
Education and nicotine dependence				
Initial model[a] (95% CI[b])	−0.17 (−0.20 to −0.14)	−0.32 (−0.59 to −0.11)	−0.39 (−0.95 to 0.33)	0.00 (−0.06 to 0.07)
Initiated smokers				
Education and nicotine dependence				
Initial model[a] (95% CI[b])	−0.05 (−0.09 to 0.00)	−0.21 (−0.80 to 0.23)	0.00 (−0.89 to 0.93)	0.02 (−0.06 to 0.11)

r_A, Additive genetic correlation; r_C, shared environmental correlation; r_E, unique environmental correlation; CI, confidence interval.

[a] Parameter estimates from Mx modeling.
[b] 95% CIs from bootstrap simulation (included a minimum of 1000 runs without convergence warnings).

compromise of model fit, although some evidence of multicollinearity was present, as there was little additional change in model fit if the T and U terms were dropped either jointly or individually, once V had been eliminated from the model. Over the 4 years reflecting the educational attainment of the majority of participants (12–16 years of education), the difference in educational attainment was associated with a reduction in total variance of 1.85 units in the liability scale (95% CI 1.75–2.32).

Nicotine dependence

For nicotine dependence in the full sample, jointly fixing the three moderation parameters to zero (ACE model) resulted in a significant deterioration of model fit, indicating significant moderation of total variance in nicotine dependence by years of education. In the backwards elimination procedure, only the moderation of the additive genetic component by educational attainment (T) contributed significantly to the model, indicating that educational attainment moderates the total variance in nicotine dependence, but that this effect is primarily driven by moderation of the additive genetic effects on nicotine dependence. In further modelling, it was also determined that the shared environment component did not contribute significantly to nicotine dependence, resulting in a final model of AE-T for nicotine dependence in the full sample.

However, among initiated smokers, the joint test for moderation of the total variance by educational attainment did not reach statistical significance ($\Delta-2LL=3.30$, df $=3$, $p=0.35$), indicating no gene × measured environment interaction effects on residual liability for nicotine dependence among initiated smokers. Further, none of the individual moderator parameters contributed significantly to model fit, indicating a lack of moderation of the additive genetic, shared environmental and unique environmental factors by educational attainment. Consistent with the full sample, shared environmental variance could also be dropped from the model, resulting in a final model of AEBF. Heritability at the mean level of education of 14 years was estimated at 0.34 (95% CI 0.26–0.40).

Discussion

The results of this study highlight the critical importance of phenotype definition and the potential for gene–environment correlation and gene × environment interaction in genetic studies of cigarette smoking. We focused on a well-documented 'environmental' risk factor for smoking, educational attainment, and uncovered several novel aspects of its association with cigarette smoking phenotypes using twin models of gene–environment correlation and interaction.

Table 4. *Comparative model fits for education level as a continuous moderator of smoking initiation and nicotine dependence*

	Model fit			Comparative model fit			
	Model[a]	-2LL	df	$\Delta-2$LL	df	p	Test
Full sample							
Smoking initiation							
1. Full	ACETUVBF	8888.53	7994				
2. No moderation	ACEBF	8910.32	7997	21.79	3	<0.001	2 *v*. 1
2a. No shared environmental moderation	ACETVBF	8904.78	7995	16.25	1	<0.001	2a *v*. 1
2b. No non-shared environmental moderation	ACETUBF	8910.05	7995	21.52	1	<0.001	2b *v*. 1
2c. No additive genetic moderation	ACEUVBF	8908.50	7995	19.97	1	<0.001	2c *v*. 1
3. Partial moderation							
3a. Shared environmental moderation	ACEUBF	8910.24	7996	21.71	2	<0.001	3a *v*. 1
3b. Non-shared environmental moderation	ACEVBF	8909.86	7996	21.33	2	<0.001	3b *v*. 1
3c. Additive genetic moderation	ACETBF	8910.12	7996	21.59	2	<0.001	3c *v*. 1
Nicotine dependence							
1. Full	ACETUVBF	12656.44	6914				
2. No moderation	ACEBF	12671.91	6917	15.47	3	0.001	2 *v*. 1
2a. No shared environmental moderation	ACETVBF	12657.33	6915	0.89	1	0.344	2a *v*. 1
2b. No non-shared environmental moderation	ACETBF	12658.60	6916	1.27	1	0.26	2b *v*. 2a
2c. No additive genetic moderation	ACEBF	12671.91	6917	13.31	1	<0.001	2c *v*. 2b
3. No shared environmental effects	AETBF	12659.49	6917	0.89	1	0.346	3 *v*. 2b
4. No non-shared environmental effects	ATBF	13700.05	6918	1040.56	1	<0.001	4 *v*. 3
5. No additive genetic moderation	AEBF	12672.48	6918	12.99	1	<0.001	5 *v*. 3
Initiated smokers							
Nicotine dependence							
1. Full	ACETUVBF	6388.90	2990				
2. No moderation	ACEBF	6392.19	2993	3.30	3	0.348	2 *v*. 1
3. No shared environmental effects	AEBF	6392.19	2994	0.00	1	0.999	3 *v*. 2
4. No non-shared environmental effects	ABF	6500.08	2995	107.89	1	<0.001	4 *v*. 3
5. No additive genetic effects	EBF	6469.80	2995	77.60	1	<0.001	5 *v*. 3

-2LL, Twice the negative log-likelihood; df, degrees of freedom.

[a] ACETUVB and F reflect model parameters as represented in Fig. 1. B and F reflect both linear and quadratic effects of the moderator on the mean and were forced into each model. A, C and E reflect additive genetic, shared environmental and non-shared environmental variance. T, U and V are path coefficients reflecting moderation of the A, C and E paths respectively by educational level.

With regard to gene–environment correlation, we observed a substantial inverse relationship between educational attainment and likelihood of smoking initiation. Twin modelling that decomposed the correlation into additive genetic, shared environmental and unique environmental effects indicated that each contributed significantly to this inverse relationship, with the strongest effects for the shared environmental and additive genetic correlations. Thus, educational attainment appeared to account not only for some of the environmental variance in smoking initiation, but also for a percentage of the heritability in smoking initiation, a result indicative of gene–environment correlation.

Although the significant correlation of educational attainment with level of nicotine dependence in the full sample was suggestive of potential gene–environmental correlation, modelling among initiated smokers revealed that the observed correlations in the full sample were primarily attributable to the association of educational attainment with smoking initiation, and not nicotine dependence *per se*. Specifically, the significant correlation between educational attainment and nicotine dependence was substantially diminished in the subset of initiated smokers. No additive genetic, shared environmental and non-shared environmental correlations between educational attainment and nicotine dependence remained significant among initiated smokers. Thus, correlations with educational attainment, be they genetically or environmentally mediated, appear to operate primarily at the level of smoking initiation.

A similar pattern of effects was observed in gene × measured environment interaction analyses that had been adjusted for potential gene–environment correlation. There was a substantial difference in the total variance of the liability for smoking initiation as a function of educational attainment, suggesting that novel additive genetic, shared environmental or unique environmental factors contribute to smoking initiation liability among persons with a lower level of education. Over the 4 years reflecting the educational attainment of the majority of participants (12–16 years of education), the increase in educational attainment was associated with a reduction in total variance of 1.85 units (95 % CI 1.75–2.32). Although twin modelling indicated that the interaction of educational attainment with additive genetic, shared environmental and unique environmental factors each contributed to the additional variance in liability to smoking initiation at lower levels of education, modelling results were suggestive of multicollinearity, so that the independent contribution of interactions with additive genetic, shared environmental and unique environment factors could not be delineated with certainty. Of note, the moderation of additive genetic, shared environmental and non-shared environmental variance by educational attainment could not be attributed to any correlation of educational attainment with smoking initiation.

Twin modelling of liability to nicotine dependence revealed no evidence of interaction with educational attainment, once smoking initiation was taken into account. In the full sample, there was significant moderation of the total variance in liability to nicotine dependence, which appeared to be driven primarily by an effect of educational attainment on additive genetic variance, or gene × educational attainment interaction. However, once smokers had initiated, there were no additional effects of educational attainment on the variance of nicotine dependence overall, or on the additive genetic variance component, suggesting that the effects of educational attainment on nicotine dependence operate primarily at the level of smoking initiation.

Our interaction results indicate increased variance in smoking initiation and the associated variance components at lower levels of education. Thus, a lower socio-economic environment, as indexed by relatively lower educational attainment, had a permissive effect on the expression of smoking initiation. Overall, these results strongly suggest that in the search for genes related to cigarette use, great care should be taken in selecting the phenotypes of interest. Further, depending on the phenotype, taking into account environmental factors, such as educational attainment or SES, may improve the ability to detect genetic effects.

To our knowledge, no prior study has examined the potential for gene–environment correlation or interaction in the association of educational attainment with smoking initiation and nicotine dependence. It is notable, however, that characteristics of family relationships, such as time spent with parents and parental monitoring, moderated the genetic effects on smoking frequency in adolescents (Dick *et al.* 2007). Several moderators of the genetic effects on alcohol use, a strong correlate of cigarette smoking, have also been identified. These include marital status (Heath *et al.* 1989), urban/rural geographic setting (Dick *et al.* 2001; Rose *et al.* 2001), religiosity (Koopmans *et al.* 1999*b*) and peer alcohol use (Dick *et al.* 2007). Interestingly, the results of these studies are inconsistent as to whether high-risk environments increase or decrease genetic variance in smoking and/or alcohol use. For example, time spent with parents was associated with greater genetic variance in frequency of smoking among adolescents; however, parental monitoring was associated with reduced genetic variance in frequency of smoking in the same sample (Dick *et al.* 2007).

It is important to note limitations to the present study. First, the VET Registry is predominantly Caucasian. The generalizability of these results to civilians, women and ethnic minorities remains to be determined. Furthermore, due to the ready access to cigarettes in the military in the Vietnam era, the rates of smoking initiation and nicotine dependence may be higher than other population-based samples with perhaps less permissive environments. In addition, Lyons *et al.* (2006), in a study of members of the VET Registry, found that veterans who served in Vietnam attained 0.7 fewer years of education by their 50s than veterans of the same era who did not serve in Vietnam.

We used educational attainment as a proxy for environmental context, as educational attainment is one of the components of SES most consistently associated with other indices of SES (Laaksonen *et al.* 2005) and cigarette smoking (Matthews *et al.* 1989; Escobedo *et al.* 1990; Winkleby *et al.* 1990, 1995, 1999; Escobedo & Peddicord, 1996; Laaksonen *et al.* 2005). However, educational attainment is also, in part, genetically influenced (Heath & Berg, 1985), consistent with our result of common genetic factors underlying educational attainment and risk for smoking initiation. Here, we accounted for gene–environment correlation by including the effects of educational attainment in the models for the means of the gene × environment interaction models. Lastly, as we included data from two different surveys of the Vietnam-era twins, we had a smaller sample size for the analyses of nicotine dependence, relative to smoking initiation,

particularly in the gene × environment interaction models.

Acknowledgements

Declaration of Interest

None.

References

Adler N, Boyce T, Chesney M, Cohen S, Folkman S, Kahn R, Syme S (1994). Socioeconomic status and health: the challenge of the gradient. *American Psychologist* **1**, 15–24.

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders, Revised*, 3rd edn. American Psychiatric Association: Washington, DC.

Dick DM, Pagan JL, Viken R, Purcell S, Kaprio J, Pulkkinen L, Rose RJ (2007). Changing environmental influences on substance use across development. *Twin Research* **10**, 315–326.

Dick DM, Rose RJ, Viken RJ, Kaprio J, Koskenvuo M (2001). Exploring gene–environment interactions: socioregional moderation of alcohol use. *Journal of Abnormal Psychology* **110**, 625–632.

Eisen S, True W, Goldberg J, Henderson W, Robinette CD (1987). The Vietnam Era Twin (VET) Registry: method of construction. *Acta Geneticae Medicae et Gemellologiae* **36**, 61–66.

Escobedo LG, Anda RF, Smith PF, Remington PL, Mast EE (1990). Sociodemographic characteristics of cigarette smoking initiation in the United States. Implications for smoking prevention policy. *Journal of the American Medical Association* **264**, 1550–1555.

Escobedo LG, Peddicord JP (1996). Smoking prevalence in US birth cohorts: the influence of gender and education. *American Journal of Public Health* **86**, 231–236.

Falconer DS (1965). The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics* **29**, 51–76.

Goldberg J, True W, Eisen S, Henderson W, Robinette CD (1987). The Vietnam Era Twin (VET) Registry: ascertainment bias. *Acta Geneticae Medicae et Gemellologiae* **36**, 67–78.

Heath AC, Berg K (1985). Effects of social policy on the heritability of educational achievement. *Progress in Clinical and Biological Research* **177**, 489–507.

Heath AC, Jardine R, Martin NG (1989). Interactive effects of genotype and social environment on alcohol consumption in female twins. *Journal of Studies on Alcohol* **50**, 38–48.

Heath AC, Martin NG (1993). Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addictive Behaviors* **18**, 19–34.

Heath AC, Martin NG, Lynskey MT, Todorov AA, Madden PA (2002). Estimating two-stage models for genetic influences on alcohol, tobacco or drug use initiation and dependence vulnerability in twin and family data. *Twin Research* **5**, 113–124.

Heath AC, Meyer J, Eaves LJ, Martin NG (1991). The inheritance of alcohol consumption patterns in a general population twin sample: I. Multidimensional scaling of quantity/frequency data. *Journal of Studies on Alcohol* **52**, 345–352.

Henderson WG, Eisen S, Goldberg J, True WR, Barnes JE, Vitek ME (1990). The Vietnam Era Twin Registry: a resource for medical research. *Public Health Reports* **105**, 368–373.

Hettema JM, Corey LA, Kendler KS (1999). A multivariate genetic analysis of the use of tobacco, alcohol, and caffeine in a population based sample of male and female twins. *Drug and Alcohol Dependence* **57**, 69–78.

Kaprio J, Koskenvuo M (1988). A prospective study of psychological and socioeconomic characteristics, health behavior and morbidity in cigarette smokers prior to quitting compared to persistent smokers and non-smokers. *Journal of Clinical Epidemiology* **41**, 139–150.

Kendler KS, Neale MC, Kessler RC, Heath, AC, Eaves LJ (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics* **23**, 21–27.

Kendler KS, Neale MC, Sullivan P, Corey LA, Gardner CO, Prescott CA (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine* **29**, 299–308.

Koenen K, Harney R, Lyons M (2002). A twin registry study of familial and individual risk factors for trauma exposure and posttraumatic stress disorder. *Journal of Nervous and Mental Disorders* **190**, 209–218.

Koopmans JR, Slutske WS, Heath AC, Neale MC, Boomsma DI (1999*a*). The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behavior Genetics* **29**, 383–393.

Koopmans JR, Slutske WS, van Baal GC, Boomsma DI (1999*b*). The influence of religion on alcohol use

initiation: evidence for genotype × environment interaction. *Behavior Genetics* **29**, 445–453.

Laaksonen M, Rahkonen O, Karvonen S, Lahelma E (2005). Socioeconomic status and smoking: analysing inequalities with multiple indicators. *European Journal of Public Health* **15**, 262–269.

Lessov CN, Martin NG, Statham DJ, Todorov AA, Slutske WS, Bucholz KK, Heath AC, Madden PA (2004). Defining nicotine dependence for genetic research: evidence from Australian twins. *Psychological Medicine* **34**, 865–879.

Lyons MJ, Kremen WS, Franz C, Grant MD, Thompson-Brenner H, Eisen S (2006). Vietnam service, combat and lifetime educational attainment: preliminary results from the Vietnam Era Twin Study of Aging (VETSA). *Journal of Research on Aging* **28**, 37–55.

Madden PA, Heath AC, Pedersen NL, Kaprio J, Koskenvuo MJ, Martin NG (1999). The genetics of smoking persistence in men and women: a multicultural study. *Behavior Genetics* **29**, 423–431.

Matheny Jr. AP (1979). Appraisal of parental bias in twin studies. Ascribed zygosity and IQ differences in twins. *Acta Geneticae Medicae et Gemellologiae* (*Roma*) **28**, 155–160.

Matthews KA, Kelsey SF, Meilahn EN, Kuller LH, Wing RR (1989). Educational attainment and behavioral and biologic risk factors for coronary heart disease in middle-aged women. *American Journal of Epidemiology* **129**, 1132–1144.

Neale M, Boker S, Xie G, Maes H (2002). *Statistical Modeling. VCU Box 900126*, 6th edn. Department of Psychiatry: Richmond, VA.

Neale M, Cardon LR (1992). *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic Publishers: Dordrecht.

Neale MC, Harvey E, Maes HH, Sullivan PF, Kendler KS (2006). Extensions to the modeling of initiation and progression: applications to substance use and abuse. *Behavior Genetics* **36**, 507–524.

Pugh H, Power C, Goldblatt P, Arber S (1991). Women's lung cancer mortality, socio-economic status and changing smoking patterns. *Social Science in Medicine* **32**, 1105–1110.

Purcell S (2002). Variance components models for gene–environment interaction in twin analysis. *Twin Research* **5**, 554–571.

Robins L, Helzer J, Cottler L, Goldberg J (1988). *National Institute of Mental Health Diagnostic Interview Schedule Version III – Revised*. Department of Psychiatry, Washington University: St Louis, MO.

Rose RJ, Dick DM, Viken RJ, Kaprio J (2001). Gene–environment interaction in patterns of adolescent drinking: regional residency moderates longitudinal influences on alcohol use. *Alcoholism Clinical and Experimental Research* **25**, 637–643.

Scarr S, Carter-Saltzman L (1979). Twin method: defense of a critical assumption. *Behavior Genetics* **9**, 527–542.

Self S, Liang K-Y (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.

Shenassa ED, McCaffery JM, Swan GE, Khroyan TV, Shakib S, Lerman C, Lyons M, Mouttapa M, Niaura RS, Buka SL, Leslie F, Santangelo SL (2003). Intergenerational transmission of tobacco use and dependence: a transdisciplinary perspective. *Nicotine and Tobacco Research* **5** (Suppl. 1), S55–S69.

Sullivan PF, Kendler KS (1999). The genetic epidemiology of smoking. *Nicotine and Tobacco Research* **1** (Suppl. 2), S51–S57; discussion S69–S70.

Swan GE, Carmelli D, Cardon LR (1997). Heavy consumption of cigarettes, alcohol and coffee in male twins. *Journal of Studies on Alcohol* **58**, 182–190.

True WR, Xian H, Scherrer JF, Madden PA, Bucholz KK, Heath AC, Eisen SA, Lyons MJ, Goldberg J, Tsuang M (1999). Common genetic vulnerability for nicotine and alcohol dependence in men. *Archives of General Psychiatry* **56**, 655–661.

Vink JM, Willemsen G, Boomsma DI (2005). Heritability of smoking initiation and nicotine dependence. *Behavior Genetics* **35**, 397–406.

Winkleby MA, Fortmann SP, Barrett DC (1990). Social class disparities in risk factors for disease: eight-year prevalence patterns by level of education. *Preventive Medicine* **19**, 1–12.

Winkleby MA, Robinson TN, Sundquist J, Kraemer HC (1999). Ethnic variation in cardiovascular disease risk factors among children and young adults: findings from the Third National Health and Nutrition Examination Survey, 1988–1994. *Journal of the American Medical Association* **281**, 1006–1013.

Winkleby MA, Schooler C, Kraemer HC, Lin J, Fortmann SP (1995). Hispanic versus white smoking patterns by sex and level of education. *American Journal of Epidemiology* **142**, 410–418.

Xian H, Scherrer JF, Madden PA, Lyons MJ, Tsuang M, True WR, Eisen SA (2003). The heritability of failed smoking cessation and nicotine withdrawal in twins who smoked and attempted to quit. *Nicotine and Tobacco Research* **5**, 245–254.