

Exploring the inter-relationship of smoking age-at-onset, cigarette consumption and smoking persistence: genes or environment?

KATHERINE I. MORLEY^{1,2*}, MICHAEL T. LYNSKEY^{1,3}, PAMELA A. F. MADDEN³,
SUSAN A. TRELOAR¹, ANDREW C. HEATH³ AND NICHOLAS G. MARTIN¹

¹ Genetic Epidemiology Group, Queensland Institute of Medical Research, Queensland, Australia; ² School of Population Health, The University of Queensland, Queensland, Australia; ³ Department of Psychiatry, Washington University St Louis, St Louis, MO, USA

ABSTRACT

Background. We investigated the genetic and environmental contributions to covariation between smoking age-at-onset, cigarette consumption and smoking persistence.

Method. Multivariate biometrical modelling methods were applied to questionnaire data from Australian twins and their siblings (14 472 individuals from 6247 families). The contributions of genetic and environmental factors to covariation between the three traits were estimated, allowing for sex differences in both trait prevalence and the magnitude of genetic and environmental effects.

Results. All traits were moderately heritable in males and females (estimates between 0.40 and 0.62), but there were sex differences in the extent to which additive genetic influences were shared across traits. Twin-specific environmental factors accounted for a substantial proportion of the variance in smoking age-at-onset in females (0.19) and males (0.12), but had little influence (<0.08) on other traits. Unique environmental factors were estimated to have a moderate influence on smoking age-at-onset (0.17 for females, 0.19 for males), but a stronger influence on other traits (between 0.39 and 0.49).

Conclusions. These results provide some insight into observed sex differences in smoking behaviour, and suggest that searching for pleiotropic genes may prove fruitful. However, further work on phenotypic definitions of smoking behaviour, particularly persistence, is warranted.

INTRODUCTION

Tobacco use makes a significant contribution to global burden of disease. In 2000, approximately 4.83 million deaths worldwide were attributable to tobacco smoking, accounting for 12% of the estimated global adult mortality for that year (Ezzati & Lopez, 2003, 2004). In Australia, tobacco smoking is the single most influential risk factor for disease burden, responsible for roughly 12% and 7% of total

disease burden in men and women respectively (Mathers *et al.* 2001). Thus tobacco use remains an important public health issue, despite the decline in smoking prevalence associated with the implementation of tobacco control programmes since the 1980s (White *et al.* 2003).

Epidemiological research indicates an inter-relationship between different aspects of smoking behaviour. Studies suggest that age at smoking initiation is related to subsequent aspects of smoking behaviour, such as cigarette consumption, nicotine dependence and smoking cessation (Breslau & Peterson, 1996; Hymowitz *et al.* 1997; Everett *et al.* 1999; Khuder *et al.* 1999; Lando *et al.* 1999). In general, such

* Address for correspondence: Ms Katherine Morley, Genetic Epidemiology Group, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Brisbane, Queensland 4029, Australia.
(Email: kateM@qimr.edu.au)

studies have shown that the younger an individual initiates smoking, the greater their cigarette consumption and risk of nicotine dependence, and the lower their likelihood of quitting. Thus some risk factors may influence more than one stage of smoking progression. What currently remains unclear is whether those risk factors are genetic or environmental in origin.

Twin studies demonstrate that both genetic and environmental factors influence variability in smoking behaviours (Sullivan & Kendler, 1999; Li, 2003). While some of these influences will be specific to certain smoking behaviours, many will be shared across traits. Multivariate twin analyses have shown some overlap between genetic and environmental factors influencing smoking initiation and persistence (Heath, 1990; Heath & Martin, 1993; Madden *et al.* 1999, 2004; Heath *et al.* 2002), initiation and cigarette consumption (Koopmans *et al.* 1999), and initiation, regular tobacco use and nicotine dependence (Kendler *et al.* 1999; Maes *et al.* 2004).

However, twin analyses of smoking are complicated by the fact that all aspects of smoking behaviour are observed conditional on the subject initiating smoking; we cannot observe nicotine dependence in those who have never smoked (for discussion see Neale *et al.* 2006*a, b*). Heath *et al.* (2002) developed a two-stage model of substance use that overcomes this limitation, incorporating non-smokers into analyses in a methodical manner. This model requires a measure of smoking initiation for the first 'stage', with other smoking-related traits making up the second (and subsequent) stages. Initiation is defined as a multiple category variable, with one category being non-smokers, having a single underlying distribution of liability. The second dimension phenotype can be considered Missing At Random (Little & Rubin, 1987) for non-smokers; as they are below the threshold for initiation, their status for the second dimension variable is unknown (Heath *et al.* 2002; Neale *et al.* 2006*b*). Using this model, the strength of genetic and environmental correlations between different stages of smoking can be estimated.

Heath and colleagues fitted this model to data on smoking age-at-onset and persistence from Australian adult twins. We have extended this

original model in several ways. First, we used a significantly larger sample. Second, we incorporated data from both twins and their siblings. Importantly, we also extended the range of behaviours assessed to look at three distinct aspects of smoking, age-at-onset, cigarette consumption, and smoking persistence, all of which have been shown to be inter-related. The aim of this study was to elucidate the interactions between these phenotypes, which have been observed previously in epidemiological studies, and to shed light on shared sources of risk.

METHOD

Subjects

Phenotypic measures of smoking behaviour were taken from four complementary self-report questionnaires mailed to adult twins and their family members between 1989 and 1993. Twins were recruited in two cohorts from the Australian National Health and Medical Research Council (NH&MRC) Twin Registry, and were asked to provide the contact details of other family members who would also be willing to participate. All studies were approved by the Queensland Institute of Medical Research Human Research Ethics Committees and the Australian Twin Registry. In total, 21 222 participants responded to the questionnaires, of which 5321 were twin pairs, 986 single twins and 3715 siblings.

Individual response rates for the twins were 81% for the first (older) cohort and 84% for the second (younger) cohort. Response rates were 60% and 56% for relatives of the older and younger cohorts respectively, although there was some variation in response rate depending upon the type of relative (Lake *et al.* 2000). As this questionnaire represented a follow-up survey for twins from the older cohort (previously surveyed in 1981), some information was available regarding the characteristics of non-responders. Current or ex-smoking as assessed in 1981 predicted non-response to the 1989 survey for males only. Non-response to the 1989 survey was not predicted by quantity smoked or smoking persistence in 1981 for either sex (Heath *et al.* 1995). Similar information was not available for the younger twin cohort or relatives of either cohort. Zygosity was

determined using standard questionnaire methods, validated by genotyping subsets of twins. More detailed descriptions of these studies can be found elsewhere (Heath & Martin, 1994; Kirk *et al.* 2000; Lake *et al.* 2000).

Phenotypic measures

Participants answered a number of questions about their own smoking behaviour and the smoking behaviour of their family members. They classified themselves and other family members as non-smokers, ex-smokers or current smokers. They also reported their average number of cigarettes consumed per day, either currently or when they previously smoked if an ex-smoker (scaled as never smoker, 1–4 cigarettes, 5–10, 11–20, 21–40, 40+), the ages at which they began and had successfully quit smoking (if they were an ex-smoker), and the total number of years for which they had smoked. Responses to the different questions were used to verify that participants' answers were internally consistent.

Three measures of smoking behaviour were defined using the questionnaire data. Following Heath *et al.* (2002), the smoking initiation dimension was defined according to self-reported age-at-onset of smoking. As most users begin in adolescence, we categorized smokers as those who began smoking by age 18 (standard onset) and those who began smoking after age 18 (late onset). Assuming a single underlying distribution of liability, these categories were ordered as never smoker, late-onset smoker, and standard-onset smoker, as those who begin smoking early are more likely to experience difficulty quitting (Lando *et al.* 1999; Lewinsohn *et al.* 1999). Average cigarette consumption was defined as a five-category variable, including all categories from the questionnaire except 'never smoker'. The third dimension, smoking persistence, was defined as whether or not the participant was a self-reported current or ex-smoker at the time of survey.

Structural equation modelling

All categorizations of smoking behaviour were treated as ordinal phenotypes, using a liability-threshold model. This model assumes that liability to a trait is normally distributed, with the boundaries between categories representing arbitrary thresholds along the distribution.

Biometrical modelling of threshold traits is discussed in more detail in Neale & Cardon (1992). In brief, classical twin analysis permits variance in liability to an ordinal trait to be decomposed into the following latent sources: additive genetic (A), non-additive genetic (D), environment shared by family members (C), and environment unique to each family member (E). However, C and D are confounded in analyses consisting of only twins reared together, and thus only one of these parameters can be estimated within a model (Grayson, 1989; Hewitt, 1989).

The standard twin design can be extended to include information from additional non-twin siblings, which are parameterized as for dizygotic (DZ) twins (Posthuma *et al.* 2003). When non-twin siblings are included, the influence of environmental factors specific to twin pairs (T) can also be estimated (Koeppen-Schomerus *et al.* 2003). Differences between males and females in the strength of latent factors can be determined by fitting a common-effects sex-limitation model, which permits the magnitude of genetic and environmental influences to differ between the sexes (Neale & Cardon, 1992; Medland, 2004; Neale *et al.* 2006*c*).

Multivariate biometrical analyses can be used to partition the covariation between phenotypes, in addition to partitioning the variance within a phenotype. In the three-stage model, a series of latent factors is assumed to explain the variance and covariance of a series of traits. The three-stage model we fitted consisted of three latent factors from four sources of variance (A, C, E and T). The first set of latent factors was hypothesized to influence smoking age-at-onset, and to explain part of the variance in the two remaining phenotypes (cigarette consumption and smoking persistence). The second set of latent factors was hypothesized to explain the remaining variance of the cigarette consumption phenotype and part of the variance in smoking persistence. The third set of latent factors was hypothesized to explain the variance in smoking persistence not already accounted for by the first two groups of latent factors.

To investigate the inter-relationship of the smoking phenotypes, we fitted a common-effects sex-limitation three-stage model including smoking age-at-onset, cigarette consumption, and smoking persistence. Model fitting was

undertaken in a stepwise manner. Prior to biometrical model fitting, a number of assumptions regarding the data were formally tested.

First, the fit of a single liability dimension (SLD) model to the age-at-onset variable was tested. This can be achieved using a contingency table script, fitted to the data using Mx, as outlined by Heath *et al.* (2002). However, when sibling data are included in the analysis, this provides an imperfect method as families with incomplete data will not be included (e.g. a family consisting of a twin and a sibling will not contain any of the requisite pairings). Consequently, a Contingent Causal Common pathway (CCC) model (Kendler *et al.* 1999) with sex-specific thresholds was fitted to the first dimension variable. For the CCC model, the age of initiation variable was split into two binary variables: initiation of regular smoking, and late *versus* standard smoking onset. This model is a constrained multivariate model that assumes that genetic and environmental factors influencing initiation only affect age-at-onset via a single common pathway. As such a model with the causal path constrained to unity indicates that all genetic and environmental influences are shared between the two traits, the estimate of the causal path can be used to determine whether the three-category definition of age-at-onset has a single distribution of liability (Neale *et al.* 2006*b*).

Second, assumptions regarding the homogeneity of phenotype prevalence and correlations for smoking age-at-onset were tested. A basic model incorporating a threshold model for smoking age-at-onset (including corrections for year of birth, sex, and an interaction between the two) and a calculation of polychoric correlations between different pairings of twins and siblings was fitted to the data. Thresholds were specified separately by sex for both twins and siblings. Correlations were allowed to vary for each zygosity group and between twin–twin, twin–sibling and sibling–sibling pairs. These model parameters were equated progressively within and between zygosity groups until the fit of the model worsened significantly. The appropriateness of including the regression terms in the threshold model was also tested by successively dropping each term from the model until the fit worsened significantly.

Table 1. *The number of families, and their structures, for individuals with relevant phenotypes*

Zygosity	Family structure						Total
	T	T-S	T-S-S	T-T	T-T-S	T-T-S-S	
MZF	156	33	11	929	407	244	1780
MZM	165	19	8	478	185	129	984
DZF	162	26	14	602	217	177	1198
DZM	158	28	19	346	110	86	747
DZOS	308	57	25	684	293	171	1538
Total	949	163	77	3039	1212	807	6247

T, Twins; S, siblings; MZ, monozygotic; DZ, dizygotic; F, female; M, male; OS, one female, one male twin.

A common-effects sex-limitation univariate model was fitted to smoking age-at-onset to assess which parameters to include in the full multivariate model. To determine the appropriate threshold models for cigarette consumption and smoking persistence in the multivariate model, the percentage of individuals in each category was estimated separately for male and female twins and siblings (including those twins from opposite-sex pairs) using Stata version 8.2 (StataCorp, 2005).

Model fitting was conducted with twins and a maximum of two additional siblings. All models were fitted to the raw data by the method of maximum likelihood as implemented in Mx (Neale *et al.* 2003). The significance of the different latent factors was assessed by using the likelihood-ratio χ^2 test to compare the fit of a model including the factor to one in which the factor was not estimated (Neale & Cardon, 1992). While estimates of the confidence intervals (CIs) for the multivariate model parameters would be desirable, this was not possible as the large sample size and complexity of the model result in prohibitively large computation time.

RESULTS

Sample demographics and phenotypic measures

For this study, analyses were conducted with a maximum of four siblings (parent, offspring and spouse information was not used). Relevant phenotypes were available for 14 472 individuals from 6247 families. Family structures and sizes are shown in Table 1.

Table 2. Percentages (with standard errors) of male and female twins and siblings in each of the phenotypic categories

Phenotype	Female twins	Female sibs	Male twins	Male sibs
Smoking age-at-onset				
Non-smoker	56.4 (0.6)	57.0 (1.2)	51.2 (0.8)	49.4 (1.4)
>18	8.9 (0.3)	10.1 (0.7)	8.6 (0.4)	8.5 (0.8)
≤18	34.7 (0.6)	33.0 (1.1)	40.2 (0.7)	42.1 (1.3)
Average daily cigarettes				
1-4	25.5 (0.8)	26.3 (1.6)	19.0 (0.9)	19.9 (1.5)
5-10	24.5 (0.8)	26.3 (1.6)	18.5 (0.8)	16.1 (1.4)
11-20	33.5 (0.9)	31.9 (1.7)	36.5 (1.0)	35.8 (1.8)
21-40	15.8 (0.7)	14.0 (1.3)	23.3 (0.9)	24.2 (1.6)
40+	0.8 (0.2)	1.5 (0.4)	2.8 (0.4)	4.1 (0.8)
Smoking persistence				
Ex-smoker	42.5 (0.9)	54.7 (1.8)	41.0 (1.1)	54.9 (1.9)
Current smoker	57.5 (0.9)	45.3 (1.8)	59.0 (1.1)	45.1 (1.9)

Females made up 60.6% of the sample. The distribution of year of birth was similar for males and females: 50% of women were born in 1958 onwards (range 1902–1974) and 50% of men were born in 1961 onwards (range 1903–1973). The prevalence of current smoking (as compared to non- and ex-smokers) in this sample was 24% for women and 28% for men, and similar to estimates from Australian general population-based samples during the same time period (1989–1992) (White *et al.* 2003). The proportion of males and females in the different phenotypic categories, shown separately for twins and siblings, is detailed in Table 2.

Overall, women were less likely to report ever having smoked: 43.5% reported themselves to be ex- or current smokers compared to 49.2% of men. The proportions of twins and siblings reporting ever having smoked were similar for females and for males. Average daily cigarette consumption was also similar between twins and siblings for each sex. However, the prevalence of current *versus* ex-smoking among twins was significantly different from that among siblings for both males and females ($p < 0.0001$ in both cases). Siblings were more likely to be ex-smokers, with approximately 55% of male and female siblings reporting this as their current smoking status. In comparison, only 43% of female twins and 41% of male twins reported themselves to be ex-smokers.

Testing model assumptions

The assumption that smoking initiation, as defined by age of onset, represented a single

underlying distribution of liability was tested using the CCC model. The estimate of the causal path for this model was very close to unity (point estimate 0.92, 95% CI 0.91–0.94). The saturated model for assumption testing for smoking age-at-onset included separate sex-specific thresholds for twins and siblings for each zygosity group. For twins, prevalence estimates were not significantly different for monozygotic (MZ) and DZ twins, for either sex ($\chi^2_4 = 7.92$, n.s.). The prevalence estimate for females from the DZ opposite-sex pairs was also not significantly different from that of the other female twins ($\chi^2_2 = 0.36$, n.s.). However, there was a significantly lower proportion of non-smokers among the males from the DZ opposite-sex group (0.46), as compared to other male twins (0.54) ($\chi^2_2 = 18.58$, $p < 0.001$).

Male and female prevalence estimates for the siblings could be equated to those of the twins (excluding the male twins from DZ opposite-sex pairs) without a significant loss of model fit ($\chi^2_4 = 8.07$, n.s.). However, the estimates for males and females were significantly different ($\chi^2_4 = 24.76$, $p < 0.001$). The regression coefficient for the interaction term between sex and year of birth (YOB) could not be dropped from the model without a significant loss of fit ($\beta_{YOB \times Sex} = 0.31$, $\chi^2_1 = 717.88$, $p < 0.001$; $\beta_{YOB} = -0.25$), suggesting that YOB influences smoking age-at-onset differently in males and females.

For the best-fitting threshold model, correlations for male and female twins could be equated both within MZ twin pairs and within

Table 3. *Univariate variance components estimates (with 95% confidence intervals) for the common-effects sex-limitation model*

A _f	C _f	T _f	E _f	A _m	C _m	T _m	E _m
0.59 (0.46–0.71)	0.04 (0.002–0.13)	0.20 (0.10–0.29)	0.17 (0.15–0.20)	0.63 (0.47–0.77)	0.07 (0.01–0.18)	0.11 (0.02–0.23)	0.19 (0.15–0.24)

A, Additive genetic; C, common environmental; T, twin-specific environmental; E, unique environmental; f, female; m, male.

same-sex DZ pairs ($\chi^2_2=0.46$, n.s.). However, equating the DZ opposite-sex correlation to that for the same-sex DZ pairs significantly reduced the fit of the model ($\chi^2_1=13.59$, $p<0.001$), suggesting that the magnitude of genetic and environmental influences on smoking age-at-onset may differ between males and females.

Consequently, further assumption testing proceeded without the DZ twin correlations equated, so that sex-specific correlation comparisons could be made. Twin–sibling and sibling–sibling correlations could be equated for the opposite-sex pairs, as well as both the male and female same-sex pairs, without a significant loss of fit ($\chi^2_1=0.52$, n.s.; $\chi^2_1=2.07$, n.s.; $\chi^2_1=0.02$, n.s.; respectively). These correlations could be equated to the respective DZ twin correlations without a significant loss of fit for the opposite-sex groups, but with a worsening of model fit for either the male or female same-sex groups ($\chi^2_1=4.18$, $p=0.04$; $\chi^2_1=8.18$, $p<0.01$, $\chi^2_1=15.41$, $p<0.001$; respectively). This was caused by the higher twin–twin correlations.

Univariate model fitting

Based upon the assumption testing results presented above, a common-effects sex-limitation model including a twin-specific environmental component was fitted for smoking age-at-onset. Male and female variance component estimates from the full common-effects sex-limitation model could not be equated without a significant loss of model fit ($\chi^2_4=16.10$, $p<0.01$). This was predominantly due to the difference in the magnitude of the twin-specific environmental component. No latent factors could be removed from the model without a significant loss of fit (all tests $p<0.01$). Thus the full common-effects sex-limitation model was the best-fitting model. Variance components estimates, with 95% CIs, for this model are shown in Table 3.

Multivariate model fitting

Based on the results from the various stages of assumption testing, a common-effects sex-limitation three-stage model including A, C, E and T parameters was fitted to the data for smoking age-at-onset, cigarette consumption and persistent smoking. The specification of thresholds for this model was complex, given the results of the preliminary analyses. For smoking age-at-onset, separate thresholds were specified for males and females, with a unique threshold model specified for the male opposite-sex twins. For cigarette consumption, this specification of thresholds was retained as there does not appear to be a significant difference between twins and siblings for average daily cigarette consumption. However, as the prevalence of current and ex-smokers differed significantly between twins and siblings in this sample, for the persistence phenotype thresholds were specified separately for male and female twins and siblings, with an additional, separate threshold again specified for the males from the opposite-sex twin pairs.

As expected, the multivariate model retrieved parameter estimates for smoking age-at-onset that were very close to the univariate model results. Results for the full model with relevant estimates for females and males are shown in Figs 1 and 2 respectively. Constraining all the paths to be equal for males and females significantly worsened the fit of the model ($\chi^2_{24}=45.84$, $p<0.01$). Equating the common environmental path coefficients for males and females significantly worsened the fit of the model ($\chi^2_6=29.65$, $p<0.001$). However, individually equating the unique environmental, twin-specific environmental or additive genetic paths did not significantly alter the model fit ($\chi^2_6=8.07$, n.s.; $\chi^2_6=3.83$, n.s.; $\chi^2_6=9.48$, n.s.; respectively).

Under the saturated model, all three phenotypes were moderately heritable, with estimates ranging from 0.40 to 0.62 (see Figs 1 and 2).

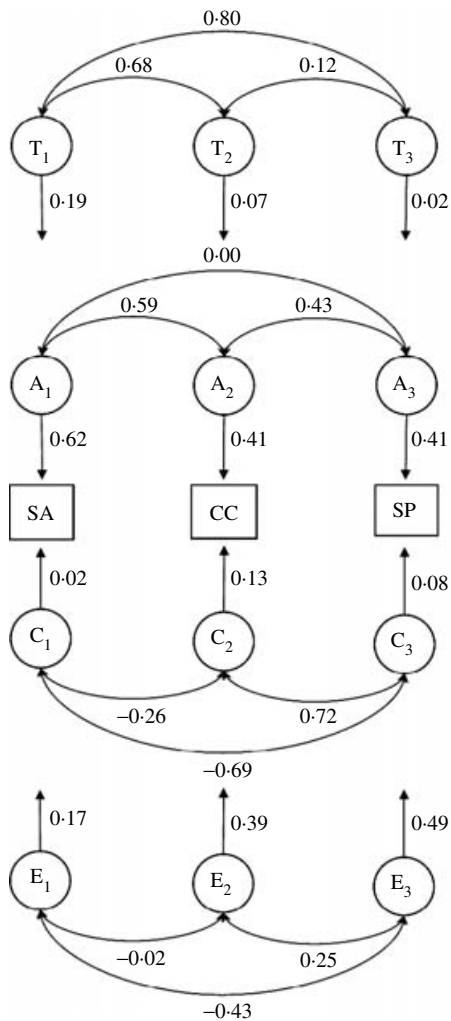


FIG. 1. Three-stage model results for smoking age-at-onset (SA), cigarette consumption (CC) and smoking persistence (SP) in females. Point estimates of additive genetic (A), twin-specific environmental (T), common environmental (C), and unique environmental (E) variances and correlations are shown.

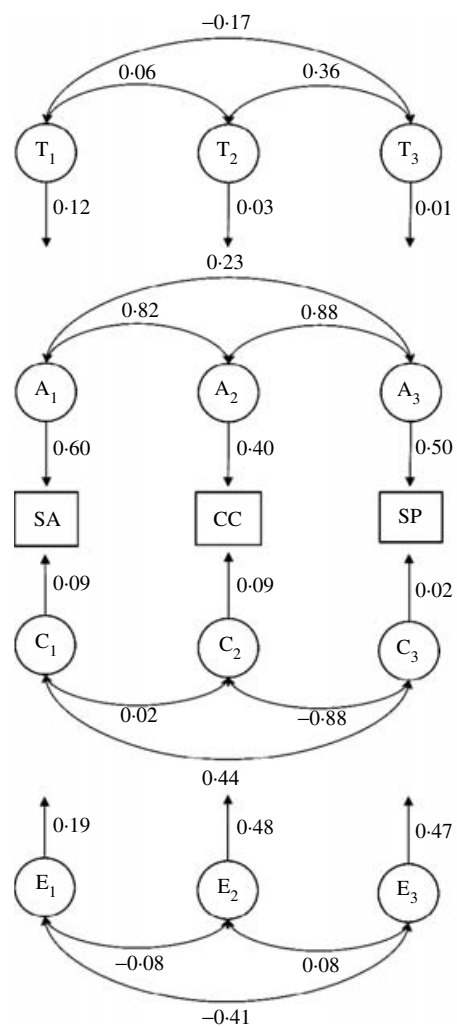


FIG. 2. Three-stage model results for smoking age-at-onset (SA), cigarette consumption (CC) and smoking persistence (SP) in males. Point estimates of additive genetic (A), twin-specific environmental (T), common environmental (C), and unique environmental (E) variances and correlations are shown.

Twin-specific environmental factors were estimated to have a reasonable influence on variability in smoking age-at-onset (estimates of 0.19 in females and 0.12 in males) but little influence on cigarette consumption and smoking persistence (less than 0.10 in both sexes). Common environmental estimates were small for all variables in both sexes, with the exception of cigarette consumption in females. Unique environmental estimates were moderate (0.17 and

0.19) for both sexes for smoking age-at-onset, but accounted for a substantial proportion of the variability in cigarette consumption and smoking persistence in both sexes, ranging from 0.39 to 0.49.

The pattern of correlations between the latent factors was somewhat different for males and females. In both sexes, there was a strong positive genetic correlation between age-at-onset and cigarette consumption, and another strong

to moderate positive genetic correlation between cigarette consumption and smoking persistence. In males the genetic correlation was small and positive between smoking age-at-onset and smoking persistence, but in females the correlation was effectively zero. Twin-specific environmental correlations were also different between the sexes; females showed strong positive correlations between age-at-onset and the other two traits, but little correlation between consumption and persistence. In males, these correlations were low to moderate, and negative in the case of smoking age-at-onset and smoking persistence. In both sexes strong negative unique environmental correlations were observed between smoking initiation and smoking persistence.

DISCUSSION

To explore the relationship between smoking age-at-onset, cigarette consumption and smoking persistence, we fitted a three-stage model allowing for sex differences in the magnitude of genetic and environmental effects. The variance estimates and correlations obtained from this multivariate model provide insight into observed sex differences in smoking behaviour, and have implications for both phenotype definition and further multivariate modelling of smoking.

All three phenotypes were estimated to be moderately heritable in both sexes, with additive genetic influences accounting for 40–60% of the variance. For females, there were moderate genetic correlations between initiation and consumption, and between consumption and persistence. In males there were moderate to high genetic correlations for all three variables. This suggests that although additive genetic factors influence variability in the traits to a similar extent in males and females, these factors are shared across traits to a much greater extent in males. As has been found in other studies (Heath *et al.* 1993, 2002; True *et al.* 1997; Hetttema *et al.* 1999; Maes *et al.* 2004), common environmental factors did not have a major influence on any of these phenotypes in either sex.

The influence of environmental factors shared by members of a twin pair on smoking age-at-onset was substantial for both sexes, although

there appeared to be a stronger influence in females. This has previously been identified for tobacco and other substance use (e.g. Rhee *et al.* 2003), and suggests that sibling interaction plays an important role in smoking initiation. This is also supported by the finding that having a sibling of the same age is a strong influence on smoking uptake (Vink *et al.* 2003). In women, there were also strong twin-specific environmental correlations between age-at-onset and cigarette consumption, and between age-at-onset and smoking persistence. This suggests that although the influence of these factors on subsequent smoking behaviours is comparatively small, many of them are shared with smoking age-at-onset. These results imply that for men the inter-relationship between different aspects of smoking behaviour is most strongly mediated by additive genetic factors. While additive genetic factors are still an important aspect of the relationship between these behaviours in women, environmental factors may make a greater contribution than is seen in men.

The unique environmental correlations from this model raise some questions about appropriate definition of smoking phenotypes. For both males and females a strong, negative unique environmental correlation was observed between smoking age-at-onset and smoking persistence. Although this result has been observed before (Heath *et al.* 2002), it seems counter-intuitive because the categorization of smoking age-at-onset was based upon the assumption that individuals who started smoking later in life would be more likely to quit smoking. Thus, these results could suggest that this parameterization of smoking initiation was incorrect.

However, it is also possible that these results were caused by censoring of the data. The mean age at survey for this sample was 34 years (median = 30 years, mode = 26 years), so individuals starting to smoke after age 18 would, on average, have had fewer opportunities to quit smoking than those who began smoking by age 18. Thus although late-onset smokers may be more likely to quit in the long term, we may have observed them at the beginning or middle of their smoking career, making them more likely to be current smokers. In comparison, the early-onset smokers will include individuals who

only experimented with smoking in their teens or early 20s, potentially resulting in a larger proportion of ex-smokers in comparison to the late-onset smokers.

The negative unique environmental correlations that we observed highlight the need for a better phenotypic definition of smoking persistence. While all measures of smoking persistence (and also smoking initiation) are subject to censoring, the binary measure of persistence used here, and in other behavioural genetic studies of smoking (e.g. Heath *et al.* 1993; True *et al.* 1997; Madden *et al.* 1999, 2004), is particularly susceptible. We suggest that future studies of smoking behaviour work towards developing more informative measures of persistence.

These results also demonstrate how twin analysis can be used to explore relationships between aspects of smoking behaviour initially observed only at a phenotypic level in epidemiological studies. The three traits analysed in the multivariate model only show a low to moderate degree of phenotypic correlation. Correlations between smoking age-at-onset and cigarette consumption, and between cigarette consumption and smoking persistence, were only moderate in either sex, at around 0.35. There was no substantial phenotypic correlation between smoking age-at-onset and persistence in males or females. However, multivariate analysis revealed strong twin-specific environmental correlations between the three variables in women. Additionally, in men, and to a lesser extent in women, there was a strong additive genetic correlation between age-at-onset and consumption, and another between consumption and persistence. These results indicate much stronger inter-relationships between certain sources of risk factors for these different aspects of smoking behaviour than might be initially expected from the phenotypic correlations. This type of information may have ramifications not only for study design but also for the development of more effective tobacco control strategies.

These results should be interpreted within the potential limitations of the modelling strategy used. They are limited by the fact that this research was based on retrospective self-reports, which can be unreliable, particularly for substance use (Johnson & Mott, 2001; Johnson &

Schultz, 2005). Additionally, if concordance for initiation is high, multivariate analyses may have insufficient power to differentiate genetic and common environmental sources of covariation between initiation and subsequent use phenotypes (Pergadia *et al.* 2006). However, given our large sample size, and the inclusion of siblings, this is unlikely to be a substantial problem for the analyses presented here. Finally, the first dimension variable used in the multivariate model did not perfectly meet the SLD assumption. However, as this variable produced an estimate for the causal path of the CCC model that was very close to unity, it is unlikely that this would have had a substantial impact on the results.

Overall, this research suggests that the inter-relationship between smoking age-at-onset, cigarette consumption and smoking persistence observed in epidemiological studies is likely to be mediated by both genetic and unique environmental factors that are shared between these traits. Despite the limitations outlined above, it is likely that similar analyses using more specialized measures of these phenotypes would produce similar results; namely, that the inter-relationship between smoking age-at-onset, cigarette consumption and smoking persistence is strongly influenced by genetic factors. Thus these results suggest that genes with pleiotropic effects on smoking behaviour are likely to exist, and that multivariate studies aimed at identifying these genes may prove informative.

ACKNOWLEDGEMENTS

We thank the twins and their families for their participation in the various studies. We also thank Drs Sarah Medland, Julie Grant, Valerie Knopik and Michele Pergadia for helpful discussions on multivariate model fitting, and Professor Wayne Hall and two anonymous reviewers for their useful comments. We acknowledge the funding sources that supported this project, and earlier studies that collected the data used: NIH grants (DA00272, DA12854, DA12540, CA75581, AA07535, and AA07728), Australian NH&MRC grants (971232 and 941177). K.I.M. is supported by an Ian Scott Fellowship from the Australian Rotary Health Research Fund.

DECLARATION OF INTEREST

None.

REFERENCES

- Breslau, N. & Peterson, E. L. (1996). Smoking cessation in young adults: age at initiation of cigarette smoking and other suspected influences. *American Journal of Public Health* **86**, 214–220.
- Everett, S. A., Warren, C. W., Sharp, D., Kann, L., Husten, C. G. & Crossett, L. S. (1999). Initiation of cigarette smoking and subsequent smoking behavior among U.S. high school students. *Preventive Medicine* **29**, 327–333.
- Ezzati, M. & Lopez, A. D. (2003). Estimates of global mortality attributable to smoking in 2000. *Lancet* **362**, 847–852.
- Ezzati, M. & Lopez, A. D. (2004). Regional, disease specific patterns of smoking-attributable mortality in 2000. *Tobacco Control* **13**, 388–395.
- Grayson, D. A. (1989). Twins reared together: minimizing shared environmental effects. *Behavior Genetics* **19**, 593–604.
- Heath, A. C. (1990). Persist or quit? Testing for a genetic contribution to smoking persistence. *Acta Geneticae Medicae et Gemellologiae* **39**, 447–458.
- Heath, A. C., Cates, R., Martin, N. G., Meyer, J., Hewitt, J. K., Neale, M. C. & Eaves, L. J. (1993). Genetic contribution to risk of smoking initiation: comparisons across birth cohorts and across cultures. *Journal of Substance Abuse* **5**, 221–246.
- Heath, A. C., Madden, P. A. F., Slutske, W. S. & Martin, N. G. (1995). Personality and the inheritance of smoking behavior: a genetic perspective. *Behavior Genetics* **25**, 103–117.
- Heath, A. C. & Martin, N. G. (1993). Genetic models for the natural history of smoking: evidence for a genetic influence on smoking persistence. *Addictive Behaviors* **18**, 19–34.
- Heath, A. C. & Martin, N. G. (1994). Genetic influences on alcohol consumption patterns and problem drinking: results from the Australian NH&MRC twin panel follow-up survey. *Annals of the New York Academy of Sciences* **708**, 72–85.
- Heath, A. C., Martin, N. G., Lynskey, M. T., Todorov, A. A. & Madden, P. A. (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.
- Hettema, J. M., Corey, L. A. & Kendler, K. S. (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.
- Hewitt, J. K. (1989). Of biases and more in the study of twins reared together: a reply to Grayson. *Behavior Genetics* **19**, 593–604.
- Hymowitz, N., Cummings, K. M., Hyland, A., Lynn, W. R., Pechacek, T. F. & Hartwell, T. D. (1997). Predictors of smoking cessation in a cohort of adult smokers followed for five years. *Tobacco Control* **6** (Suppl. 2), S57–S62.
- Johnson, E. O. & Schultz, L. (2005). Forward telescoping bias in reported age of onset: an example from cigarette smoking. *International Journal of Methods in Psychiatric Research* **14**, 113–129.
- Johnson, T. P. & Mott, J. A. (2001). The reliability of self-reported age of onset of tobacco, alcohol and illicit drug use. *Addiction* **96**, 1187–1198.
- Kendler, K. S., Neale, M. C., Sullivan, P., Corey, L. A., Gardner, C. O. & Prescott, C. A. (1999). A population-based twin study in women of smoking initiation and nicotine dependence. *Psychological Medicine* **29**, 299–308.
- Khuder, S. A., Dayal, H. H. & Mutgi, A. B. (1999). Age at smoking onset and its effect on smoking cessation. *Addictive Behaviors* **24**, 673–677.
- Kirk, K. M., Birley, A. J., Statham, D. J., Haddon, B., Lake, R. I. E., Andrews, J. G. & Martin, N. G. (2000). Anxiety and depression in twin and sib pairs extremely discordant and concordant for neuroticism: prodromus to a linkage study. *Twin Research* **3**, 299–309.
- Koepfen-Schomerus, G., Spinath, F. M. & Plomin, R. (2003). Twins and non-twin siblings: different estimates of shared environmental influence in early childhood. *Twin Research* **6**, 97–105.
- Koopmans, J. R., Slutske, W. S., Heath, A. C., Neale, M. C. & Boomsma, D. I. (1999). The genetics of smoking initiation and quantity smoked in Dutch adolescent and young adult twins. *Behavior Genetics* **29**, 383–393.
- Lake, R. I., Eaves, L. J., Maes, H. H., Heath, A. C. & Martin, N. G. (2000). Further evidence against the environmental transmission of individual differences in neuroticism from a collaborative study of 45,850 twins and relatives on two continents. *Behavior Genetics* **30**, 223–233.
- Lando, H. A., Thai, D. T., Murray, D. M., Robinson, L. A., Jeffery, R. W., Sherwood, N. E. & Henrikus, D. J. (1999). Age of initiation, smoking patterns, and risk in a population of working adults. *Preventive Medicine* **29**, 590–598.
- Lewinsohn, P. M., Rohde, P. & Brown, R. A. (1999). Level of current and past adolescent cigarette smoking as predictors of future substance use disorders in young adulthood. *Addiction* **94**, 913–921.
- Li, M. D. (2003). The genetics of smoking related behavior: a brief review. *American Journal of the Medical Sciences* **326**, 168–173.
- Little, R. J. A. & Rubin, D. (1987). *Statistical Analysis with Missing Data*. Wiley: New York.
- Madden, P. A., Heath, A. C., Pedersen, N. L., Kaprio, J., Koskenvuo, M. J. & Martin, N. G. (1999). The genetics of smoking persistence in men and women: a multicultural study. *Behavior Genetics* **29**, 423–431.
- Madden, P. A., Pedersen, N. L., Kaprio, J., Koskenvuo, M. J. & Martin, N. G. (2004). The epidemiology and genetics of smoking initiation and persistence: crosscultural comparisons of twin study results. *Twin Research* **7**, 82–97.
- Maes, H. H., Sullivan, P. F., Bulik, C. M., Neale, M. C., Prescott, C. A., Eaves, L. J. & Kendler, K. S. (2004). A twin study of genetic and environmental influences on tobacco initiation, regular tobacco use and nicotine dependence. *Psychological Medicine* **34**, 1–11.
- Mathers, C. D., Vos, E. T., Stevenson, C. E. & Begg, S. J. (2001). The burden of disease and injury in Australia. *Bulletin of the World Health Organization* **79**, 1076–1084.
- Medland, S. E. (2004). Alternate parameterization for scalar and non-scalar sex-limitation models in Mx. *Twin Research* **7**, 299–305.
- Neale, M. C., Aggen, S. H., Maes, H. H., Kubarych, T. S. & Schmitt, J. E. (2006a). Methodological issues in the assessment of substance use phenotypes. *Addictive Behaviors* **31**, 1010–1034.
- Neale, M. C., Boker, S. M., Xie, G. & Maes, H. H. (2003). *Mx: Statistical Modeling* (6th edn). Department of Psychiatry, VCU Box 900126: Richmond, VA 23298.
- Neale, M. C. & Cardon, L. R. (1992). *Methodology for Genetic Studies of Twins and Families*. Kluwer Academic: Dordrecht.
- Neale, M. C., Harvey, E., Maes, H. H., Sullivan, P. F. & Kendler, K. S. (2006b). Extensions to the modeling of initiation and progression: applications to substance use and abuse. *Behavior Genetics* **36**, 507–524.
- Neale, M. C., Roysamb, E. & Jacobson, K. (2006c). Multivariate genetic analysis of sex limitation and G×E interaction. *Twin Research and Human Genetics* **9**, 481–489.
- Pergadia, M. L., Heath, A. C., Martin, N. G. & Madden, P. A. (2006). Genetic analyses of DSM-IV nicotine withdrawal in adult twins. *Psychological Medicine* **36**, 963–972.
- Posthuma, D., Beem, A. L., de Geus, E. J. C., van Baal, G. C. M., van Hjelmborg, J. B., Iachine, I. & Boomsma, D. I. (2003). Theory and practice in quantitative genetics. *Twin Research* **6**, 361–376.
- Rhee, S. H., Hewitt, J. K., Young, S. E., Corley, R. P., Crowley, T. J. & Stallings, M. C. (2003). Genetic and environmental influences on substance initiation, use, and problem use in adolescents. *Archives of General Psychiatry* **60**, 1256–1264.
- StataCorp (2005). *Stata Statistical Software: Release 8.2*. Stata Corporation: College Station, TX.

- Sullivan, P. F. & Kendler, K. S.** (1999). The genetic epidemiology of smoking. *Nicotine and Tobacco Research* **1** (Suppl. 2), S51–S57; discussion S69–S70.
- True, W. R., Heath, A. C., Scherrer, J. F., Waterman, B., Goldberg, J., Lin, N., Eisen, S. A., Lyons, M. J. & Tsuang, M. T.** (1997). Genetic and environmental contributions to smoking. *Addiction* **92**, 1277–1287.
- Vink, J. M., Willemsen, G. & Boomsma, D. I.** (2003). The association of current smoking behavior with the smoking behavior of parents, siblings, friends and spouses. *Addiction* **98**, 923–931.
- White, V., Hill, D., Siahpush, M. & Bobevski, I.** (2003). How has the prevalence of cigarette smoking changed among Australian adults? Trends in smoking prevalence between 1980 and 2001. *Tobacco Control* **12**, ii67–ii74.