Addictive Behaviors, Vol. 18, pp. 19–34, 1993
Printed in the USA.

0306-4603/93 $6.00 + .00

GENETIC MODELS FOR THE NATURAL HISTORY OF SMOKING: EVIDENCE FOR A GENETIC INFLUENCE ON SMOKING PERSISTENCE

A. C. HEATH[1] and N. G. MARTIN[2]

[1]Department of Psychiatry, Washington University School of Medicine
[2]Queensland Institute of Medical Research, Australia

Abstract — We reanalyze data from the 1981 mailed questionnaire survey of the Australian twin register, to test for a genetic effect on smoking persistence (whether or not a smoker quits smoking). In the young cohort, aged 18–30 years, there are too few ex-smokers to permit resolution of genetic and non-genetic models. In the older cohort, we find a significant and substantial genetic effect on smoking persistence, accounting for 53% of the variance. This genetic effect on smoking persistence is independent of genetic effects on smoking initiation.

Much is known about the contribution of genetic factors to the natural history (Vaillant, 1983) of alcohol use and abuse. Selective breeding experiments in rodents have demonstrated an influence on "ethanol consumption, initial central nervous system sensitivity, ability to acquire tolerance, and expression of withdrawal symptoms" (McClearn & Erwin, 1982; see also Li et al., 1988). Studies of adoptees (Bohman et al., 1981; Cadoret, Cain, & Grove, 1980; Cloninger, Bohman, & Sigvardsson, 1981; Goodwin et al., 1974), of half-siblings (Schuckit, Goodwin, & Winokur, 1972), and of twins (Kaij, 1960; Hrubec & Omenn, 1981; Kaprio et al., 1987; McGue et al., 1989) have, with rare exception (Gurling, Murray, & Clifford, 1981; Murray, Clifford, & Gurling, 1983), confirmed a significant influence on the development of alcoholism. Surveys of drinking practices in general community samples of twins have also demonstrated a strong genetic influence on alcohol consumption patterns (Cederlof, Friberg, & Lundman, 1977; Clifford, Fulker, Gurling, & Murray, 1981; Heath, Jardine, & Martin, 1989; Heath, Meyer, Eaves, & Martin, 1991; Heath, Meyer, Jardine, & Martin, 1991; Jardine & Martin, 1984; Kaprio, Koskenvuo, & Sarna, 1981; Kaprio et al., 1987; Partanen, Bruun, & Markkanen, 1966). Such findings have generated laboratory-based twin studies which have demonstrated the role of genetic factors in determining the change in psychomotor performance, and differences in subjective ratings of intoxication, in response to a standard challenge dose of alcohol (Martin et al., 1985; Neale & Martin, 1989; Martin & Boomsma, 1989). High-risk studies of the sons of alcoholics and controls have demonstrated a decreased reactivity to alcohol in the former group which, it is hypothesized, may mediate the genetic influence on alcoholism (Schuckit, 1990; Schuckit & Gold, 1988).

Rather little is known from human studies, in contrast, about the role of genetic factors in the natural history of the smoking habit (Hughes, 1986). Several large-sample twin surveys have obtained data about smoking (Cederloff et al., 1977; Eaves &

Data collection was supported by grants from the Australian National Health and Medical Research Council and the Australian Associated Brewers. Data analysis was supported by ADAMHA Grants DA05588, AA07535 and AA07728. We thank Dr. John Mathews, Dr. John Gibson, Dr. Rosemary Jardine, and Marilyn Olsen for assistance with data collection.

Reprint requests should be sent to A. C. Heath, Dept. of Psychiatry, Washington Univ. School of Medicine, Medical School Box 8134, 4940 Audubon Ave., St. Louis, MO 63110.

Eysenck, 1980; Hannah, Hopper, & Mathews, 1984; Kaprio et al., 1981; Meyer, Heath, Martin, & Eaves, 1990; Pedersen, 1981), in most cases in order to permit investigation of the effects of smoking on health. Few genetic analyses have attained the level of sophistication characteristic of analyses of genetic influences on alcohol use and abuse. These suggest a significant genetic influence on initiation of the smoking habit (Eaves & Eysenck, 1980; Hannah et al., 1984; Heath, Meyer, & Martin, 1993; Meyer et al., 1993), the age at which onset of smoking is reported to occur (Heath, Meyer, & Martin, 1993), and the average daily consumption reported by smokers (Eaves & Eysenck, 1980; Meyer et al., 1993). Results in these studies, however, have not always been consistent across sexes (Eaves & Eysenck, 1980; Heath, 1990) or across age-cohorts (Heath et al., 1990).

In this paper, we address the question of whether there is any genetic influence on persistence of the smoking habit in those who are smokers. Two research groups have conducted careful genetic analyses and reported a significant genetic influence (Eaves & Eysenck, 1980; Hannah et al., 1984). Unfortunately, the two genetic analyses were based on radically different assumptions. In an analysis of data from a relatively small sample of London twin pairs (n = 547 pairs), Eaves and Eysenck used a parametric model which assumed that the same genetic and environmental factors determine smoking persistence as determine smoking initiation. In contrast, Hannah et al. analyzed data from a much larger sample of Australian twin pairs (Jardine & Martin, 1984; Jardine, 1985), but used a non-parametric approach which assumed complete independence of the genetic and environmental factors which determine persistence, and those which determine initiation. Incorrect assumptions about the determinants of initiation and persistence would lead to a serious bias in the estimate of genetic and environmental parameters. If persistence and initiation have independent determinants, then the approach of Eaves and Eysenck would confound two different modes of inheritance (Heath, Meyer, Eaves, & Martin, 1991). If persistence and initiation have common determinants, then the Hannah et al. analysis, which considered only pairs concordant for being smokers, would lead to truncation of the underlying liability distribution through exclusion of the non-smokers (Heath, Meyer, Eaves, & Martin, in press). Such truncation can lead to very serious biases in estimates of genetic effects from twin data (Martin & Wilson, 1982; Neale, Eaves, Kendler, & Hewitt, 1989).

In a reanalysis of the London data, we compared the results of fitting the Eaves and Eysenck model, a parametric version of the Hopper et al. model (Eaves, Eysenck, & Martin, 1989), and a more general model which included the two others as special cases (Heath, 1990). Our reanalyses confirmed the appropriateness of the original assumptions used by Eaves and Eysenck, since the parametric version of the Hopper model gave a very poor fit to the data. This was an unexpected finding. If the only genetic and environmental risk-factors which influenced smoking persistence were those which also determined initiation of the habit, then this appeared difficult to reconcile with the notion that smoking persistence is largely related to the development of nicotine dependence (U.S. DHHS, 1988). Since we know, from analyses of other data, that the genetic factors which determine average daily consumption are independent of those which determine smoking initiation (Meyer et al., 1993), this would imply no effect of level of consumption on smoking persistence! Sample sizes in the London study were small, however, and the sample was rather youthful (Eaves & Eysenck, 1980; Eaves, Eysenck, & Martin, 1989), so that there were relatively few ex-smokers in the sample. We therefore turn our attention in this paper to the much larger Australian sample, which had a wider age-range, and would therefore be expected to be much more informative about the casues of smoking persistence versus smoking cessation.

METHODS

Sample measures

The 1981 Australian twin survey has been described in detail elsewhere (Hannah et al., 1984; Heath & Martin, 1988; Jardine & Martin, 1984; Jardine, 1985). A self-report questionnaire was mailed to all 5967 twin pairs enrolled on the Australian National Health and Medical Research Council volunteer twin panel, aged 18 years or greater, between November, 1980 and March, 1982. Completed questionnaires were returned by both members of 3810 twin pairs, for a 64% *pairwise* response rate. These included 1233 monozygotic (MZ) female and 751 dizygotic (DZ) female like-sex pairs, 567 MZ male and 352 DZ male like-sex pairs, and 907 unlike-sex pairs. This excess of monozygotic twin pairs, and of female twin pairs, has commonly been found in studies of volunteer twin samples (Lykken, Tellegen, & DeRubeis, 1978; Martin & Wilson, 1982).

Included in the survey were questions about whether the respondent had ever smoked ("Have you EVER been a smoker?"); the ages at which the respondent stopped smoking cigarettes, pipe or cigars, if a former smoker ("If you have stopped smoking, how old were you when you stopped?"); and the respondent's present (or former) average daily consumption ("How many CIGARETTES do (or did) you usually smoke in a day?"). On the basis of their answers to these items, respondents were classified with respect to their use of cigarettes as either non-smokers, ex-smokers, or current smokers.

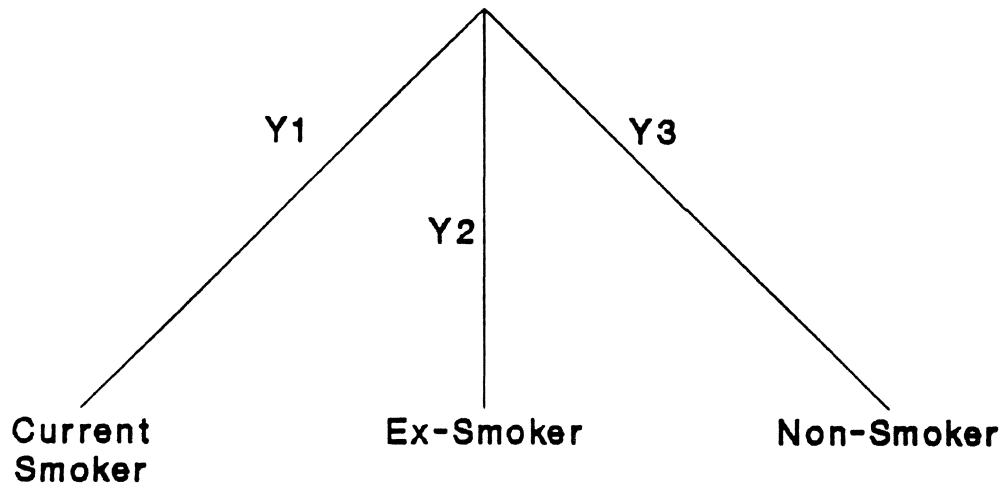
Data summary

The total sample was subdivided into those twin pairs aged 30 years or less at the time of response ("young cohort"), and those twin pairs aged greater than 30 years ("older cohort"). We considered that twins from the young cohort were more likely to have been exposed to information about the harmful effects of smoking (Royal College of Physicians, 1962; National Health and Medical Research Council, 1962) during adolescence, the age at which most smokers reported starting to smoke (Heath et al., 1993). Thus, we anticipated that there might be important differences in smoking behavior between the young and older cohorts. For each twin group, from each cohort, a two-way contingency table was computed cross-classifying the smoking status (non-smoker; ex-smoker; current smoker) of the first twin from each pair by the smoking status of the co-twin. Respondents from like-sex pairs were assigned as first or second twins on the basis of birth order where this information was available, but otherwise at random. Twins from unlike-sex pairs were reordered so that the first twin was always the female twin, the second twin her male co-twin.

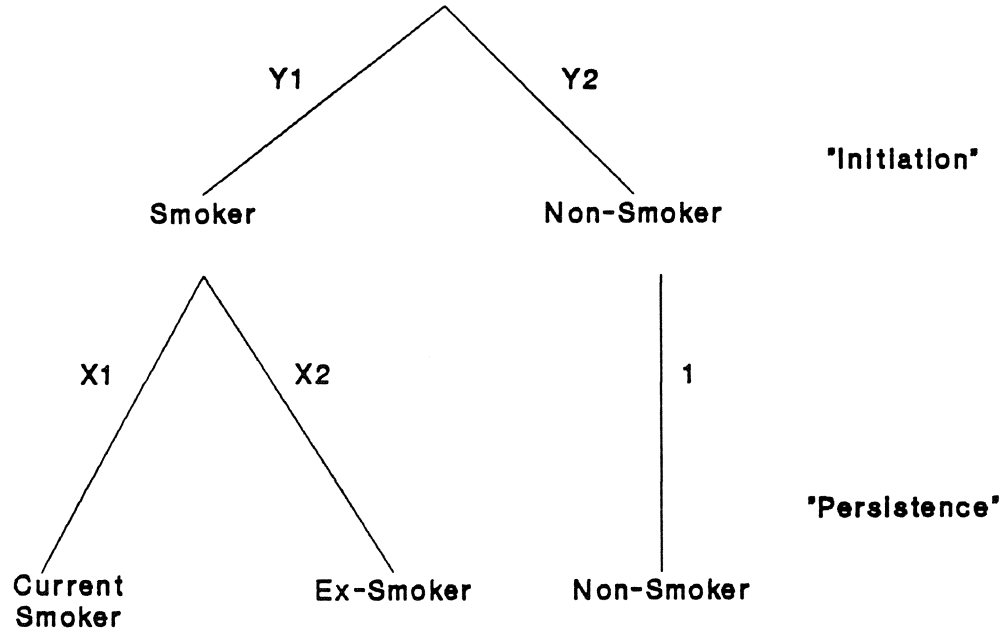
Models

We considered three alternative models for the relationship between the genetic and environmental determinants of smoking initiation, and of smoking persistence versus smoking cessation. These models are elaborated in Heath (1990), where they were applied to the London twin data of Eaves and Eysenck. The critical assumptions of these models are represented diagrammatically, in the form of probability trees, in Figure 1. Model 1(a), the Single Liability Dimension (SLD) model, assumes that the same genetic and environmental risk-factors which determine initiation of the smoking habit also determine smoking persistence. Persistent smokers are assumed to be more extreme in genetic or environmental risk than smokers who are able to quit smoking. Under this model, y_1, y_2 and y_3 denote the unconditional probabilities that a male respondent will be a non-smoker, ex-smoker or current smokers (with separate proba-

(a) Single liability dimension model



(b) Independent liability dimensions model



(*continued*)

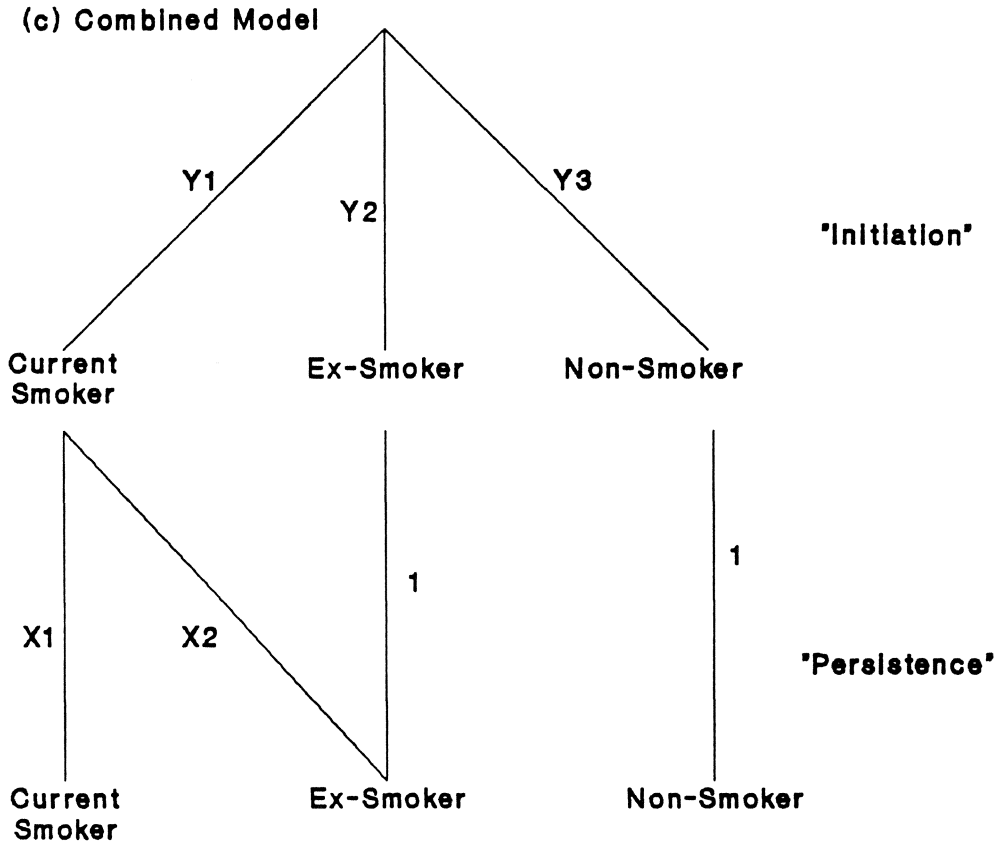


Fig. 1. Probability trees for single liability dimension (a), independent liability dimensions (b), and combined (c) models.

bilities y_1', y_2' and y_3' estimated for female respondents). This is the model which was assumed by Eaves and Eysenck (1980) in their analysis of smoking status in the London twin sample. An important prediction of this model is that the co-twins of persistent smokers are more likely to themselves be smokers than are the co-twins of ex-smokers.

Model 1(b), the Independent Liability Dimensions (ILD) model, postulates the existence of independent Initiation and Persistence dimensions, each determined by separate genetic and environmental risk-factors. The Initiation dimension determines the probability that a respondent will become a smoker (y_1 or y_1') or never start smoking (y_2 or y_2'). The Persistence dimension determines the conditional probability that an individual will persist in smoking (x_1 or x_1') or quit smoking (x_2 or x_2'), given that that individual has become a smoker. If the respondent is a non-smoker, the second dimension will clearly have no effect. This model was the one assumed by Hannah et al. (1984) in their nested non-parametric analysis of the smoking status data, pooled across cohorts, for this sample. A version of the same model was also considered by Eaves and Eysenck (1980) in a discussion of the determinants of age-of-onset of smoking, although not applied in their analysis of smoking status. This model predicts that the co-twins of persistent smokers are no more likely to be smokers than are the co-twins of ex-smokers.

The third model (1c), the combined model, derives from work on the scaling and

genetic analysis of alcohol consumption patterns (Heath, Meyer, Eaves, & Martin, 1991; Heath, Meyer, Jardine, & Martin, 1991). This is a more general model which includes both the two previous models as special cases. Like the ILD model, it postulates the existence of separate Initiation and Persistence dimensions. It recognizes the possibility that there are some genetic and environmental risk-factors which influence both smoking initiation and smoking persistence (as assumed by the SLD model) and that there are other genetic and environmental risk-factors which only influence smoking persistence, and which do not influence initiation (as assumed by the ILD model). Some individuals may become ex-smokers because of low levels of exposure to risk-factors which also promote smoking initiation while others may become ex-smokers because of low levels of exposure to risk-factors which influence smoking persistence, but not smoking initiation. This combined model predicts that smokers will be more common among the co-twins of persistent smokers than among the co-twins of ex-smokers, but not as common as would be expected under the ILD model.

Model fitting analyses

We follow Eaves and Eysenck (1980; Eaves, Last, Young, & Martin, 1978) and assume, in the case of the SLD model, a single underlying normal "liability" distribution, which determines smoking status; and, in the case of the ILD and combined models, we assume independent normal Initiation and Persistence distributions (Heath, 1990). The joint distribution of twin pairs for these liability distributions is assumed to be bivariate normal. Abrupt thresholds, whose values are to be determined by model-fitting, subdivide the continuous distributions into discontinuous categories, corresponding to the categories of Figure 1. Thus, in the case of the SLD model, we have four thresholds for each sex, with $t_0 = t'_0 = -\infty$, $t_3 = t'_3 = +\infty$, and the values of t_1, t_2 (in males) and t'_1, t'_2 (in females) are to be estimated. Male respondents whose liability value lies between t_0 and t_1 will be current smokers; between t_1 and t_2 will be ex-smokers; and between t_2 and t_3 will be non-smokers. Hence, the marginal probabilities y_1, y_2 and y_3 are obtained by integrating the standardized normal distribution between the appropriate threshold values. In the case of the ILD and combined models, we must also estimate a second set of threshold values, s_1 and s'_1, with $s_0 = s'_0 = -\infty$ and $s_2 = s'_2 = +\infty$, for the Persistence dimension: respondents who are smokers (ILD model) or current smokers (combined model) on the Initiation dimension and have liability values $< s_1$ will be current smokers; those with liability values $> s_1$ will be ex-smokers. Under the combined model, the number of thresholds to be estimated for the first, Initiation dimension will be the same as under the SLD model. Under the ILD model, there will be two fewer thresholds, since we will have $t_2 = t'_2 = +\infty$.

Under these assumptions we can compute, for the i-th zygosity group, for given threshold values and twin correlation(s), the probability P(i,j,k) of observing a twin pair in the j,k-th cell of the i-th two-way contingency table. Let Y(i,j,k) denote the unconditional probability of the twin pair falling in the j,k-th cell of the two-way contingency table for smoking status (SLD model) or smoking Initiation (ILD, combined models). Let X(i,j,k) denote the conditional probability of the twin pair falling on the j,k-th cell of the two-way contingency table for smoking Persistence (ILD, combined models only). For all models, we have:

$$Y(i,j,k) = \Phi(t_j,t_k) - \Phi(t_{j-1},t_k) - \Phi(t_j,t_{k-1}) + \Phi(t_{j-1},t_{k-1}) \quad (1)$$

where Φ is the bivariate normal distribution function with twin correlation r_i for smoking status (SLD model) or for Initiation (ILD, combined models). For the SLD model,

P(i,j,k) = Y(i,j,k) for all i,j,k. For the combined and ILD models, we can obtain X(i,j,k) in similar fashion, substituting s_i for t_i, r'_i (the twin correlation for Persistence) for r_i, and X(i,j,k) for Y(i,j,k) in function (1) above. Expressions for deriving P(i,j,k) from the Xs and Ys under the three models are summarized in Table 1.

The log-likelihood of a set of observations, under a given model with given threshold values and twin correlations, is given by

$$L = \ln(c) + \Sigma\Sigma\Sigma\, f(i,j,k) \ln P(i,j,k)$$

where c is a constant, and f(i,j,k) is the observed frequency of twin pairs from the i-th twin group in the j,k-th cell of the observed contingency table. Maximum-likelihood estimates of the model parameters are obtained by maximizing this function with respect to the parameter values. The sampling covariance matrix of the parameter estimates, from which standard errors may be obtained, can be derived as the inverse of the Fisher Information matrix, whose m,n-th element will be (Tallis, 1962; Olsson, 1979):

$$I_{m,n} = \sum_{i\ j\ k}\Sigma\Sigma \frac{N_i}{P(i,j,k)} \frac{dP(i,j,k)}{d\Theta_m} \frac{dP(i,j,k)}{d\Theta_n}$$

where Θ is the vector of parameter estimates. The fit of a given model can be assessed by the usual chi-square goodness-of-fit statistic,

$$C = \Sigma\Sigma\Sigma(f(i,j,k) - e(i,j,k))^2/e(i,j,k)$$

where e(i,j,k) is the expected frequency for the j,k-th cell of the i-th contingency table under the model. The statistic C is approximately distributed as chi-square, with number of degrees of freedom equal to $\Sigma(n_i^2 - 1) - m$, where n_i is the number of response categories of the i-th contingency table, and m is the number of model parameters estimated. The goodness-of-fit of a more general model, and a submodel which fixes some of the parameters of the general model, can also be compared by chi-square test. The difference between the chi-square values obtained under the two models is itself distributed as chi-square, with numbers of degrees of freedom equal to the number of free parameters in the more general model which have been fixed equal in the first model.

Table 1. Predicted probabilities P(j,k) under the single liability dimension (SLD), independent liability dimensions (ILD), and combined models

Twin 1/female twin	Model	Twin 2/Male twin		
		Current smoker	Ex-smoker	Non-smoker
Current smoker	SLD	Y_{11}	Y_{12}	Y_{13}
	ILD	$Y_{11}X_{11}$	$Y_{11}X_{12}$	$Y_{12}(X_{11} + X_{12})$
	Combined	$Y_{11}X_{11}$	$Y_{11}X_{12} + Y_{12}(X_{12} + X_{11})$	$Y_{13}(X_{11} + X_{12})$
Ex-smoker	SLD	Y_{21}	Y_{22}	Y_{23}
	ILD	$Y_{11}X_{21}$	$Y_{11}X_{22}$	$Y_{12}(X_{22} + X_{21})$
	Combined	$Y_{11}X_{21} + Y_{21}(X_{21} + X_{11})$	$Y_{22} + Y_{21}(X_{12} + X_{22}) + Y_{11}X_{21} + Y_{12}(X_{21} + X_{22})$	$Y_{23} + Y_{13}(X_{21} + X_{22})$
Non-smoker	SLD	Y_{31}	Y_{32}	Y_{33}
	ILD	$Y_{21}(X_{11} + X_{21})$	$Y_{21}(X_{22} + X_{12})$	Y_{22}
	Combined	$Y_{31}(X_{11} + X_{21})$	$Y_{32} + Y_{31}(X_{12} + X_{22})$	Y_{33}

Table 2. Expectations for correlations between twin pairs under a full genetic model

MZ female pairs	$VA_f + EC_f$
MZ male pairs	$VA_m + EC_m$
DZ female pairs	$\frac{1}{2}VA_f + EC_f$
DZ male pairs	$\frac{1}{2}VA_m + EC_m$
Unlike-sex pairs	$\frac{1}{2}(VA_mVA_f)^{1/2} + (EC_mEC_f)^{1/2}\, r_c$

VA_f, EC_f; VA_m, EC_m denote additive genetic and shared environmental components of variance in females, and in males, respectively; r_c denotes the correlation between shared environmental effects in male and female twins from unlike-sex pairs.

Comparison of genetic models

We can reparameterize the twin correlations of equation (1) in terms of genetic and environmental parameters, and thus test a variety of genetic and non-genetic models. For each liability dimension, we can decompose the total variance into components due to additive gene action (VA), shared environment (EC) and non-shared environment (ES). Non-additive genetic effects, such as dominance or epistasis, will tend to be masked by shared environmental effects in twin data (Martin, Eaves, Kearsey, & Davis, 1978). For each liability dimension, we compared the fit of a simple additive genetic model, under which twin resemblance is entirely due to the additive effects of genes; a simple shared environment model, under which twin resemblance is entirely due to shared environmental effects; a model which allowed for both genetic and shared environmental contributions to twin pair resemblance (VA EC model); and a full model which allowed for sex-differences in the magnitude of genetic and shared environmental variance components, and which allowed the correlation between shared environmental components in unlike-sex pairs to take values less than unity. In the case of the ILD and combined models, therefore, we fitted a total of sixteen models, testing every combination of models for the Initiation and Persistence dimensions. Since we were fitting models to correlations, implying a total variance of unity for each dimension, values of the non-shared environmental components were obtained by subtracting the sum of the additive genetic and shared environmental components from unity. Expectations for twin correlations in terms of genetic and environmental parameters are summarized in Table 2.

RESULTS

Table 3 gives the two-way contingency tables for smoking status, computed separately for each twin group from each birth cohort. Also shown are the proportions of non-smokers, ex-smokers and current sokers for first and second twins from each twin group. In the young cohort, there is no consistent sex difference in the proportions of non-smokers, ex-smokers and current smokers. The percentage of respondents who are ex-smokers is somewhat small, implying that there will be relatively little power for detecting genetic effects on smoking persistence in this cohort. In the older cohort, the percentages of female respondents who are non-smokers, or ex-smokers, are not much higher than was observed for the young cohort. The proportion of older males who are current smokers is slightly lower than in the young cohort, but the proportion who are ex-smokers is much higher.

Table 3. Twin concordance for smoking status, as a function of zygosity and age cohorts. Proportions of first or second twins who are non-smokers, ex-smokers, or current smokers are given in parentheses

	Twin 2/Male Twin[1]							
	Young cohort				Older cohort			
Twin 1/female twin[1]	Non-smoker	Ex-smoker	Current smoker		Non-smoker	Ex-smoker	Current smoker	
MZ female pairs		(n = 570)				(n = 663)		
Non-smoker	282	24	21	(57%)	348	26	35	(62%)
Ex-smoker	23	19	33	(13%)	37	45	29	(16%)
Current smoker	28	23	117	(30%)	40	30	73	(22%)
	(58%)	(12%)	(30%)		(64%)	(15%)	(21%)	
MZ male pairs		(n = 274)				(n = 293)		
Non-smoker	128	11	15	(56%)	93	29	16	(47%)
Ex-smoker	13	16	13	(15%)	24	54	17	(33%)
Current smoker	18	13	47	(29%)	12	18	30	(20%)
	(58%)	(15%)	(27%)		(44%)	(35%)	(22%)	
DZ female pairs		(n = 351)				(n = 400)		
Non-smoker	140	19	39	(56%)	170	27	47	(61%)
Ex-smoker	17	10	18	(13%)	37	23	17	(19%)
Current smoker	31	9	68	(31%)	31	17	31	(20%)
	(53%)	(11%)	(36%)		(60%)	(16%)	(24%)	
DZ male pairs		(n = 206)				(n = 146)		
Non-smoker	85	11	22	(57%)	36	11	12	(40%)
Ex-smoker	12	8	13	(16%)	11	20	14	(31%)
Current smoker	19	10	26	(27%)	8	16	18	(29%)
	(56%)	(14%)	(30%)		(38 %)	(32%)	(30%)	
DZ unlike-sex pairs		(n = 510)				(n = 397)		
Non-smoker	169	36	79	(56%)	98	84	66	(62%)
Ex-smoker	30	20	34	(18%)	22	21	19	(16%)
Current smoker	47	22	73	(28%)	14	28	45	(22%)
	(48%)	(16%)	(36%)		(34%)	(33%)	(33%)	

[1]Female twin, male twin in the case of unlike-sex pairs.

Table 4 compares the results of fitting single liability dimension (SLD), independent liability dimension (ILD), and combined models, estimating separate twin correlations for each twin group. In the young cohort, both the SLD and the ILD models are clearly rejected by the chi-square goodness-of-fit test. The combined model gives a good fit to the data, and a substantially better fit than either reduced model. In the older cohort, no model gives a good fit to the data. However, the SLD model gives a fit which is substantially worse than that of the combined model, whereas the ILD model does not give a significantly worse fit ($X_2^2 = 3.68$, $p = 0.16$). Thus the combined model gives the best

Table 4. Results of fitting single liability dimension (SLD), independent liability dimensions (ILD), and combined models, estimating a separate correlation for each twin group

	Young cohort			Older cohort		
Model	d.f.	χ^2	p	d.f.	χ^2	p
SLD	31	57.61	0.003	31	94.22	<0.001
ILD	26	52.48	0.002	26	41.51	0.03
Combined	24	28.75	0.23	24	37.83	0.04

fit to the data from the young cohort, but the independent liability dimensions model gives the best fit to the older cohort data.

Figure 2 summarizes the marginal probabilities (Xs and Ys) estimated for each cohort under the full combined model. In the older and young males, and the older females, a relatively small proportion of ex-smokers arises from the Initiation dimension (27%, 26% and 12%, respectively). In the young females, however, some 55% of ex-smokers appear to have quit because of their low liability on the Initiation dimension. It must be this group which is causing the combined model to give a significantly better fit than the independent liability dimensions model in the young cohort.

Table 5 gives maximum-likelihood estimates of polychoric correlations for the Initiation and Persistence dimensions, and the standard errors of these correlations, for the combined model (young cohort) and ILD model (older cohort). In the young

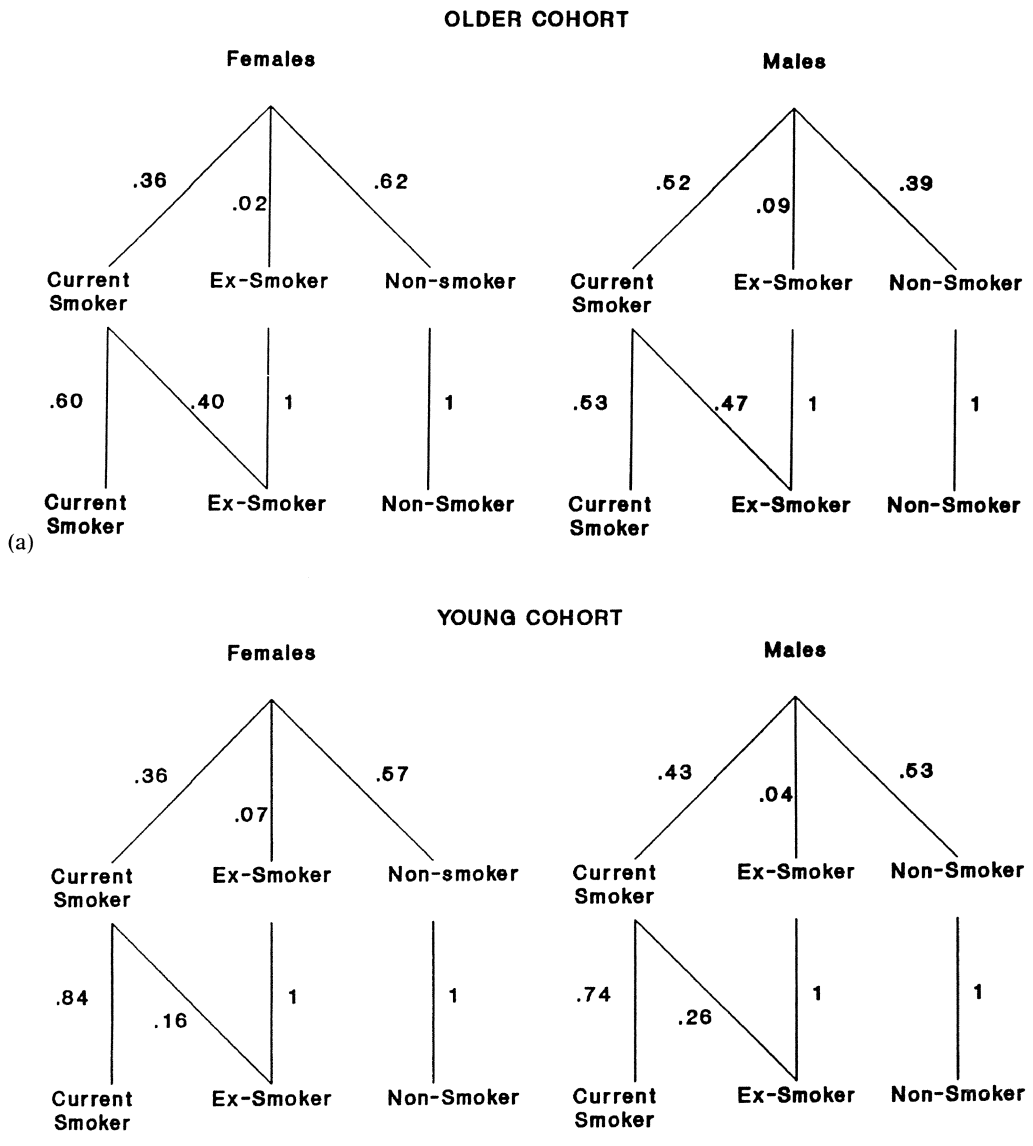


Fig. 2. Estimated probabilities under full combined model.

Table 5. Maximum likelihood estimates of polychoric correlations (r) and their standard errors (SE) under best fitting model

	Young cohort (Combined model)				Older cohort (ILD model)			
	Initiation		Persistence		Initiation		Persistence	
	r	SE	r	SE	r	SE	r	SE
MZ female pairs	0.86	0.03	0.48	0.20	0.77	0.03	0.48	0.10
MZ male pairs	0.79	0.05	0.60	0.17	0.64	0.06	0.58	0.10
DZ female pairs	0.57	0.06	0.76	0.22	0.40	0.07	0.34	0.16
DZ male pairs	0.55	0.08	0.21	0.27	0.59	0.10	0.19	0.19
DZ unlike-sex pairs	0.38	0.06	0.35	0.24	0.26	0.08	0.27	0.15

cohort, the monozygotic correlations for Initiation are significantly higher than the same-sex dizygotic correlations, implying an important genetic influence. For the Persistence dimension, whilst the MZ male correlation is substantially greater than the DZ male correlation, the MZ female correlation is actually lower than the DZ female correlation. However, the standard errors of the estimated correlations are extremely large, reflecting the relatively small proportion of ex-smokers in this age-group. The standard errors of the female like-sex correlations are especially large, despite the larger number of female than male like-sex twin pairs; this is a consequence of the relatively small number of young female ex-smokers who arise from the Persistence dimension.

In the older cohort, the twin correlations indicate a substantial genetic effect on Initiation in older females, but a negligible genetic effect on Initiation in older males. The older cohort polychoric correlations for the Persistence dimension have rather smaller standard errors than was the case for the young cohort. The monozygotic correlations for Persistence are higher, in both sexes, than the corresponding like-sex dizygotic correlations, implying a genetic influence.

Table 6 compares the results of fitting different genetic or environmental submodels for the Initiation and Persistence dimensions. Once again, for the young cohort we report the results obtained under the combined model, whereas for the older cohort we report results under the independent liability dimensions model.

For the young cohort, we find that there is no power for resolving genetic and non-genetic models for the Persistence dimension. For a given initiation genetic model, neither the simple additive genetic (VA) model nor the simple shared environment (EC) model for the Persistence dimension give significantly worse fits than the full model. For the Initiation dimension, a simple EC model can be rejected. A simple VA model gives an adequate fit to the data, and a VA EC model does not give a significant improvement in fit, by likelihood-ratio chi-square test. However, the full model does give a significant improvement in fit ($X_{42} = 10.54$, $p = 0.03$), indicating significant genotype x sex interaction for the Initiation dimension.

For the older cohort, all models which do not allow for genotype x sex interaction for the Initiation dimension are rejected by goodness-of-fit test. Among models which estimate a full genetic model for Initiation, the model which specifies a simple additive genetic model for the Persistence dimension gives an adequate fit to the data, and a fit which is not significantly worse than that of the VA EC ($X_1^2 = 0.5$, $p = 0.82$) or full models ($X_4^2 = 0.92$, $p = 0.33$). The simple EC model for the Persistence dimension, in contrast, is rejected by goodness-of-fit test, and gives a significantly worse fit than the VA EC model ($X_1^2 = 3.95$, $p < 0.05$). Thus for the older cohort we find that a model

Table 6. Comparison of goodness-of-fit of genetic and environmental models

Genetic model[1]		Chi-square goodness-of-fit tests					
		Young cohort (Combined model)			Older cohort (ILD model)		
Initiation	Persistence	d.f.	χ^2	p	d.f.	χ^2	p
Full model	Full model	24	28.75	0.23	26	41.51	0.03
Full model	VA	28	31.38	0.30	30	42.43	0.07
Full model	EC	28	31.88	0.28	30	46.33	0.03
Full model	VA EC	27	31.03	0.27	29	42.38	0.05
VA EC	Full model	27	36.91!	0.10	29	51.21	0.007
VA EC	VA	31	38.62	0.16	33	52.13	0.02
VA EC	EC	31	39.13	0.15	33	56.02	0.007
VA EC	VA EC	30	38.28	0.14	32	52.08	0.01
VA	Full model	28	39.29	0.09	30	51.23	0.009
VA	VA	32	41.68	0.12	34	52.15	0.02
VA	EC	32	41.96	0.11	34	56.06	0.01
VA	VA EC	31	41.23	0.10	33	52.10	0.02
EC	Full model	28	98.67!	< 0.001	30	92.38	< 0.001
EC	VA	32	98.55	< 0.001	34	93.29	< 0.001
EC	EC	32	100.37	< 0.001	34	97.21	< 0.001
EC	VA EC	31	98.49	< 0.001	33	93.24	< 0.001

! indicates a parameter estimate has gone to a boundary.
[1] VA = additive genetic model; EC = shared environmental model; VA EC = additive genetic plus shared environmental model for twin pair resemblance.

which specifies a full genetic model for the Initiation dimension, and the simple VA model for the Persistence dimension, gives the best fit to the data.

Table 7 summarizes the variance components estimated under the best-fitting model for each cohort. For the Initiation dimension, we report parameter estimates under a full genetic model. For the older cohort, we report parameter estimates when a simple VA model was fitted for Persistence. For the young cohort, we give estimates separately for the cases where the VA model, and the EC model, were fitted for Persistence. Esti-

Table 7. Estimates of genetic and environmental variance components under best-fitting model

	Young cohort (Combined model)		Older cohort (ILD model)
	VA model	EC model	
	Smoking initiation		
Additive genetic			
males	48.7%	49.4%	11.3%
females	55.7%	56.3%	74.1%
Shared environment			
males	30.7%	29.9%	53.2%
females	29.9%	29.4%	3.0%
Non-shared environment			
males	20.6%	20.7%	35.5%
females	14.4%	14.3%	22.9%
Unlike-sex environmental correlation	0.16	0.16	0.47
	Smoking persistence		
Additive genetic	58.4%	—	52.6%
Shared environment	—	48.8%	—
Non-shared environment	41.6%	51.2%	47.4%

mated variance components for the Initiation dimension change little according to whether the VA or EC model was fitted for the Persistence dimension. For the young cohort, we find that both shared environment and additive gene action are having an important influence on Initiation. The correlation between shared environmental effects in unlike-sex pairs is rather low (0.16), implying that many of these effects are sex-specific. Non-shared environmental influences are accounting for 42–51% of the variance in Persistence, but we cannot determine whether the remaining familial variance is attributable to shared environmental or additive genetic effects. For the older cohort, we find an important influence of shared environment, but only a slight influence of genotype, on Initiation in males; but a major genetic influence, and rather minor shared environmental influence, on Initiation in females. Genetic factors account for a substantial proportion of the variance on the smoking Persistence dimension.

DISCUSSION

For the older cohort of Australian twins, aged 31 years or greater at the time of the 1981 survey, we have found a significant and substantial genetic effect on smoking persistence, accounting for 53% of the variance in liability. This genetic effect on persistence appears to be unrelated to effects on smoking initiation, since the independent liability dimensions model gave the best fit to the data. These findings support the report of Hannah et al. (1984) of a higher MZ than DZ correlation for persistence, and provide a justification for their use of a nested model. As noted elsewhere, in an analysis of the determinants of age-of-onset of smoking (Heath, Meyer, & Martin, 1993), we have found little evidence for a genetic effect on smoking initiation in males from this cohort, despite finding substantial heritability of smoking initiation in females. This failure to find genetic effects on smoking initiation in older males, together with the relative absence of genetic effects on male alcohol consumption patterns in this older cohort (Jardine & Martin, 1984), must be considered an unexplained anomaly. Speculative explanations would be inappropriate until this cohort- or age-related difference is replicated in other studies.

For the young cohort of Australian twins, aged 18–30 years at the time of the 1981 survey, the combined model gave a significantly better fit than both single liability dimension and independent liability dimensions models. This implies that, in this cohort, there are some genetic or environmental risk-factors which influence both smoking initiation and smoking persistence; and others which influence only persistence. For the Initiation dimension, we found substantial genetic influence, accounting for between 49–56% of the variance in liability. Since, under the combined model, some respondents are expected to be ex-smokers because of their low liability on the Initiation dimension (55% of the young female ex-smokers, and 26% of the young male ex-smokers), we can thus say that we also have some evidence for genetic effects on smoking persistence in the young cohort. For the second Persistence dimension, however, there was too little information in the young cohort to be able to resolve the relative contributions of genetic and shared environmental factors.

Our findings for the Australian data-set, especially for the older cohort, are clearly inconsistent with the findings of the Eaves and Eysenck (1980; Heath, 1990) analyses of the London twin data. One possible explanation of the difference between these two studies is that the samples differ in age structure. Compared to the Australian sample, the English sample included younger respondents (ages 16+ years) and included a

higher proportion of adolescents or young adults (Eaves & Eysenck, 1980; Eaves, Eysenck, & Martin, 1989; Jardine, 1985). Results from the Eaves and Eysenck analyses more closely parallel those for the young Australian cohort, in finding some genetic and environmental risk-factors influencing both initiation and persistence of the smoking habit. When a combined model was fitted to the London data, we found that a sizeable proportion of respondents were ex-smokers because of their position on the Initiation liability dimension (42.6% of male ex-smokers, and 72.7% of female ex-smokers: Heath, 1990).

From the very limited data on the respondents' smoking histories available in the English and Australian twin studies, we cannot determine with certainty the explanation for this apparently age-related difference in findings. It is, however, likely that many of the "non-smokers" in each sample will have at least experimented with one or more cigarettes. Since we are relying on retrospective data, it is possible that some of the younger respondents may be reporting themselves as "ex-smokers" on the basis of a brief period of experimentation which never progressed to regular smoking; while those older respondents who never progressed beyond this stage, further removed in time from the period of experimentation, would be more likely to report themselves as non-smokers. If these assumptions are correct, then it is possible that genetic differences in acute sensitivity to nicotine (Marks, Burch, & Collins, 1983) account for some of the variance in the Initiation dimension. Individuals who are especially sensitive to the adverse effects of nicotine would be less likely to progress beyond the stage of experimentation, reporting themselves as non-smokers or, if young, as ex-smokers. Many of those individuals who are genetically insensitive to the acute effects of nicotine would progress to become regular smokers, and in these smokers new genetic factors, perhaps influencing the degree to which nicotine dependence develops, would determine the persistence of the smoking habit.

REFERENCES

Bohman, M., Sigvardsson, S., & Cloninger, C. R. (1981). Maternal inheritance of alcohol abuse: Cross-fostering analysis of adopted women. *Archives of General Psychiatry,* **38**, 965–969.

Cadoret, R. J., Cain, C. A., & Grove, W. M. (1980). Development of alcoholism in adoptees raised apart from alcoholic biologic relatives. *Archives of General Psychiatry,* **37**, 561–563.

Cederlof, R., Friberg, L., & Lundman, T. (1977). The interactions of smoking, environment and heredity and their implications for disease aetiology: a report of epidemiological studies on the Swedish twin registries. *Acta Medica Scandinavica Supplement,* **612**, 1–128.

Clifford, C. A., Fulker, D. W., Gurling, H. M. D., & Murray, R. M. (1981). Preliminary findings from a twin study of alcohol use. In L. Gedda, P. Parisi, W. E. Nance (Eds.), *Twin research 3: C. Epidemiological and clinical studies.* (Progress in Clinical and Biological Research, V. 69C) (pp. 47–52). New York: Alan R. Liss, Inc.

Cloninger, C. R., Bohman, M., & Sigvardsson, S. (1981). Inheritance of alcohol abuse: Cross-fostering analysis of adopted men. *Archives of General Psychiatry,* **38**, 861–868.

Eaves L. J., Last, K., Young, P. A., & Martin, N. G. (1978). Model-fitting approaches to the analysis of human behavior. *Heredity,* **41**, 249–320.

Eaves, L. J., & Eysenck, H. J. (1980). New approaches to the analysis of twin data and their application to smoking behavior. In H. J. Eysenck (Ed.), *The causes and effects of smoking* (pp. 140–314). London: Maurice Temple Smith.

Eaves, L. J., Eysenck, H. J., & Martin, N. G. (1989). *Genes, culture and personality: An empirical approach.* New York: Academic Press.

Goodwin, D. W., Schulsinger, F., Moller, N., Hermansen, L., Winokur, G., & Guze, S. B. (1974). Drinking problems in adopted and nonadopted sons of alcoholics. *Archives of General Psychiatry,* **31**, 164–169.

Gurling, H. M. D., Murray, R. M., & Clifford, C. A. (1981). Investigations into the genetics of alcohol dependence and into its effects on brain function. In L. Gedda, P. Parisi, & W. E. Nance (Eds.), *Twin research 3: Epidemiological and clinical studies* (Progress in Clinical and Biological Research, V. 69C) (pp. 77–87). New York: Alan R. Liss.

Hannah, M. C., Hopper, J. L., & Mathews, J. D. (1984). Twin concordance for a binary trait. II. Nested analysis of ever-smoking and ex-smoking traits and unnested analysis of a committed-smoking trait. *American Journal of Human Genetics,* **37**, 153–165.

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

Heath, A. C., & Martin, N. G. (1988). Teenage alcohol use in the Australian twin register: Genetic and social determinants of starting to drink. *Alcoholism: Clinical and Experimental Research,* **12**, 735–741.

Heath, A. C., Meyer, J., Eaves, L. J., & Martin, N. G. (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.

Heath, A. C., Meyer, J., Jardine, R., & Martin, N. G. (1991). The inheritance of alcohol consumption patterns in a general population twin sample: II. Determinants of consumption frequency and quantity consumed. *Journal of Studies on Alcohol.*

Heath, A. C., Meyer, J. M., & Martin, N. G. (1993). *Smoking in the Australian twin register: Genetic and social determinants of initiation of smoking.* Manuscript submitted for publication.

Hrubec, Z., & Omenn, G. S. (1981). Evidence of genetic predisposition to alcoholic cirrhosis and psychosis: Twin concordances for alcoholism and its biological end points by zygosity among male veterans. *Alcoholism: Clinical and Experimental Research,* **5**, 207–215.

Hughes, J. R. (1986). Genetics of smoking: A brief review. *Behavior Therapy,* **17**, 335–345.

Jardine, R., & Martin, N. G. (1984). Causes of variation in drinking habits in a large twin sample. *Acta Geneticae Medicae Gemellologeae,* **33**, 435–450.

Jardine, R. (1985). *A twin study of personality, social attitudes and drinking behavior.* Unpublished doctoral thesis, Australian National University, Canberra.

Kaij, L. (1960). *Alcoholism in twins: Studies on the etiology and sequels of abuse of alcohol.* Stockholm: Almqvist & Wiksell International.

Kaprio, J., Koskenvuo, M., & Sarna, S. (1981). Cigarette smoking, use of alcohol, and leisure-time physical activity among same-sexed adult twins. In L. Gedda, P. Parisi, & W. E. Nance (Eds.), *Twin research 3: Epidemiological and clinical studies* (Progress in clinical and biological research, V. 69C) (pp. 37–46). New York: Alan R. Liss, Inc.

Kaprio, J., Koskenvuo, M. D., Langinvainio, H., Romanov, K., Sarna, S., & Rose, R. J. (1987). Genetic influences on use and abuse of alcohol: A study of 5638 adult Finnish brothers. *Alcoholism: Clinical and Experimental Research,* **11**, 349–356.

Li, T. K., Lumeng, L., Doolittle, D. P., McBridge, W. J., Murphy, J. M., Froelich J. C., & Mozorati, S. (1988). Behavioral and neurochemical association of alcohol seeking behavior. In K. Kuriyama, A. Takada, & H. Ishii (Eds.), *Biomedical and social aspects of alcohol and alcoholism* (pp. 435–439). New York: Excerpta Medica.

Lykken, D. T., Tellegen, A., & DeRubeis, R. (1978). Volunteer bias in twin research: The rule of two-thirds. *Social Biology,* **25**, 1–9.

McClearn, G., & Erwin, V. (1982). Mechanisms of genetic influence on alcohol related behaviors. In NIAAA (Ed.), *Alcohol consumption and related problems—Alcohol and Health Monograph, 1,* 263–285 (DHHS Publication No. ADM 82-1190). Washington, DC: U. S. Government Printing Office.

McGue, M., Pickens, R. W., & Svikis, D. S. (1989). *Sex differences in the inheritance of alcoholism.* Paper presented at the 19th Annual Meeting of the Behavior Genetics Association, Charlottesville, VA.

Marks, M. J., Burch, J. B., & Collins, A. C. (1983). Genetics of nicotine response in four inbred strains of mice. *Journal of Pharmacology and Experimental Therapeutics,* **226**, 291–302.

Martin, N. G., Eaves, L. J., Kearsey, M. J., & Davies, P. (1978). The power of the classical twin study. *Heredity,* **40**, 97–116.

Martin, N. G., & Boomsma, D. I. (1989). Willingness to drive when drunk and personality: A twin study. *Behavior Genetics,* **19**, 97–111.

Martin, N. G., & Wilson, S. R. (1982). Bias in the estimation of heritability from truncated samples of twins. *Behavior Genetics,* **12**, 467–472.

Martin, N. G., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., Perl, J., & Wilks, A. V. (1985). A twin study of psychomotor and physiological responses to an acute dose of alcohol. *Behavior Genetics,* **15**, 305–347.

Meyer, J. M., Heath, A. C., Martin, N. G., & Eaves, L. J. (1993). *Genetic and environmental influences on the onset of the smoking habit and the quantity smoked.* Manuscript submitted for publication.

Murray, R. M., Clifford, C. A., & Gurling, H. M. D. (1983). Twin and adoption studies: How good is the evidence for a genetic role? In M. Galanter (Ed.), *Recent developments in alcoholism, vol. 1* (p. 25). New York: Plenum Press.

National Health and Medical Research Council. (1962). *Smoking and lung cancer.* Report of the 53rd Session of NH & MRC. Canberra, Australia, Commonwealth Government Printer.

Neale, M. C., Eaves, L. J., Kendler, K. S., & Hewitt, J. K. (1989). Bias in correlations from selected samples of relatives: The effects of soft selection. *Behavior Genetics,* **19**, 163–169.

Neale, M. C., & Martin, N. G. (1989). The effects of age, sex, and genotype on self-report drunkenness following a challenge dose of alcohol. *Behavior Genetics,* **19**, 63–78.

Olsson, U. (1979). Maximum-likelihood estimation of the polychoric correlation coefficient. *Psychometrika,* **44**, 443–460.

Partanen, J., Bruun, K., & Markkanen, T. (1966). *Inheritance of drinking behavior: A study on intelligence, personality, and use of alcohol of adult twins.* Helsinki: The Finnish Foundation for Alcohol Studies (Distributed by Rutgers Center of Alcohol Studies, New Brunswick, NJ).

Pedersen, N. (1981). Twin similarity for usage of common drugs. In L. Gedda, P. Parisi, & W. E. Nance (Eds.), *Twin research 3: Epidemiological and clinical studies* (Progress in clinical and biological research, V. 69C) (pp. 53–59). New York: Alan R. Liss, Inc.

Royal College of Physicians of London (1962). *Smoking and health.* New York: Pitman Publishing Corp.

Schuckit, M. A. (1990). A prospective study of children of alcoholics. In C. R. Cloninger & M. Begleiter (Eds.), *Banbury report 33: Genetics and biology of alcoholism* (pp. 183–194). New York: Cold Spring Harbor Laboratory Press.

Schuckit, M. A., & Gold, E. O. (1988). A simultaneous evaluation of multiple markers of ethanol/placebo challenges in sons of alcoholics and controls. *Archives of General Psychiatry,* **45**, 211–216.

Schuckit, M. A., Goodwin, D. W., & Winokur, G. (1972). A study of alcoholism in half siblings. *American Journal of Psychiatry,* **128**, 1132–1136.

Tallis, G. M. (1962). The maximum-likelihood estimation of correlation from contingency tables. *Biometrics,* **18**, 342–353.

U. S. Department of Health and Human Services (DHHS) (1988). *The health consequences of smoking: Nicotine addiction: A report of the Surgeon General* (DHHS Publication No. (CDC) 88-8406). U. S. Department of Health and Human Services, Public Health Service, Centers for Disease Control.

Vaillant, G. E. (1983). *The natural history of alcoholism.* Cambridge: Harvard University Press.