Acta Genet Med Gemellol 39: 447-458 (1990)

Sixth International Congress
on Twin Studies

Persist or Quit?
Testing for a Genetic Contribution to Smoking Persistence

A.C. Heath

Department of Psychiatry, Washington University School of Medicine, St. Louis, Missouri, USA

Abstract. We consider three alternative parametric models to describe genetic and environmental influences on smoking initiation and smoking persistence. Under the single liability dimension model, the same genetic and environmental influences which determine smoking initiation also influence smoking persistence. Under the independent liability dimensions model, independent initiation and persistence dimensions determine onset of smoking, and persistence in those who become smokers. The combined model also postulates separate initiation and persistence dimensions, but allows for the possibility that some smokers are so low on liability to smoke on the initiation dimension that they become ex-smokers for this reason. Reanalyses of London twin data published by Eaves and Eysenck support the single liability dimension model. We discuss the difficulty of reconciling this finding with the hypothesis that nicotine dependence is a major determinant of smoking persistence, but caution that sample sizes in the London twin study were small.

Key words: Twins, Smoking persistence, Genetic models

INTRODUCTION

Several papers have explored the contributions of genetic and social factors to the initiation of substance use [eg, for alcohol: 7,10; or tobacco use: 1,4,14]. Others have examined their contribution to total level of consumption by those who are substance users [eg, quantity, frequency, or overall total level of alcohol consumption: 8,9,11,13; or average daily cigarette consumption: 12,14]. Analyses of the determinants of cessation of, versus persistence in, substance use have been much rarer. Such analyses may make an important contribution to our understanding of why some individuals experience so much more difficulty than others, for example, in quitting smoking [17].

Attempts to resolve the contribution of genetic factors to smoking cessation have been reported by Eaves and Eysenck [1] and by Hannah et al [6]. Analyses in the two studies were based on diametrically opposite assumptions. Hannah et al performed a nested analysis in which the determinants of smoking persistence in those who were smokers were assumed to be independent of the determinants of smoking initiation. Eaves and Eysenck assumed that the same liability continuum determined whether or not an individual started to smoke, and whether or not a smoker persisted in the smoking habit. While the analyses of Hannah et al are easy to reconcile with the hypothesis that nicotine dependence is a major determinant of smoking persistence [17], those of Eaves and Eysenck are less easily reconcilable. If the same genetic and environmental risk factors which influence smoking persistence also influence initiation of the smoking habit, as the Eaves and Eysenck model requires, then these presumably predate the onset of smoking.

As we have shown elsewhere in analyses of the inheritance of alcohol consumption patterns [8,9], the classical twin design provides a very powerful basis for resolving alternative models (eg, nested vs non-nested) for the natural history of substance use. Thus, the contrasting assumptions implicit in the analyses of Hannah et al and of Eaves and Eysenck can be tested. In this paper we show how this is possible, by reanalyzing the Eaves and Eysenck data.

METHODS

Overview of the Models

Fig. 1 presents three different hypotheses about the relationship between the determinants of smoking initiation and smoking persistence, each illustrated by a probability tree. The single liability dimension (SLD) model is the model used by Eaves and Eysenck [1] in their analyses. It assumes that the same liability dimension affects smoking onset and smoking persistence, and that persistent smokers are simply more extreme in liability (eg, more deviant in their personality scores [4]) than smokers who quit smoking. The independent liability dimensions (ILD) model is the model assumed by Hannah et al [6]. It specifies that the probability of being a current or ex-smoker is determined by two independent processes, the first of which ("initiation") determines the probability of being a smoker, and the second of which ("persistence") determines the conditional probability of continuing to smoke, given that an individual has become a smoker. The third model, which we have referred to as a Combined model, includes both the other models as special cases. It allows for the possibility that both the "initiation" process and the "persistence" process may generate ex-smokers. Ex-smokers from the initiation phase might include individuals who were only transiently smokers, and who never developed tolerance to the effects of nicotine. Ex-smokers from the persistence phase would include those who had become "committed" smokers [6] but had subsequently succeeded in quitting the habit.

The analyses of Eaves and Eysenck were conducted using a multiple threshold model, by the method of maximum likelihood [3]. These analyses were based on the

standard assumptions used in the estimation of a polychoric correlation [15], ie, that the observed discontinuous response distribution is determined by an underlying continuous normal "liability" distribution, and that the joint distribution of the liabilities of twin pairs is bivariate normal. The analyses of Hannah et al were performed within a non-

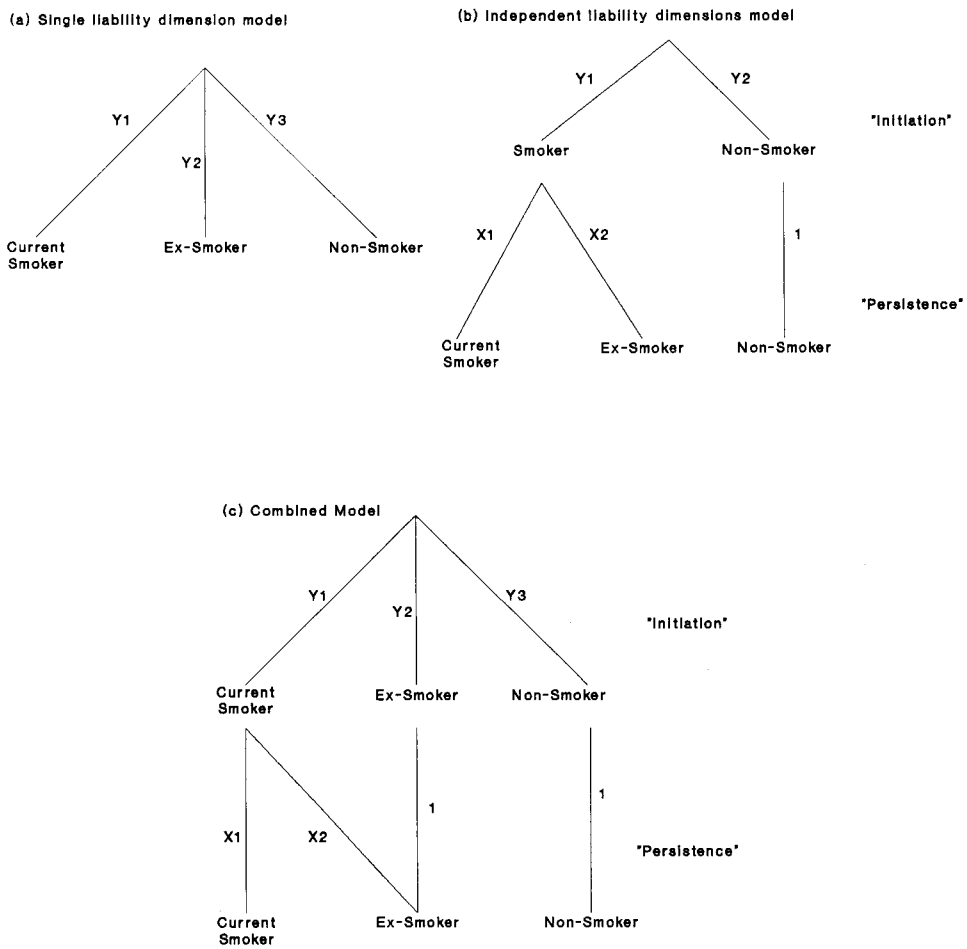

Fig. 1. Probability tree representations of single liability dimension, independent liability dimensions, and combined models.

parametric framework [see also 5]. Nonetheless, both the independent liability dimensions model and the combined model can be tested under the same parametric assumptions used by Eaves and Eysenck, again using the method of maximum likelihood [1,2,9].

https://doi.org/10.1017/S0001566000003676 Published online by Cambridge University Press

Data Summary

For each zygosity group, our data summary will consist of a two-way contingency table, cross-classifying the smoking status (current smoker, ex-smoker or non-smoker) of the first twin by the smoking status of the cotwin (see Table 1, which corrects an error in the original table of Eaves and Eysenck [1]). In like-sex pairs, twins were assigned as first or second twins on the basis of birth order, where this information was available, or otherwise at random. In unlike-sex pairs, the twins were reordered so that the female twin's status was cross-classified by that of the male twin.

Table 1 - Twin concordance for smoking status (current vs ex- vs non-smoker) in the London twin sample (after Eaves and Eysenck [1])

Twin 1/Female twin	Twin 2/Male twin		
	Current smoker	Ex-smoker	Non-smoker
MZ female pairs (n = 236)			
Current smoker	58	11	10
Ex-smoker	15	25	14
Non-smoker	11	25	67
MZ male pairs (n = 80)			
Current smoker	21	3	5
Ex-smoker	7	7	5
Non-smoker	6	7	19
DZ male pairs (n = 123)			
Current smoker	22	5	7
Ex-smoker	13	14	15
Non-smoker	9	10	28
DZ male pairs (n = 50)			
Current smoker	15	11	4
Ex-smoker	5	3	1
Non-smoker	3	3	5
DZ unlike-sex pairs (n = 58)			
Current smoker	13	6	4
Ex-smoker	6	4	3
Non-smoker	11	3	8

Note: We have corrected an error in Eaves and Eysenck [1], who switched the ordering of male and female twins from unlike-sex pairs.

https://doi.org/10.1017/S0001566000003676 Published online by Cambridge University Press

Maximum-Likelihood Estimation

Let $P_{i,j,k}$ denote the probability, under a given model, that a twin pair from the i-th zygosity group will fall in the j,k-th cell of the i-th contingency table. Thus $P_{1,1,1}$ will denote the probability that a female monozygotic twin pair will be concordant current smokers; $P_{1,2,1}$ will denote the probability that the first twin from a female MZ pair will be an ex-smoker and the second twin a current smoker, and so on. Let $f_{i,j,k}$ denote the observed frequency of twin pairs from the i-th twin group in the j,k-th cell of the i-th contingency table. The log-likelihood of a set of observations, under the given model, is given by

$$L = \ln (c) + \Sigma\Sigma\Sigma\ f_{i,j,k} \ln (P_{i,j,k}) \tag{1}$$

where c is a constant. Provided that we can generate the probabilities $P_{i,j,k}$ as a function of the parameters of a model, and provided that the model is identified, we can obtain maximum-likelihood estimates of the model parameters by maximizing this function with respect to the parameter values.

The fit of a given model can be assessed using 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 of the j,k-th cell of the i-th contingency table under the model. Provided that the expected cell frequencies are not too small (eg, no less than 2), the statistic C is 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. A significant chi-square indicates rejection of the model. If two models both give an adequate fit to the data, and one is a special case of the other, ie, some of the free parameters of the second model have been fixed to zero (or unity) in the first model, the improvement in fit of the more general model over the reduced model may be tested for significance by likelihood-ratio chi-square test. The difference between the chi-square values obtained under the two models is itself distributed as chi-square, with number of degrees of freedom equal to the number of free parameters in the second model which have been fixed in the first model.

Single Liability Dimension Model

Under the single liability dimension model [3,15,16], it is assumed that a normal liability distribution, with abrupt thresholds t_0, $t_1 \ldots t_3$ (in males) and t_0', $t_1' \ldots t_3'$ (in females) superimposed, determines an individual's smoking status. Threshold values $t_0 = t_0' = -\infty$, $t_3 = t_3' = +\infty$, and the values of t_1, t_2, t_1' and t_2' are to be estimated. No differences in threshold value are allowed as a function of zygosity or birth order, since these would not be predicted under any simple genetic model. A male respondent with

liability falling between t_0 and t_1 will be a current smoker; between t_1 and t_2 will be an ex-smoker; and between t_2 and t_3 will be a non-smoker. Thus the marginal probabilities that a male respondent will be a current smoker, ex-smoker or non-smoker (Y_1, Y_2 and Y_3 in Fig. 1a) are obtained by integrating a standardized normal distribution between the corresponding threshold values. Probabilities for female respondents are obtained in similar fashion.

As applied by Eaves and Eysenck [1], the single liability dimension model assumes that the joint distribution of twin pairs for the underlying latent "liability" variable is bivariate normal, with correlation r_i (which must be estimated as a separate parameter for each zygosity group). Thus the probability that two MZ female cotwins will be concordant current smokers, $Y_{1,1,1,}$ is obtained by integrating the bivariate normal distribution (with correlation r_1) from t_0', t_0', to t_1', t_1'. In general, the probability that a twin pair from the i-th zygosity group falls into the j,k-th cell of the i-th contingency table is given by

$$Y_{i,j,k} = \Phi\,(t_j,t_k) - \Phi\,(t_{j-1},t_k) - \Phi\,(t_j,t_{k-1}) + \Phi\,(t_{j-1},t_{k-1}) \tag{2}$$

where Φ is the bivariate normal distribution function with correlation r_i. Since, under the single liability dimension model, there is a direct one-to-one correspondence between the categories into which the underlying liability distribution is divided by the thresholds, and the response categories, $P_{i,j,k} = Y_{i,j,k}$ for all i,j,k, and we can obtain maximum-likelihood estimates of the twin correlations and threshold values by maximizing (1) above.

Independent Liability Dimension Model

Eaves et al [2] also considered the application of an independent liability dimensions version of the threshold model, although not applying it in the specific case of smoking data. For smoking status, we may postulate the existence of independent normal liability dimensions which determine the probability of becoming a smoker (we have called this the "initiation" dimension), and the conditional probability of quitting smoking, given that an individual has become a smoker (the "persistence" dimension). In non-smokers, the persistence dimension will have no effect. For the initiation dimension, there will now be only two threshold values to be estimated, t_1 and t_1', and we will have $t_0 = t_0' = -\infty$, as before and $t_2 = t_2' = +\infty$. Respondents with liability values below t_1 (or t_1') are predicted to be current or former smokers, those with higher liability values, to have never smoked. For the persistence dimension, we must estimate two further threshold values, s_1 and s_1', and again $s_0 = s_0' = -\infty$, $s_2 = s_2' = +\infty$. Smokers with liability values on the second dimension greater than s_1 (or s_1') are predicted to be quitters, the others to be persistent smokers. The probabilities that a male respondent will be a current smoker, ex-smoker, or non-smoker, will now be Y_1X_1, Y_1X_2 and Y_2, respectively, where Y_1 and Y_2 are the unconditional probabilities of initiating the smoking habit, or remaining a non-smoker, and X_1 and X_2 are the conditional probabilities of persisting in smoking, or quitting the habit, given that an individual has become a smoker, respectively (see Fig. 1).

Under the independent liability dimensions model, we assume that the joint distribution of twin pairs is bivariate normal, both for the "initiation" dimension and for the "persistence" dimension. Thus it will be necessary to estimate separate twin correlations, $r_{i,1}$ and $r_{i,2}$, for initiation and for persistence, separately for each twin group. $Y_{1,1,1}$ will now denote the probability that two female MZ cotwins will both be (current or former) smokers, $Y_{1,2,1}$ the probability that the first twin will be a non-smoker and the second twin a smoker, $Y_{1,2,2}$ the probability that both twins will be non-smokers, and so on. These probabilities will be obtained from function (2) as before, replacing r_i by r_{i1}. $X_{1,1,1}$ will be the conditional probability that two cotwins will both be smokers, $X_{1,1,2}$ the conditional probability that first twin will be a current smoker and the second twin an ex-smoker, and so on. These latter probabilities will be obtained by substituting threshold values s_i for t_i, $r_{i,2}$ for r, and $X_{i,j,k}$ for $Y_{i,j,k}$ in function (2).

Table 2 - Probability of observing a twin pair in the i,j-th cell of a two-way contingency table, under the independent liability dimensions model

Twin 1/Female twin	Twin 2/Male twin		
	Current smoker	Ex-smoker	Non-smoker
Current smoker	$Y_{11}\ X_{11}$	$Y_{11}\ X_{12}$	$Y_{12}\ (X_{11}+X_{12})$
Ex-smoker	$Y_{11}\ X_{21}$	$Y_{11}\ X_{22}$	$Y_{12}\ (X_{22}\ X_{21})$
Non-smoker	$Y_{21}\ (X_{11}+\ X_{21})$	$Y_{21}\ (X_{22}+X_{12})$	Y_{22}

The probability that two female MZ cotwins will be concordant ex-smokers, under the independent liability dimensions model, will be $Y_{1,1,1}\ X_{1,2,2}$. The probability that the first twin will be a non-smoker and the second twin an ex-smoker will be $Y_{1,2,1}(X_{1,1,2}+X_{1,2,2})$, *ie, the product of the unconditional probability that the twin pair* will be discordant for smoking status, and the marginal probability that a female smoker will be an ex-smoker rather than a current smoker. The probabilities $P_{i,j,k}$ are summarized in terms of Xs and Ys in Table 2. To simplify presentation, the first subscript identifying the zygosity group has been omitted from this table. As before, maximum likelihood estimates of threshold values and twin correlations (for persistence, and for initiation) are obtained by maximizing (1) with respect to the model parameters.

Combined Model

The SLD and ILD models have the same number of parameters. Thus, if they both give an adequate fit to the data by chi-square goodness-of-fit test, they cannot be directly compared to each other by likelihood-ratio chi-square test. Each model could be compared, however, to a more general model which includes both as special cases. The combined model of Fig. 1 is one example of a more general model. Like the ILD model, the combined model allows for independent initiation and persistence dimensions, and

so requires the estimation of two twin correlations for each zygosity group. Like the SLD model, the combined model allows for the possibility that some ex-smokers are ex-smokers by virtue of their position on the first, initiation dimension. Thus, the number of threshold values to be estimated for the initiation dimension will be the same as under the single liability dimension model, and the full model will require the estimation of two more thresholds than either the SLD or ILD models. The remaining ex-smokers are predicted to be those who are smokers on the initiation dimension, but become ex-smokers by virtue of their position on the persistence dimension. Probabilities $Y_{i,j,k}$ and $X_{i,j,k}$ are obtained by integrating the bivariate normal distributions for initiation, and for persistence, respectively. Expressions for the overall probabilities $P_{i,j,k}$ in terms of the Xs and Ys are summarized in Table 3, again omitting the zygosity group subscript for simplicity. Maximum-likelihood estimates of parameters are obtained by maximizing function (1) as before.

Table 3 - Probability of observing a twin pair in the i, j-th cell of a two-way contingency table, under the combined model

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

RESULTS

Table 4 summarizes the goodness-of-fit chi-squares obtained when the SLD, ILD and combined models were fitted to the Eaves and Eysenck [1] data. The independent liability dimensions model is rejected by chi-square test of goodness-of-fit. The single liability dimension model gives a barely adequate fit to the data. The combined model also gives an adequate fit, but does not give a significantly better fit than the SLD model, by likelihood-ratio chi-square ($\chi^2 = 10.92$, $df = 7$, $p = 0.14$). These data do not give any grounds for rejecting the single liability dimension model favored by Eaves and Eysenck in their original analysis.

Table 4 - Comparison of goodness-of-fit chi-squares, obtained under different models

		SLD model	ILD model	Combined model
Test of goodness-of-fit	df	31	26	24
	χ^2	37.44	49.37	26.52
	p	0.20	0.004	0.33

https://doi.org/10.1017/S0001566000003676 Published online by Cambridge University Press

Fig. 2 summarizes, in the form of two probability trees, the marginal unconditional (Y) and conditional (X) probabilities estimated for females and for males under the combined model. These estimates help explain why the independent liability dimensions model gave such a poor fit to the data. It appears that some 42.6% of the male ex-smokers, and 72.7% of the female ex-smokers, are ex-smokers by virtue of their position on the initiation dimension. The persistence dimension is having a relatively minor impact, especially in female smokers.

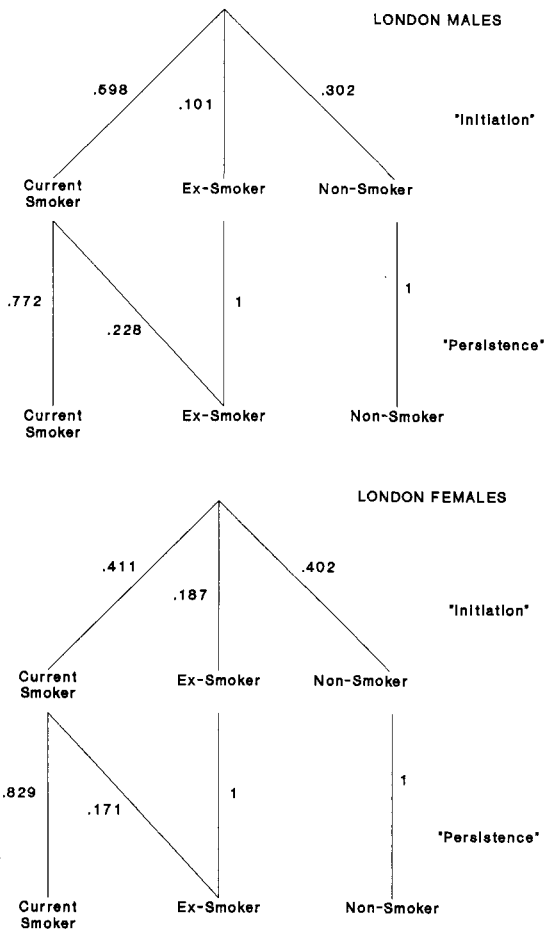

Fig. 2. Probability estimates under the full combined model.

Table 5 compares maximum-likelihood estimates of twin polychoric correlations under the three models. Comparison of the estimates obtained under the different models emphasizes the importance of testing the underlying assumptions about the relationship between the determinants of smoking initiation and the determinants of smoking persistence. Under the single liability dimension model, which does fit the data, we

https://doi.org/10.1017/S0001566000003676 Published online by Cambridge University Press

Table 5 - Maximum likelihood estimates of polychoric correlations under different models

	SLD model	ILD model		Combined model	
	Smoking status	Initiation	Persistence	Initiation	Persistence
MZ female pairs	0.70	0.68	0.71	0.70	0.99[a]
MZ male pairs	0.59	0.59	0.60	0.62	0.79
DZ female pairs	0.51	0.46	0.48	0.50	0.93
DZ male pairs	0.37	0.61	−0.13	0.57	−0.33
DZ unlike-sex pairs	0.18	0.32	0.13	0.26	0.26

[a] Parameter is fixed on its upper bound of 0.99.

observe higher MZ than DZ correlations in both like-sex groups, consistent with a moderate genetic effect on the smoking initiation/persistence dimension. Had we just fitted the independent liability dimensions model, which does not fit these data, we would have observed a higher MZ than DZ correlation, for the initiation dimension, in female like-sex pairs only, and concluded that twin concordance for smoking initiation in male pairs was a purely environmental phenomenon. We would, however, have inferred an independent genetic effect on persistence, in both sexes. Under the more general combined model, we again observe only a very slight difference between the male like-sex MZ and DZ correlations for the initiation dimension, but more pronounced evidence for genetic effects in females. Estimates of correlations for the persistence dimension, however, are unreasonably high for the two MZ and the female like-sex DZ groups, and strongly negative for the male like-sex group. In the female MZ group, the twin correlation for persistence has gone to its upper bound of 0.99! These nonsensical values, taken together with the non-significant improvement in fit compared to the single liability dimension model, suggest that in fitting the combined model to these data we are trying to fit a more elaborate model than the observed pattern of twin concordances (and the small sample sizes) will justify.

By reparameterizing the correlations r_i (or $r_{i,2}$) under a given model as a function of genetic and environmental parameters, it is possible to compare the fit of genetic and non-genetic models [3,9]. In the present case, since the single liability dimension model used by Eaves and Eysenck was found to give the best fit to the data, their original conclusions, that there is significant evidence for genetic effects on smoking status (accounting for approximately 68% of the variance in the initiation/persistence dimension), will still stand.

DISCUSSION

Our findings justify the original decision of Eaves and Eysenck [1] to fit a conventional threshold model (our single liability dimension model) to these data. They do not justify the alternative approach of Hannah et al [6], admittedly using a different data set, to

analyze smoking status under a nested model. The latter model would be equivalent to our independent liability dimensions model, which was rejected by the London data. It thus appears that the same genetic and environmental factors which influence smoking initiation also determine smoking persistence. If replicated in other data-sets, this finding will be difficult to reconcile with the notion that smoking persistence is largely related to the development of nicotine dependence [17]. Under such a hypothesis, we might expect to find that different factors determine the initiation of smoking, and the extent to which nicotine dependence develops once smoking has started.

Sample sizes in the London study were small, and we would not therefore wish to give undue emphasis to this conclusion. It is possible that in larger samples, involving larger numbers of ex-smokers, the combined model would be found to give a significant improvement in fit over either simple model. The models which we have presented in this paper, however, help to illustrate how genetic models can be used to study the cessation of substance use, not just initiation of use and quantity and frequency of use.

Acknowledgments. Supported by ADAMHA grants DA05588, AA07535 and AA07728.

REFERENCES

1. Eaves LJ, Eysenck HJ (1980): New approaches to the analysis of twin data and their application to smoking behavior. In Eysenck HJ (ed): The Causes and Effects of Smoking. London: Maurice Temple Smith, pp 158-235.
2. Eaves LJ, Eysenck HJ, Martin NG (1989): Genes, Culture and Personality: An Empirical Approach. London: Academic Press.
3. Eaves LJ, Last K, Young PA, Martin NG (1978): Model-fitting approaches to the analysis of human behavior. Heredity 41:249-320.
4. Eysenck HJ (1973): Personality and the maintenance of the smoking habit. In Dunn WL (ed): Smoking Behavior: Motives and Incentives. Washington, DC: VH Winston and Sons, pp 113-146.
5. Hannah MC, Hopper JL, Mathews JD (1983): Twin concordance for a binary trait. I: Statistical models illustrated with data on drinking status. Acta Genet Med Gemellol 32:127-138.
6. Hannah MC, Hopper JL, Mathews JD (1984): Twin concordance for a binary trait. II: Nested analysis of ever-smoking and ex-smoking traits and umested analysis of a committed-smoking trait. Am J Hum Genet 37:153-165.
7. Heath AC, Martin NG (1988): Teenage alcohol use in the Australian twin register: Genetic and social determinants of starting to drink. Alcoholism Clin Exp Res 12:735-741.
8. Heath AC, Meyer J, Eaves LJ, Martin NG (1991): The inheritance of alcohol consumption patterns in a general population twin sample: I. Multidimensional scaling of quantity/frequency data. J Stud Alc (in press).
9. Heath AC, Meyer J, Jardine R, Martin NG (1991): The inheritance of alcohol consumption patterns in a general population twin sample: II. Determinants of consumption frequency and quantity consumed. J Stud Alc (in press).
10. Heath AC, Meyer J, Martin NG (1991): The inheritance of alcohol consumption patterns in the Australian Twin Survey, 1981. In: Genetics and Biology of Alcoholism, Banbury Report 33 (in press).

11. Jardine R, Martin NG (1984): Causes of variation in drinking habits in a large twin sample. Acta Genet Med Gemellol 33:435-450.
12. Kaprio J, Koskenvuo M, Sarna S (1981): Cigarette smoking, use of alcohol, and leisure-activity among same-sexed adult twins. In Gedda L, Parisi P, Nance WE (eds): Twin Research 3: Part C. New York: Alan R Liss, pp 37-46.
13. Kaprio J, Koskenvuo MD, Langinvainio H, Romanov K, Sarna S, Rose RJ (1987): Genetic influences on use and abuse of alcohol: A study of 5,638 adult Finnish brothers. Alcoholism Clin Exp Res 11:349-356.
14. Meyer JM, Heath AC, Martin NG, Eaves LJ (1191): Genetic and environmental influences on onset of the smoking habit and quantity smoked. (In review).
15. Olsson H (1979): Maximum-likelihood estimation of the polychoric coefficient, Psychometrika 44:443-460.
16. Tallis GM (1962): The maximum-likelihood estimation of correlation from contingency tables. Biometrics 18:342-353.
17. US Department of Health and Human Services (1988): The Health Consequences of Smoking: Nicotine Addiction. A Report of the Surgeon General, Public Health Service, Centers for Disease Control. DHHS Publication No. (CDC) 88-8406.

Correspondence. Dr. Andrew C. Heath, Department of Psychiatry, Washington University School of Medicine, 4940 Audubon Avenue, St. Louis, MO 63110, USA.