Am J Hum Genet 37:153–165, 1985

Twin Concordance for a Binary Trait. II. Nested Analysis of Ever-Smoking and Ex-Smoking Traits and Unnested Analysis of a "Committed-Smoking" Trait

MURRAY C. HANNAH,[1] JOHN L. HOPPER, AND JOHN D. MATHEWS

SUMMARY

Twin concordance rates for a binary trait can provide information about causes of trait variation. However, if trait prevalence varies with age (or birth cohort) or between the sexes, trait concordance rates will be artificially inflated because of the matching within pairs of twins. Our previous paper showed how to minimize the effects of such confounding by using logistic regression to model trait prevalence as a function of age and sex and that the binary correlation coefficient was useful as a measure of concordance that can be adjusted for trait prevalence. This method is extended here to allow for nested analyses and is applied to the smoking habits of a sample of 3,807 pairs of adult twins. For monozygotic (MZ) twins, the correlation coefficients for the binary trait of "ever-smoking" (males: .50 ± .04; females: .60 ± .02) were significantly greater than for dizygotic (DZ) twins (males: .37 ± .05; females: .31 ± .04; unlike-sex pairs: .21 ± .03). For "giving-up smoking," given that both twins were previously smokers, the correlations for MZ twins (males: .37 ± .07; females: .29 ± .05) were also greater than for DZ twins (males: .11 ± .09; females: .26 ± .08; unlike-sex pairs: .13 ± .06), although the difference was not statistically significant for females.

Current smokers who had been smoking for at least 10 years were arbitrarily defined as "committed-smokers." The binary trait of "committed-smoking" was more strongly correlated in MZ twins (males: .41 ± .06; females: .41 ± .04) than in DZ twins (males: .22 ± .08; females: .18 ± .05; unlike-sex pairs: .16 ± .05). These observations suggest that as well as depending on socially determined environmental factors,

Received March 20, 1984; revised July 10, 1984.

This work was supported by the National Health and Medical Research Council of Australia, the Australian Tobacco Research Foundation, the Victor Hurley Fund of The Royal Melbourne Hospital, and the Anti-Cancer Council of Victoria.

[1] All authors: University of Melbourne, Department of Medicine, Royal Melbourne Hospital, Victoria 3050, Australia.

 0002–9297/85/3701–0014$02.00

smoking behavior is influenced by genetic factors and/or by environmental factors unique to the MZ twin environment, which are of particular importance as determinants of "committed-smoking." There is a need for further research to investigate the personal characteristics of "committed-smokers" and to seek intervention strategies that are more suited to the needs of individual smokers.

INTRODUCTION

Under the assumption of homogeneity of environment, concordance rates in MZ and DZ twins can be used to indicate the relative importance of genetic and environmental determinants of a binary trait. However, many traits are dependent on variables such as age (or birth cohort) and sex, and this confounding can interfere with the estimation of twin concordance rates. Logistic regression techniques have been introduced to measure and to make adjustment for the effects of these confounding variables while estimating concordance rates in twins [1]. Such models provide a useful descriptive summary of the data, facilitate detailed statistical inference, and provide a basis for biological interpretation.

Our study extends these methods to an analysis of the smoking behavior of adult twins. To deal with the problem posed by individuals who have been smokers but have stopped, the model treats smoking status as a "nested" binary trait. All individuals are classified as belonging to one of two mutually exclusive categories depending on whether they have ever or never been a smoker, thereby defining a "primary binary trait." "Ever-smokers" are then subclassified into one of two mutually exclusive subcategories: "current" or "ex-smoker," defining a "secondary binary trait."

This nested scheme is shown to provide a useful framework for the analysis of smoking behavior in twins. Interpretation of these results is facilitated by an additional analysis of the unnested binary trait of "committed-smoking," arbitrarily defined as current smoking for a duration of at least 10 years.

MATERIALS AND METHODS

Statistical Model

Consider a nested binary trait measured on a single pair of twins. Let X_k be 1 if the primary trait is present (ever-smoker) and 0 if it is absent (never-smoker) in twin k, $k = 1, 2$. Let the conditional random variable $Y_k|(X_k = 1)$ be 1 if the secondary trait is present (ex-smoker), 0 if it is absent (current-smoker), and undefined if $X_k = 0$, $k = 1, 2$.

Define $\pi_1 = P\{X_k = 1\}$ and $\pi_2 = P\{Y_k = 1|X_k = 1\}$, $k = 1, 2$. For simplicity of exposition, we have assumed that π_1 and π_2 are the same for both twins. This is not a necessary assumption and easily can be relaxed for the application of the model to unlike-sex DZ pairs. The probabilities P_{ij} $(i, j = 1, 2, 3)$ of all possible outcomes may be summarized in a 3×3 table (table 1).

Following [1], each P_{ij} may be expressed in terms of π_1, π_2, and the within pair (binary) correlation coefficients: ρ_1 for the primary trait and ρ_2 for the secondary trait conditional on $X_1 = 1$ and $X_2 = 1$. Let $\delta_i = \rho_i\pi_i(1 - \pi_i)$, $i = 1, 2$. Then

TABLE 1

PROBABILITIES FOR A NESTED BINARY TRAIT MEASURED ON A TWIN PAIR

		Twin 1			
			$X_1 = 1$		
Twin 2		$X_1 = 0$	$Y_1 = 0$	$Y_1 = 1$	
$X_2 = 0$		P_{11}	P_{12}	P_{13}	$1 - \pi_1$
$X_2 = 1$	$Y_2 = 0$	P_{21}	P_{22}	P_{23}	$\pi_1(1 - \pi_2)$
	$Y_2 = 1$	P_{31}	P_{32}	P_{33}	$\pi_1\pi_2$
		$1 - \pi_1$	$\pi_1(1 - \pi_2)$	$\pi_1\pi_2$	1

NOTE: X_k is 1 if the primary trait is present, and 0 otherwise, and Y_k is 1 if the secondary trait is present, and 0 otherwise (conditional on $X_k = 1$), for twin $k = 1, 2$. π_1 is the prevalence of the primary trait, and π_2, the prevalence of the secondary trait given the primary trait.

$$
\begin{aligned}
P_{11} &= P\{X_1 = 0, X_2 = 0\} \qquad (1)\\
&= (1 - \pi_1)^2 + \delta_1\\
P_{22} &= P\{Y_1 = 0, Y_2 = 0 | X_1 = 1, X_2 = 1\}P\{X_1 = 1, X_2 = 1\}\\
&= [(1 - \pi_2)^2 + \delta_2][\pi_1^2 + \delta_1]\\
P_{23} &= P_{32} = [\pi_2(1 - \pi_2) - \delta_2][\pi_1^2 + \delta_1]\\
P_{33} &= [\pi_2^2 + \delta_2][\pi_1^2 + \delta_1]\\
P_{12} &= P_{21} = (1 - \pi_2)[\pi_1(1 - \pi_1) - \delta_1]\\
P_{13} &= P_{31} = \pi_2[\pi_1(1 - \pi_1) - \delta_1] \ .
\end{aligned}
$$

Note that P_{12} and P_{13} involve π_1, π_2, and ρ_1, but not ρ_2.

π_1 and π_2 may depend upon a vector of explanatory variables, $\mathbf{z}$, such as age and sex that can be modeled as logistic functions:

$$\pi_i = \frac{\exp(\boldsymbol{\alpha}_i' \cdot \mathbf{z}_i)}{1 + \exp(\boldsymbol{\alpha}_i' \cdot \mathbf{z}_i)} \ , \qquad (2)$$

where $\boldsymbol{\alpha}_i$ is a vector of constants, $i = 1, 2$. The log likelihood for a given parameter vector $(\boldsymbol{\rho}, \boldsymbol{\alpha})$ can be formed by simple summation over the $\log(P_{ij})$ values for the observed data on each twin pair in the sample. With the aid of a suitable computer and maximization routine (we have used MAXLIK; [2]), the parameter values that maximize the log likelihood (that is, the maximum likelihood estimates or MLEs) may be calculated, and statistical inference based on well-known asymptotic likelihood theory (see, e.g., [3]). For example, the statistic

$$Z = \frac{\hat{\rho}_{mz} - \hat{\rho}_{dz}}{\{\text{Var}\ \hat{\rho}_{mz} + \text{Var}\ \hat{\rho}_{dz} - \text{Cov}(\hat{\rho}_{mz}, \hat{\rho}_{dz})\}^{1/2}} \qquad (3)$$

(where "^" indicates MLEs, and variance-covariance estimates are derived from the observed information matrix) is asymptotically normal in distribution and may be used to test the

difference between correlation coefficients. Simulations [4] based on effectively 10,000 repetitions of Z over each of a wide variety of π and ρ values gave 95% confidence intervals for the actual type I error rate of (.049, .052) at the nominal .05 level and (.010, .012) at the nominal .01 level, provided all expected cell counts (i.e., the expected number of twin pairs in each of the outcome categories) were at least 18. Since the minimum observed cell count is 21 (tables 2 and 4), equation (3) should provide a highly reliable test in the present study.

Data

Data on smoking habits were obtained by a postal questionnaire sent to all twins on the Australian National Health and Medical Research Council Twin Registry who were at least 18 years of age. The registry was established by voluntary recruitment of twins in 1976–1980; questionnaires were mailed between November 1980 and March 1982 and completed and returned by both members of 3,807 pairs, representing a 64% pairwise response rate. The data for each individual contains information on age, sex, zygosity, smoking status, and general health and psychological factors. Diagnosis of zygosity was given by self-report unless twins differed, in which case a recent photograph was sought (see [5]).

Individuals were classified as "ever-smokers" or "never-smokers" according to their answers to the questionnaire, and each "ever-smoker" was subclassified as a "current-smoker" or an "ex-smoker." The data are summarized by smoking status, sex, and zygosity in table 2.

To simplify the model, the initial analysis was restricted to the 2,901 like-sex twin pairs. Although the total number of parameters in the model was thereby reduced by only two (viz., the "ever-smoking" and "ex-smoking [given ever-smoking]" correlation coefficients for unlike-sex DZ twins), this simplification enabled estimation to proceed independently for each sex, reducing the dimensionality of each likelihood surface by more than one-half and, hence, greatly diminishing the amount of work time and space required for computation. Correlation coefficients for DZ unlike-sex pairs were computed subsequently using the like-sex adjustments for age and sex.

ANALYSIS AND RESULTS

Plots of the proportion of ever-smokers against age are given in figure 1. For males, the proportion rises progressively with age, whereas for females, it increases sharply at age 18 and then falls gradually, indicating the presence of a strong cohort effect. Figure 2 shows that the proportion of ex-smokers from among ever-smokers increases progressively with age for each sex. These plots suggest

TABLE 2

SMOKING STATUS, SEX, AND ZYGOSITY OF 3,807 TWIN PAIRS

	MALE		FEMALE		UNLIKE-SEX	
	MZ	DZ	MZ	DZ	DZ	TOTAL
Both never smoked	221	121	630	308	266	1,546
One never–One current	58	60	124	147	206	595
One never–One ex-smoker	80	46	110	102	172	510
Both current smokers	74	43	190	99	116	522
One current–One ex-smoker	59	53	114	61	105	392
Both ex-smokers	74	29	65	33	41	242
Total	566	352	1,233	750	906	3,807

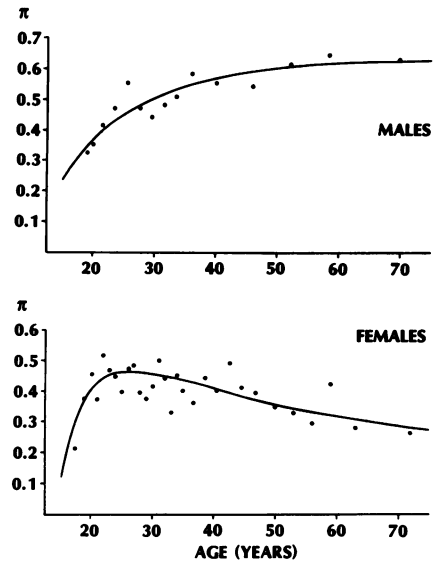

FIG. 1.—The proportion of individuals who have ever-been smokers (π) against age. *Plotted points* represent the observed proportion in age groups of approximately 100 individuals against the mean age of the group. *The continuous curve* is given by the fitted logistic model (table 3A) with age power transformation (table C).

that age-adjustment terms should be incorporated into the model. Accordingly, quadratic logistic models were used to adjust the frequencies of ever-smokers for age and linear logistic models for the frequencies of ex-smokers. A Box-and-Cox-type power transformation [6] of the age scale was incorporated to allow for asymmetry of the "quadratic" effects and nonlinearity of "linear" effects on the natural scale [1].

Table 3 presents the results of the nested analysis of smoking habits for male and female like-sex and unlike-sex twin pairs; correlation coefficients and age-adjustment parameters for the proportions of ever-smokers (primary trait) and ex-smokers (secondary trait) were estimated concurrently.

In the analysis of smoking status the following were observed: (1) Both proportions of ever-smokers and of ex-smokers among ever-smokers were highly age dependent (figs. 1 and 2). The likelihood-ratio chi-square statistic (on 4 degrees of freedom) for all the age adjustment parameters (i.e., testing H_0: $\alpha_1 = \alpha_2 = \beta_1 = 0$ and $\lambda = 1$) is 63.3 for males and 94.0 for females; highly significant in each case. (2) Both proportions of ever-smokers and of ex-smokers among ever-smokers tended to be higher in males than in females (table 3 and figs. 1 and 2). Taking into account the parameters π_i, α_i ($i = 1, 2$), β, and λ, the likelihood-ratio test for a difference between the sexes gives $\chi^2_6 = 112.5$ ($P < .001$). (3) The estimates of within pair correlation for ever having been a smoker were significantly higher among MZ than DZ pairs for both males and females. Using estimates from table 3, the statistic (3) for testing differences between MZ and DZ within pair correlations becomes $Z = 2.03$ ($P = .02$) for

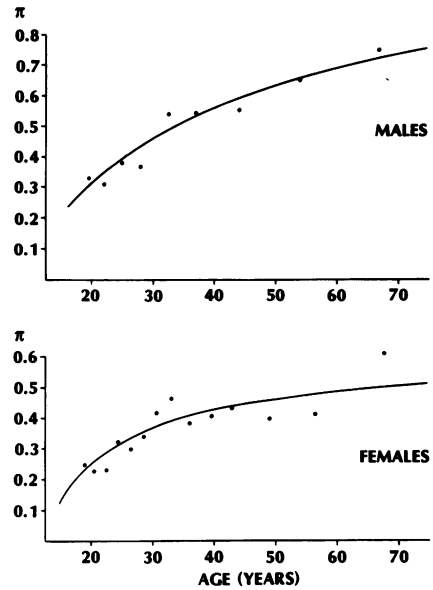

FIG. 2.—The proportion of individuals who have stopped smoking from among those who have ever-been smokers (π) against age. *Plotted points* represent the observed proportion in age groups of approximately 100 individuals against the mean age of the group. *The continuous curve* is given by the fitted logistic model (table 3B) with age power transformation (table 3C).

males, and $Z = 6.89$ ($P < .001$) for females (one-tailed tests). (4) Similarly, it can be shown that the MZ correlation for ever-smoking was higher for females than for males ($P = .02$, two-tailed test). (5) The estimated correlation for giving-up smoking in pairs where both were smokers was higher for MZ than for DZ pairs. The difference was significant for males ($P = .01$), but not for females ($P = .38$, one-tailed tests). (6) The correlation for ever-smoking was lower for unlike-sex DZ pairs than for like-sex DZ pairs ($P < .03$ males, $P < .02$ females). No significant differences were found between correlations in DZ twins for giving-up smoking.

Committed-Smokers

In view of the preceding results (see DISCUSSION), a new binary trait of "committed-smoking" was arbitrarily defined for current smokers who had been smoking for at least 10 years. For analysis of this binary trait, the data were restricted to individuals of at least 28 years of age in order to allow time for the trait to be realized. The resulting smaller sample is summarized in table 4 according to committed-smoking status, sex, and zygosity. Figure 3 shows the proportion of committed-smokers against age for males and females. A simple (unnested) model [1] was used to analyze these data, and parameter estimates for the saturated model are given in table 5. The similarity between the sexes for this trait is striking (fig. 3 and table 5). There are no significant differences between males and females, whether this be measured on the parameter estimates individually

using asymptotic normality of MLEs (table 5) or by the likelihood-ratio criterion over the entire model; ($\chi^2_5 = 2.74$).

Table 4 indicates a significant difference between male MZ and DZ twins in the *prevalence* of committed-smoking ($\chi^2_1 = 10.4$). This difference cannot be explained easily in terms of differences in age distribution, which are not dissimilar for the two groups ($\chi^2_{10} = 14.3$, $P \simeq .2$). The age-dependent prevalence rates were modeled separately for male MZ and DZ twins, but the resulting increase in log likelihood failed to indicate a significant improvement in fit ($\chi^2_3 = 5.6$). Indeed, most of the change in likelihood was accounted for by a difference in the constant term (π) alone ($\chi^2_1 = 5.4$, $P = .02$), giving $\hat{\pi}_{mz} = .16$ and $\hat{\pi}_{dz} = .25$ (when allowing no zygosity difference for the α's and β's). The properly (age and zygosity) adjusted correlation coefficients for males are then $\hat{\rho}_{mmz} = .414 \pm .060$ and $\hat{\rho}_{mdz} = .220 \pm .079$ (significantly different, $P < .05$), altering slightly the values reported in table 5, but leaving the inference essentially unchanged.

For unlike-sex twins, the correlation coefficient for committed-smoking was $.16 \pm .05$, a value comparable with those for the like-sex DZ twins ($.22 \pm .08$ for males and $.18 \pm .05$ for females).

TABLE 3

NESTED MODEL PARAMETER ESTIMATES

PARAMETER		MALES		FEMALES	
		MLE*	SE†	MLE*	SE†
A. Primary trait (ever-smoking)					
MZ correlation	ρ_{mz}	.500	.037	.600	.023
DZ correlation	ρ_{dz}	.374	.050	.312	.035
DZ unlike-sex correlation	ρ_{udz}	.126	.062	...	...
Constant term	π_1	.372	.026	.405	.020
Linear-age coefficient	α_1	0.884	0.347	1.077	0.479
Quadratic-age coefficient	β_1	−0.185	0.156	−1.255	0.812
B. Secondary trait (giving-up smoking) conditional on primary trait					
MZ correlation	ρ_{mz}	.369	.068	.287	.054
DZ correlation	ρ_{dz}	.110	.088	.258	.075
DZ unlike-sex correlation	ρ_{udz}	.126	.062	...	...
Constant term	π_2	.313	.033	.250	.023
Linear-age coefficient	α_2	0.842	0.337	0.934	0.312
C. Age-scale power transformation					
	λ	0.190	0.438	−0.487	0.440

* Maximum likelihood estimate.

† Asymptotic standard error from observed inverse information matrix. In terms of equation (2), $\boldsymbol{\alpha}_1' \cdot \mathbf{z}_1 = \kappa_1 + \alpha_1 t + \beta_1 t^2$ and $\boldsymbol{\alpha}_2' \cdot \mathbf{z}_2 = \kappa_2 + \alpha_2 t$, where $t = \{[(\text{age in years} - 10)/10]^\lambda - 1\}/\lambda$ is the scaled age, and the constant term $\pi_i = e^{\kappa_i}/(1 + e^{\kappa_i})$ may be interpreted as the prevalence for 20-year-olds.

TABLE 4

COMMITTED-SMOKING STATUS OF 1,788 TWIN PAIRS

	MALES		FEMALES		UNLIKE-SEX	
	MZ	DZ	MZ	DZ	DZ	TOTAL
Neither committed	250	104	545	305	277	1,204
One committed	65	55	151	131	160	402
Both committed	37	21	88	36	45	182
Total	352	180	784	472	482	1,788
Prevalence (%)	19.7	26.9	20.9	21.5	25.9	21.4

DISCUSSION

Cigarette smoking has changed dramatically since World War I; it appears to have been influenced by personal, social, and economic factors [7–9], which have themselves changed with the passage of time. Consequently, in any cross-sectional (prevalence) study of smoking behavior, the age and sex distribution of ever-smoking and ex-smoking traits will depend on the differential effects of social and economic factors acting on different birth cohorts. In a cross-sectional analysis, the effects of birth cohort on trait prevalence are completely confounded with the effects of age. With the methods of analysis used here, it is possible to estimate the effects of age/cohort and to calculate a twin correlation for each binary trait that is less subject to influence by these confounding factors.

Sample Selection

The wider validity of any generalized conclusions from this analysis must depend on the extent to which the twin sample can be regarded as being representative of the Australian population. The Australian National Health and Medical Research Council Twin Registry was established by asking twins to volunteer their names. Registered adult twins were asked to complete the postal questionnaire also on a voluntary basis; this resulted in a pairwise response rate of 64%. Thus, the questionnaire respondents were subject to a number of potential selective biases, and it is unlikely that they represent a random sample of Australians of the same age and sex distribution. In particular, it is possible that the questionnaire respondents are somewhat less likely to have been ever-smokers, and more likely to be ex-smokers, than the general Australian population; this may reflect the underrepresentation of economically disadvantaged individuals among the respondents.

Such a sampling bias, however, is unlikely to seriously invalidate the conclusions from the study unless the sampling fraction depended on zygosity and on twin concordance for smoking, as well as on smoking habits per se. In this connection, there does appear to be a net zygosity difference in the proportions of male committed-smokers (see RESULTS) that should alert us to the possibility of such a selection bias. The observed difference *could* reflect an increased ascertainment probability or relative response rate for male MZ concordant never- or ex-smokers.

Nevertheless, even if this is so, there is little evidence of it greatly affecting overall twin concordance.

An artifactual increase in the observed concordance for smoking habits in MZ pairs of twins *could* occur without any manifest zygosity difference in smoking prevalence if the positive selective bias for concordant pairs (both smokers or both nonsmokers) were greater in magnitude for MZ pairs than for DZ pairs. Although we cannot formally exclude such a possibility, we believe that it is a simpler hypothesis to suppose that the observed zygosity differences in smoking concordance are real rather than artifactual, and we note that similar zygosity differences in smoking concordance have been reported in several other twin studies [7, 10].

Advantages of the Statistical Model

The "nested" analysis of "ever-smoking" and "ex-smoking" traits appears to be a useful innovation. If the nesting is ignored, and the two traits are analyzed separately, information is lost because pairs discordant for the primary trait would be excluded from the analysis of the secondary trait even though they contribute information on its marginal distribution and, hence, indirectly, on trait concordance. This additional information is of particular relevance in the presence of explanatory variables that may affect the prevalence rate; the marginal effect of a variable such as age (or cohort) is obviously of intrinsic interest, as well as providing a means for estimation of an age-adjusted twin correlation.

A number of methods for the analysis of pedigree (including twin) data have been proposed in the literature [7, 11–13] that depend heavily upon assumptions of distribution, scale, or mode of gene action. In practice, these assumptions may be difficult or impossible to test. As in our earlier paper [1], we have used the binary correlation coefficient as the underlying measure of twin concordance.

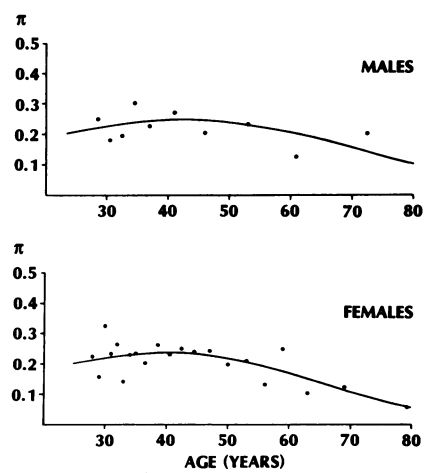

FIG. 3.—The proportion of individuals who are committed-smokers (π) against age. *Plotted points* represent the observed proportion in age groups of approximately 100 individuals against the mean age of the group. *The continuous curve* is given by the fitted logistic model (table 5).

TABLE 5

ESTIMATES AND STANDARD ERRORS FOR COMMITTED-SMOKING MODEL

Parameter*	MLE	SE
ρ_{mmz}	.434	.060
ρ_{fmz}	.411	.040
ρ_{mdz}	.210	.075
ρ_{fdz}	.175	.050
ρ_{udz}	.161	.048
π_m	.185	.053
π_f	.173	.035
α_m	0.318	0.299
α_f	0.405	0.210
β_m	−0.071	0.053
β_f	−0.103	0.038

NOTE: The age-transformation parameter was estimated as $\hat{\lambda} = 1.010 \pm 0.894$ for all the data (pooled over sex) and was subsequently taken to be unity.

* Subscripts refer to sex and zygosity.

This parameterization is justified in terms of its descriptive value, its conceptual simplicity, the ease of estimation of marginal distributions, and, heuristically, in terms of the stability of the correlation coefficient after allowance has been made for the confounding effect, on the marginal probability, of a covariate such as age. Furthermore, it should be noted that parameterizations of this form can be extended to allow for more explicit causal models, for the estimation of the fixed (marginal) effects of measured causal variables, and for the estimation of random effects of several causes, not explicitly measured, through a combination of one or more correlation coefficients.

Interpretation of Results

The present results indicate that the MZ twin correlations are significantly greater than DZ correlations for "ever-smoking" in both males and females and for giving-up smoking in males but not in females. Furthermore, the prevalences of ever-smoking, and of ex-smoking given previous smoking, are greater for (older) males than for females. These observations could be explained by postulating that past environmental pressures to smoke have been somewhat less effective in females than in males; if the *intrinsic* predisposition to smoke were equal for males and females, this would have led to a greater proportion of (highly) predisposed individuals among (older) female ever-smokers than among male ever-smokers. This, in turn, could explain why proportionally fewer females than males appear to give up smoking. Furthermore, those females who do give up smoking would tend to do so for reasons other than those that are attributable to the effects of genetic factors or to factors in the shared MZ twin (family) environment that might influence the intrinsic predisposition to smoke.

It is also of particular interest to compare the age-specific proportion of ever-smokers in males and females (fig. 1). For females, the maximum prevalence

occurs around the age of 25, whereas the male prevalence approaches a maximum in old age; this reflects the tendency for smoking to become more frequent in later-born cohorts of women and contrasts with the slightly reduced prevalence rates of ever-smoking in men born since World War II. These trends, in turn, could reflect the relaxation of the environmental and social constraints on women smoking and the effect of health-education programs in reducing smoking in young men.

Although the DZ correlation for ever-smoking in unlike-sex pairs is less than for like-sex pairs, it would be presumptuous to attribute this to differences in gene expression between the sexes when the observed age and sex differences in the marginal distribution (smoking prevalence) are more plausibly interpreted as being due to interactions between sex and the changing social environment. Rather, it seems likely that the lower correlation for unlike-sex DZ twins is due to an interaction between sex and factors in the family environment.

Interpretation is simpler for the less volatile trait of "committed-smoking." All within pair correlation coefficient estimates are significantly greater than zero, indicating a significant family (or twin pair) effect that may be due to genetic and/or shared environmental factors. The MZ correlation coefficients are much less than unity, which implies a considerable role for individual nongenetic factors in determining committed-smoking behavior. The difference between MZ and DZ twin correlation estimates is highly significant for males ($P < .01$, one-sided test) and for females ($P < .001$, one-sided test). Furthermore, for each sex, the DZ correlation estimate is close to one-half the estimated value for MZ pairs. Although this relationship between the MZ and DZ correlations has not been given a precise parametric interpretation in terms of a particular genetic model, these observations suggest that genetic predisposition plays an influential role in determining committed-smoking status. The observed similarity between unlike-sex and like-sex DZ correlation estimates is consistent with homogeneous gene expression in both sexes.

These conclusions are in close agreement with those of Eysenck and Eaves [7] in their analysis of "never," "used to," and "currently do" smoke as a quantitative trait in liability. However, in both studies, it is formally impossible to exclude the alternative, but less parsimonious, explanation that in MZ twins there is a greater effect of shared environment for becoming a committed-smoker.

The details of the age dependence of the proportion of committed-smokers are of particular interest. Although not significantly different for males and females, the age effect itself is statistically significant for females ($\chi^2_2 = 18.70, P < .01$) but not for males ($\chi^2_2 = 3.42, P > .2$). This is due not only to an *apparently* stronger age dependence in females, as reflected by *slightly* larger parameter estimates (table 7), but, more importantly, to a considerably larger sample size.

The quadratic age-dependence in the proportion of committed-smokers (fig. 3) could indicate that: (1) Some potential committed-smokers in younger age groups did not commence smoking sufficiently early for the status of committed-smoker to be realized. (2) In the older-age groups, there is a greater opportunity for long-term smokers to have given up. (3) There is possibly a higher mortality rate for committed-smokers in the older-age groups. (4) Individuals born about

1940 (particularly females) are more likely to have become committed-smokers because of social and economic values prevailing during their early-adult life.

These findings appear to have important implications for public health. First, they are consistent with the common-sense conclusion that the probability of becoming a smoker is influenced by social attitudes, peer-group influences, and economic circumstances prevailing in the environment during adolescence and early-adult life [8, 9, 14]. Second, after adjusting for these sex- and cohort-dependent differences, the correlations for becoming a smoker are much higher for MZ twins than for DZ twins; this finding is consistent with the notion that the decision to start smoking is influenced by genetic factors. More importantly, it seems that in the contemporary Australian environment, the state of "committed-smoking" (i.e., having decided to start and then not to have given up after at least 10 years of smoking) not only is influenced by (environmental) factors particular to the individual but also behaves as a trait under considerable genetic influence in both males and females. There is a growing body of evidence to suggest that smoking behavior is associated with the personality dimensions extraversion, neuroticism, and psychoticism [7, 15, 16]. It may be that "committed-smokers," as a group, are characterized by genetically determined differences in the rewards obtained from smoking or by genetically influenced differences in personality dimensions [17] that make them less responsive to conventional health-education programs. In either case, substantial individual variability indicates the need for further research to investigate the personal characteristics of "committed-smokers" and to seek intervention strategies that are more suited to the characteristics of individual smokers.

ACKNOWLEDGMENTS

We thank Jan Temperley and Rosemary Jardine for their valuable help, and we gratefully acknowledge the work of Dr. N. G. Martin in establishing the Twin Registry and coordinating the postal survey.

REFERENCES

1. HANNAH MC, HOPPER JL, MATHEWS JD: Twin concordance for a binary trait. I. Statistical models illustrated with data on drinking status. *Acta Genet Med Gemellol (Roma)* 32:127–138, 1983
2. KAPLAN B, ELSTON RC: *A Subroutine Package for Maximum Likelihood Estimation* (MAXLIK). Chapel Hill, Univ. of North Carolina, Institute of Statistics Mimeo Series No. 823, 1972
3. COX DR, HINKLEY DV: *Theoretical Statistics*. London, Chapman and Hall, 1973, pp 279–363
4. HANNAH MC: *Simulation of a Test Statistic for the Difference between Binary Correlation Coefficients*. Melbourne, Australia, Univ. of Melbourne, Faculty of Medicine Epidemiology Unit Technical Report No. 1, 1984
5. MARTIN NG, MARTIN PG: The inheritance of scholastic abilities in a sample of twins. I. Ascertainment of the sample and diagnosis of zygosity. *Ann Hum Genet* 39:213–218, 1975
6. BOX GEP, COX DR: An analysis of transformations. *J R Stat Soc Ser B* 26:211–252, 1964
7. EYSENCK HJ: *The Causes and Effects of Smoking*. London, Maurice Temple Smith, 1980

8. O'Connell DL, Alexander HM, Dobson AJ, et al.: Cigarette smoking and drug use in schoolchildren. II. Factors associated with smoking. *Int J Epidemiol* 10:223–231, 1981
9. Higgins MW, Kjelsberg M, Metzner H: Characteristics of smokers and nonsmokers in Tecumseh, Michigan. I. The distribution of smoking habits in persons and families and their relationship to social characteristics. *Am J Epidemiol* 86:45–59, 1967
10. Kaprio J, Hammar N, Koskenvico M, Floderus-Myrhed B, Langinvainio H, Sarna S: Cigarette smoking and alcohol use in Finland and Sweden: a cross-national twin study. *Int J Epidemiol* 11:378–386, 1983
11. Eaves LJ: Inferring the causes of human variation. *J R Stat Soc A* 140:324–355, 1977
12. Hopper JL, Mathews JD: Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet* 46:373–383, 1982
13. Smith C: Concordance in twins: methods and interpretation. *Am J Hum Genet* 26:454–466, 1974
14. Clausen JA: Adolescent antecedents of cigarette smoking: data from the Oakland growth study. *Soc Sci Med* 1:357–379, 1968
15. Cherry N, Kiernan K: Personality scores and smoking behaviour; a longitudinal study. *Br J Prev Soc Med* 30:123–131, 1976
16. Spielberger CD, Jacobs GA: Personality and smoking behavior. *J Pers Assess* 46:396–403, 1982
17. Martin NG, Eaves LJ, Fulker DW: The genetical relationship of impulsiveness and sensation seeking to Eysenck's personality dimensions. *Acta Genet Med Gemellol (Roma)* 28:197–210, 1979