Acta Genet Med Gemellol 28:173–195 (1979)
The Mendel Institute/Alan R. Liss, Inc.

Received 13 April 1979
Final 4 July 1979

A Twin Methodology for the Study of Genetic and Environmental Control of Variation in Human Smoking Behavior

David W. Crumpacker,[1,2] Rune Cederlöf,[3] Lars Friberg,[3] William J. Kimberling,[1,4] Stefan Sörensen,[3] Steven G. Vandenberg,[1,5] James S. Williams,[1,6] Gerald E. McClearn,[1,2,5] Britt Grevér,[3] Hari Iyer,[6] Margaret J. Krier,[6] Nancy L. Pedersen,[1,5] Richard A. Price,[1,5] Ingegärd Roulette[3]

[1]*Institute for Behavioral Genetics, University of Colorado, Boulder;* [2]*Department of Environmental, Population and Organismic Biology, University of Colorado, Boulder;* [3]*Department of Environmental Hygiene, The Karolinska Institute, and Department of Environmental Hygiene, Swedish National Environmental Protection Board, Stockholm;* [4]*Formerly, Department of Pediatrics, University of Colorado Medical Center, Denver; Current address: The Boys Town Institute for Communication Disorders in Children, Creighton University Medical School, Omaha;* [5]*Department of Psychology, University of Colorado, Boulder;* [6]*Department of Statistics, Colorado State University, Fort Collins*

A method is presented for partitioning the variance associated with human smoking behavior into additive genetic, nonadditive genetic, prenatal environmental, postnatal familial environmental, and postnatal extrafamilial environmental components. Estimations can also be made of additive genetic and residual correlations between spouses and of the correlation between parental additive genetic effect and progeny nonadditive genetic and environmental effect. The variance estimates are free of the biases that might result from these correlations. The statistical genetic analysis is being applied to a large group of MZ and DZ twins, their spouses, and their adult children who live in southern Sweden. Blood samples from each subject will be used to identify their genetic constitution for a number of biochemical polymorphisms, some of which may be considered a priori to have possible relationships to smoking. Associations and genetic linkages between biochemical marker loci and quantitative behavioral traits will be sought. Traits of interest include a wide array of tobacco-use variables, motives for smoking, personality and cognitive variables, and other variables associated with drug use and health. Zygosity determinations based on biochemical polymorphisms have indicated MZ to DZ and DZ to MZ misclassification rates of 0% and 6.15%, respectively, when based solely on external morphology and ques-

Supported by grant CFTR-1066 from The Council for Tobacco Research, USA, Inc.

0001-5660/79/2803-0173$04.10

tionnaire data. The nonpaternity ratio of the fathers with respect to their supposedly biological children is estimated to be 0.28%. Gene frequency estimates for 21 marker loci show that the sample of twins and their relatives is quite representative of the Swedish population at large. All loci were in Hardy-Weinberg-Castle equilibrium, with no evidence of assortative mating for biochemical traits. The MZ twins are significantly more concordant than the DZ twins with respect to whether they have ever had a smoking habit.

Key words: Twins, Genetics, Pre- and postnatal environment, Assortative mating, Smoking behavior, Psychological variables, Biochemical polymorphisms, Concordance, Zygosity, Paternity

INTRODUCTION

Smoking behavior is a complex trait with many aspects. These include the type and amount of tobacco consumed, the method of consumption (cigarettes, pipes, etc), the depth of inhalation, age and sex patterns, time trends, the psychological and social settings, and relationship to the use of drugs other than nicotine. Several interesting models have been proposed by psychologists and others in their attempts to explain why people smoke [eg: 21, 33, 38, 40, 55, 63]. Comparisons of smoking behavior in MZ and DZ twins in several countries suggest an hereditary influence [eg: 15, 17, 18, 24, 48, 58]. A comprehensive analysis of the degree to which a complex trait such as smoking is inherited might lead not only to an increased understanding of its causes but also to improvement of models and analytical techniques in human behavioral genetics. We chose to study the genetics of smoking behavior of twins and their relatives for these and for other reasons as well. For example, people are generally willing to discuss their smoking habits because, at least until very recently, there has been little stigma or controversy associated with the use of tobacco. Another factor that facilitates the study of smoking is that large amounts of population data have been accumulated by epidemiologists interested in the association of smoking with disease [for references see Cederlöf et al, 15].

This report is limited to 1) a general description of the twin methodology we have developed for the study of genetic and environmental control of variation in smoking behavior; 2) a discussion of the make-up of the sample of twins and their relatives, which will be used to provide data on smoking behavior and other variables; 3) a description of the age structure, general smoking behavior, and general representativeness of the sample; and 4) an analysis of concordance between the twins for their general smoking status. This will provide the basis for future, detailed analyses of a wide range of smoking behavior traits and related variables collected from the sample, which consists of Swedish subjects identified and located with the aid of twin registries compiled at the Department of Environmental Hygiene of the Karolinska Institute in Stockholm [13, 43].

Although twin studies provide a classical and potentially powerful means of studying genetic influences on human behavior, they can be criticized on the grounds of possible biases in parameter estimates caused by the assumptions needed for previous types of analyses. For example, an estimate of broad-sense heritability can be obtained from twice the difference between the intraclass correlation of MZ and DZ pairs [for discussion see McClearn and DeFries, 39] assuming an absence of assortative mating and identical environmental variances of MZ and DZ pairs. Narrow-sense heritability, h^2, is a significant concept in evolutionary theory because it refers to that fraction of the phenotypic variance in a trait which is most susceptible to modification by natural selection. Its more practical usefulness stems from the fact that it provides a minimal estimate of broad-sense heritability.

Regression of offspring on midparent values can be used to estimate h^2 if several assumptions are made, including that of no environmental covariance between the two types of relatives. Elston and Gottesman [20] have shown how data on twins, their sibs, and their parents can be used to obtain improved estimates of h^2, but their method depends on a series of simplifying assumptions that include no appreciable correlation between genetic and environmental effects, and no assortative mating. Nance and Corey [45] have proposed using MZ twin pairs, their spouses, and their children to obtain estimates of additive, dominance and epistatic genetic variances, and hence of h^2. Their method can be used to detect the effects of assortative mating, but it provides no means for removing this bias or that associated with certain types of cultural transmission, such as learning by a child from a parent or a sibling.

The method we have developed for our analysis of smoking behavior combines the power of MZ and DZ twin comparisons with family studies in which there are several pairs of relatives with the same genetic relationship but different environmental relationships. For example, the additive genetic relationship between an aunt and a niece is the same as that between female cousins whose mothers or fathers are MZ twins, but their environments differ. Our study is based on information obtained from MZ and DZ twins, their spouses, and their adult progeny. Adult progeny are required because the smoking behavior phenotype is not usually expressed before adolescence. Enough relationships are provided by this approach to permit estimates of narrow- and broad-sense heritabilities, additive-genetic assortative mating, a specific type of cultural transmission (effect of parental genotype on progeny environment), and a relatively detailed set of prenatal and postnatal environmental parameters. The method was designed for use with any adequate group of twins and their relatives and can be expanded to include additional types of relatives. It incorporates some features of the theoretical model of Cavalli-Sforza and Feldman [12], in which parental genotype can affect rearing environment of offspring, and the empirical model of Rao et al [50], in which allowance is made for contributions from various genetic and environmental sources and interactions.

In order to generalize from a twin study to the population that contains the twins, one must assume that the joint distribution of phenotypes and genotypes is not atypical. Since the present testing of our analytical method depends on data obtained from Swedish twins and their relatives, it is important to note that Cederlöf [13] found no large, consistent differences between Swedish twins and the general Swedish population for an array of medical and socioeconomic variables after the data had been standardized for age. Our analysis extends this type of comparison to include consideration of gene frequencies for a wide range of biochemical polymorphisms in blood and saliva.

Previous twin studies that have provided suggestive evidence for the genetic control of smoking behavior dealt with this complex trait in an elementary fashion. In most cases, the only variation considered was that of "smoking" vs "nonsmoking." A somewhat more detailed analysis was provided by Conterio and Chiarelli [17], who compared MZ and DZ concordance for smokers vs nonsmokers, regular vs irregular smokers, and inhalers vs noninhalers. Our analysis is based on a wide range of smoking variables (collectively constituting a "smoking behavior composite"), as well as on a group of psychological, physiological, and environmental variables that may be related to smoking. The smoking variables were chosen so as to provide a number of useful classifications of smokers, thereby providing a much broader concept of smoking behavior for our analysis.

Association between the qualitative variation determined by alleles of a gene controlling a polymorphism and the variation for a certain quantitative trait may provide an alternative type of evidence (linkage or pleiotropy) for genetic control of the quantitative trait [37]. This procedure can be used, then, to supplement results obtained from other methods

such as our analysis of correlations. We have incorporated this approach into our study by obtaining information on biochemical marker phenotypes (ie, blood groups, serum protein, red cell enzymes) in each Swedish subject. A search for associations between the marker phenotypes and quantitatively varying smoking behavior traits is now being conducted. Any association found will imply such things as linkage, population stratification, or pleiotropism. Hence, it will be important to follow these analyses with an analysis designed to detect linkage [47].

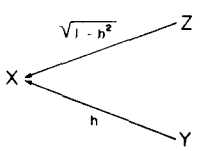
Although the "linkage" method is a potentially powerful means of demonstrating genetic control of a quantitative trait, it is very inefficient [53]. The probability of locating any particular gene or gene complex governing smoking behavior, by virtue of its linkage with a marker gene, is low. However, this probability will increase with a corresponding increase in the number of marker genes available for analysis, the number of independent aspects of smoking behavior investigated, and the number of families included [for discussion, see Elston, 19]. The probability will also depend upon the degree to which the quantitative trait is genetically controlled.

We are using two approaches which will increase the probability of detecting associations due either to linkage or pleiotropy. First, we will be especially interested to search for linkage between marker genes and those smoking traits that have been shown in advance by our analysis of correlations to have genetic components. Second, we have included in our blood survey a subset of polymorphic loci that have an a priori, theoretical possibility of affecting smoking behavior in a pleiotropic fashion because of their predisposing or exacerbating effects on respiratory diseases. The loci selected are Pi (α-1-antitrypsin), Gm (IgG), Sec (secretor status), ABO blood group, ADA (adenosine deaminase), CDA (cytosine deaminase), and UMPK (uridine monophosphate kinase). Several of these loci have been described only in single reports because the deleterious alleles rarely occur in the homozygous state [25, 26, 27, 28, 32, 41, 42, 62]. Heterozygotes for these alleles, with a less deleterious effect on respiratory function and a much higher population frequency, would provide a more likely basis for detection of an association between marker genes and smoking behavior. For example, a heterozygote with decreased lung function due to pulmonary emphysema may be unable to maintain a smoking habit because of the increased personal discomfort which it causes. In fact, α-1-antitrypsin has been reported in several studies to be associated with pulmonary emphysema [eg: 1, 8, 34, 35], and recent studies have shown a tendency towards decreased lung function and decreased α-1-antitrypsin levels in heterozygotes for deleterious alleles of the Pi locus [61].

METHOD OF ANALYSIS: I. THE THEORETICAL MODEL

All quantitative theories that allow separation of genetic and environmental effects are based on analyses of covariances and hence must be mathematically equivalent to path analyses. We use the latter for the presentation of our model because of the facility they afford for displaying relationships through pedigree. We will present here only an outline of the model and its associated analysis. The model assumes autosomal, diploid inheritance with linkage equilibrium. Details will be provided in a separate publication.

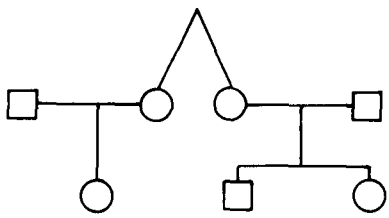
Emphasis is placed on estimation of narrow-sense heritability, which permits use of the following basic model:



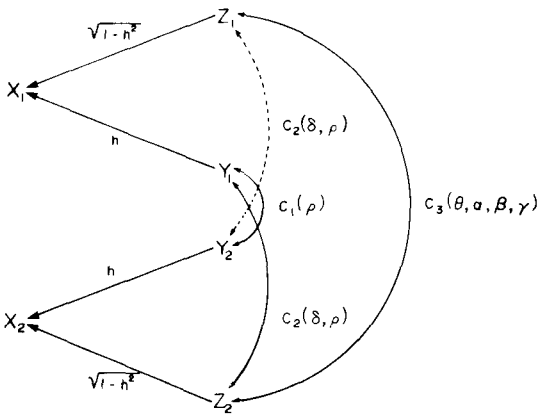
where X, Y, Z are phenotypic, additive genetic, and residual scores, respectively; h and $\sqrt{(1-h^2)}$ are additive genetic and residual regression coefficients of X on Y and Z, respectively; and

$$h^2 = \frac{\text{additive genetic variance}}{\text{phenotypic variance}} = \text{narrow-sense heritability coefficient.}$$

We define a twin unit as a pair of like-sex twins, their spouses, and one or more adult progeny (⩾20 years of age) *in at least one* of the two twin families, eg:



In this unit the twins are female and there are seven people. Most of the twin units in the sample we obtained had at least one child in each family. There are four types of twin units: MZ male, MZ female, DZ male, and DZ female. One can distinguish 33 pairwise relationships from within these units when sex of an individual and zygosity type of a twin pair are used as distinguishing characters. Examples of consanguineous relationships are a pair of male cousins, a pair of female cousins, and a mixed pair of male and female cousins, related in each case through DZ twin mothers. Examples of nonconsanguineous relatives are a pair of husbands of MZ twin sisters and a pair of wives of MZ twin brothers. There are fewer within-pair correlations when phenotypic values are adjusted within twin units to eliminate effects of sex and generation. For example, one correlation value is then appropriate for all three of the above-listed pairs of cousins related through DZ twin mothers, and a single correlation value applies to the two types of relationships involving spouses of MZ twin pairs. There are 9 and 13 meaningful pairings, respectively, for nonconsanguineous and consanguineous relatives. In theory, their correlation coefficients can differ substantially because of genetic and environmental differences among relatives. We have modeled the 22 correlation coefficients with the functions of 14 unknown, but interpretable parameters. A generalized path diagram in which these parameters and functions are displayed for a pair of relatives is shown below:



https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

The generalized phenotypic correlation, r_{12}, between the pair is:

$$r_{12} = h^2 c_1(\rho) + h\sqrt{1-h^2}\ 2^{\upsilon_{12}} c_2(\delta,\rho) + (1-h^2) c_3(\theta,\alpha,\beta,\gamma)$$

The definition of additive genetic score is such that (Y_1,Z_1) and (Y_2,Z_2) are uncorrelated within pairs. The three c functions are correlations that depend on the degree of biological or cultural relationships within a pair. The parameters on which these functions depend have the following properties:

1. Non-zero ρ is a consequence of phenotypic assortative mating or inbreeding due to isolation by distance on additive genetic scores.
2. Non-zero δ is a consequence of genetically influenced cultural transmission.
3. Non-zero θ is a consequence of assortative mating or phenotypic convergence (change in phenotypic similarity of spouses following marriage) on residual scores derived from nonadditive genetic and environmental effects.
4. Non-zero α is a consequence of prenatal environmental effects; non-zero β is a consequence of postnatal environmental effects; and non-zero γ is a consequence of nonadditive genetic effects.

There are three levels of α, four of β, and three of γ, which lead to ten parameters (α_1, $\alpha_2, \ldots, \gamma_3$). Levels 1 through 4 refer to levels of increasing relatedness in pedigree and rearing environment. The parameter υ, which is known, is an indicator of whether the relatives are in the same or different generations. The dashed line between Z_1 and Y_2 means that this path may or may not occur if the relatives are in different generations. The correlations provide 22 distinct equations that involve 14 parameters. This leaves ample degrees of freedom to test the fit of the model, as well as the validity of the rules by which the functions of ρ, θ, α, β, and γ are derived. In case of a poor fit to the data, up to eight parameters can be added to improve the model. It is also possible to simplify the model by deleting superfluous parameters, thereby improving the estimates of those which remain.

A weighted least-squares procedure based on Z transformations of sample phenotypic correlations can be used to obtain estimates of the genetic, mating, and environmental parameters. In large samples this is an approximately maximum-likelihood method, which will account fully for unequal precision of the various sample correlations.

We originally estimated that 30 of each of the four types of twin units would provide enough information to obtain satisfactory coefficients of variation for the estimates of the Z transformations of the 22 distinct phenotypic correlation coefficients in our sample. This would provide at least 112 degrees of freedom for each such estimate. The better of these estimates, such as spouse with progeny or twin with progeny, are actually based on 129 or more degrees of freedom, since our realized sample of complete twin units of each type exceeded our original goal of 30 each (see Table 9). The coefficients of variation for the better estimates will range from 15% for those with a 0.05 correlation coefficient down to 1% or less for those with a correlation coefficient of 0.50 or more. The coefficient of variation is below 10% for all correlation coefficients greater than, or equal to, 0.08. Some of the Z-transformed sample correlation coefficients from less important, widely related pairs of relatives will be based on many fewer than 129 degrees of freedom. In the extreme

https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

cases, such as the relationship of first cousins who are children of MZ twin mothers, the coefficients of variation will range from 59% (0.05 correlation coefficient) down to 12% (0.25 correlation coefficient).

A preliminary survey of our data indicates that the coefficients of variation for estimates of the model parameters such as h^2 and products of $(1 - h^2)$ with α, β, and γ, range from 30% to 50%. With values like these, the power of likelihood tests for non-zero values of these parameters will range from 0.64 to 0.95. The ρ, θ, and δ parameters will be estimated less precisely and, therefore, be more difficult to demonstrate as being non-zero.

A special problem associated with the statistical analysis concerns the generation differences to be expected with traits as labile as those associated with smoking. At least three possibilities exist for handling this problem. They are 1) to enlarge the model to include generation-specific parameters, 2) to model the effects of intergeneration environmental changes on the 14 basic parameters, and 3) to use numerical indicators and descriptions of the environment (eg, socioeconomic status) in each generation to adjust the smoking behavior phenotypes prior to the path analysis. Generation-specific parameters are not feasible for the present Swedish study because the total number of parameters would have to be doubled, leading to the need for at least a doubled sample size. Modeling the effects of generation change with not more than seven parameters might be feasible, providing that the new set of "change" or transition parameters could be estimated from our data and other sources of Swedish population information. Adjustment of the data prior to genetic analysis is feasible if the generation changes are not too large. This can be done first by calculation of all sample correlations within subsets of the data indexed by sex, generation, and family type, and then pooling estimates of correlation with the same functional form and values over the subsets. More refined adjustment for the environmental effects specific to an individual subject can then be calculated from environmental indices, such as peer pressure to use drugs during teenage years (discussed in section on Ancillary Variables), which have been estimated for each subject.

METHOD OF ANALYSIS: II. ASSOCIATION WITH BIOCHEMICAL MARKER GENES

We plan to test for 57 genetically determined biochemical polymorphisms in our Swedish subjects. Each of these is controlled by a single locus with two or more alleles. The polymorphisms are distributed among tissues as follows: red cells, 29; white cells, 6; serum, 14; and whole saliva, 8. The analysis seeks to detect associations between the marker locus phenotypes and smoking behavior traits, and then to determine if the associations are caused by linkage. Association between a pair of traits occurs when their phenotypes are not randomly distributed with respect to one another in a population of genetically independent individuals (eg, a population consisting of one member from each twin pair, or a population consisting of one spouse from each of a group of twin pairs). One-way analyses of variance will be used to compare the quantitative variation of a behavioral trait between and within marker locus phenotypes in our Swedish sample. A multiple linear regression model will also be used, with each quantitative trait as the dependent variable and the phenotypes of each marker locus as the independent variables. If significant associations are found, the relevant variables will then be subjected to linkage analysis using Ott's computer program [47] modified for quantitative variables. This method consists of a likelihood ratio test that compares the probabilities of cosegregation of two traits, given the hypotheses of linkage and no linkage. It will be applied to the twin family unit, since detection of linkage requires the use of family data to demonstrate the cosegregation of two traits during their transmission from parent to progeny.

THE SMOKING BEHAVIOR COMPOSITE

A Swedish subject was classified as a "current smoker" if he or she met any of the following criteria:

1) Has smoked at least 100 cigarettes, 60 cigar-cigarettes (cigarillos), 20 cigars, or a 50-gram pack of pipe tobacco in a month or less; and has smoked at least six cigarettes, six cigar-cigarettes, two cigars, or two pipefuls in the past month.
2) Currently smokes more or less regularly but less than in (1).
3) Currently smokes "now and then," for example at parties.

Class (1) subjects are, then, "regular" current smokers, whereas class (2) and (3) subjects are "very light" current smokers.

A subject was classified as an "ex-smoker" if he or she met either of the following criteria:

1) Fulfilled criterion (1) above but had not smoked more than six cigarettes, six cigar-cigarettes, two cigars, or two pipefuls in the past month.
2) Former very light smoker who no longer smokes.

A subject was classified as a "nonsmoker" if he or she met none of the criteria for current or ex-smoker; ie, if he or she had never smoked at all, or at most a few trial cigarettes, cigar-cigarettes, cigars, or pipefuls.

The remainder of the smoking behavior composite was determined primarily by two sets of questionnaires, one which obtained information on aspects of tobacco use and another which attempted to determine reasons or motives for smoking. The tobacco-use questionnaire consisted of similar subsets of questions on current and lifetime use of cigarettes, cigars (and cigar-cigarettes), and pipes, and dealt with such items as types and brands of tobacco products, amount smoked per day, number of puffs per unit, degree to which each unit was smoked, depth of inhalation, age at which smoking started, number of years of smoking, number of attempts to stop smoking, and self-assessment of smoking type (heavy, moderate, etc). A few questions dealt with the use of snuff and chewing tobacco. Current smokers completed the current and lifetime parts of the questionnaires, whereas ex-smokers completed only the latter.

The main "smoking motives" questionnaire, completed by current smokers and ex-smokers, was based on the work of Russell and his colleagues in England [55 and Russell, personal communication]. Our 44-item version contained most of the original 34 items used by Russell et al. Multivariate analysis of the Swedish responses to our questionnaire showed a factor structure remarkably like that obtained previously with similar questionnaires in London, in Denver, Colorado, and Fort Collins, Colorado (Williams, Crumpacker, and Krier, unpublished). The same basic factor structure was maintained in different sexes, age groups, and classes of smokers (current and ex), as well as with response scales differing in length and degree of symmetry. Six first-order factors were identified by Russell et al as smoking motives: psychosocial, indulgent, sensorimotor, stimulation, addictive, and automatic. These intercorrelated in the London sample to produce two second-order factor motives denoted by Russell et al as pharmacological (the last three factors just listed) and nonpharmacological (the first three). We found that a three-factor description, viz pharmacological, nonpharmacological, indulgent, fit the Swedish data better.

We used two additional motivational questionnaires in the Swedish study, which we designed to assess motives for "starting to smoke" and "never starting to smoke." Current smokers and ex-smokers completed the former whereas nonsmokers completed the latter. Analysis of the Swedish responses agreed with an earlier analysis of Denver data (unpublished) in providing evidence for factor motives, which we named as follows: starting to smoke – independence, boredom, and social acceptability; never starting to smoke – attitudes of friends and family, and health.

Shortly after testing began in Sweden, we realized that the use of snuff was considerably more extensive than had been assumed. Therefore, we devised motives questionnaires for "snuff habits" (36 items) and "starting to use snuff" (11 items), both patterned after the smoking questionnaires. The factor structures of the Swedish responses to these questionnaires have not yet been determined.

ANCILLARY VARIABLES

We collected information on a number of ancillary variables of several types, primarily personality, cognitive, medical, and environmental. This was done for the following reasons: 1) to construct indices for prediction of the degree to which a person is predisposed towards various types of smoking behavior; 2) to characterize in as much detail as possible the nature of environmental changes between generations; 3) to improve our understanding of the environmental and residual parameters estimated by analysis of correlations (α's, β's, δ, and θ); and 4) to obtain preliminary estimates of genetic correlations between traits.

Personality was assessed in Sweden by means of Comrey, Eysenck, Locus of Control, Concern for Health, and Teenage questionnaires. Our Comrey questionnaire was based on four of the eight original Comrey Personality Scales [16]. The scales we chose (Orderliness, Emotional Stability, Extraversion, and Activity) were related to personality dimensions which, on the basis of earlier studies [64], were suspected of having some genetic component. Our Eysenck questionnaire included two (Neuroticism and Extraversion) of the four original Eysenck Personality Scales [22]. The original Locus of Control Scale [54] attempted to assess how much control a person believes he or she has over the events that affect his or her life. We devised a group of items, based on Rotter's questionnaire, that we believed to be most representative of this trait. The Concern for Health questionnaire consisted of items that we devised, because we could find no suitable published scale designed to measure this trait. We developed the Teenage questionnaire as a result of a pilot study which showed that 72% of the cigarette smokers in a Denver suburb began smoking between the ages of 13 and 19. The questionnaire retrospectively assesses the personality dimensions of neuroticism, rebelliousness, compulsiveness, and impressionability.

Multivariate analyses of the Swedish data (unpublished) produced very satisfactory replications of the dimensions just described for the various personality assessments in both the twin and progeny generations. Several of the factors showed correlations between their scale scores as expected; eg, the Comrey Emotional Stability scale correlated −0.69 with Eysenck neuroticism and −0.41 with Teenage neuroticism, and the Comrey extraversion scale correlated 0.76 with Eysenck extraversion.

We included only one measure of cognition, a shortened version of the Raven Progressive Matrices Test [51]. The Raven test is generally regarded as a reasonably culture-fair measure of nonverbal reasoning. It was included because of earlier reports of a negative correlation between cognitive abilities and smoking [65], and also because the Swedish

study provided an exceptional opportunity to investigate the importance of genetic and environmental effects on a general measure of cognitive ability. Multivariate analysis of the Swedish responses to the Raven test clearly replicated earlier results by identifying a major factor that has been referred to by previous workers as a general ability factor.

The Swedish medical or "Health" questionnaire included a series of questions designed to provide information on symptoms and medical histories for the diseases most commonly associated with smoking (cardiovascular and respiratory diseases), and also data on frequency of medical examinations, dyslexia, birth order of twins, age, height, weight, optimal time of functioning during the day, handedness, and use of drugs other than tobacco (alcohol, coffee, tea, tranquilizers, and sleeping pills). Alcohol-use questions were designed to provide quantity-frequency estimates of alcohol consumption; they were based on three similar subsets of questions that pertained to beer, wine, and liquor. Each subject was also asked to provide the following information for all first-degree relatives: relationship, date of birth, date and cause of death (if applicable), a brief medical history for cardiovascular and respiratory diseases, smoking habits, and alcohol use. Wives in the twin generation were asked a detailed set of questions about the prenatal, perinatal, and early postnatal aspects of each pregnancy. This provided information on pregnancy outcome, delivery, birthweight and sex of the infant, congenital problems of the infant, and use of various drugs throughout pregnancy. During the period of blood sampling, each subject was questioned about cardiac infarct, angina pectoris, medicines taken during the past 24 hours, and, if female, whether or not she was pregnant or taking oral contraceptives.

The Environmental questionnaire given to each subject requested information on education; length of work week; physical and mental stress, and fatigue associated with work; work rules pertaining to smoking; exercise during leisure time; irritation due to pollutants at place of residence; time spent in presence of smokers; parental occupations, place, type, size, and crowdedness of residence, and standard of living during the subject's teenage years.

Each pair of spouses in the twin generation, as well as progeny living outside their parents' homes, was asked to complete a Family questionnaire during the test session. This provided current information on type and size of residence; number of persons living in the residence; ownerships of cars, television sets, boats, pets, and summer houses; vacation and business trips abroad; occupation of family wage earners and education or training required for the occupations.

The Teenage questionnaire described earlier with respect to psychological variables also included items that attempted to determine the nature of the surrounding environment during each subject's teenage years and, in particular, its influence on the development of smoking and drinking habits. We devised the environmental items, modeling them to some extent on the work of Schaefer [56, 57]. Separate multivariate analyses were conducted on the Swedish twin and progeny generations. Three factors, which we call pressure to use tobacco and alcohol, unpleasant home environment, and financial status, were found to be quite similar in both generations. However, the assessment of parental child-rearing attitudes by the two groups differed somewhat. The following factors were identified: twin generation – parental encouragement of work and responsibility, and nurturance-encouragement of autonomy; progeny generation – parental encouragement of achievement, parental discipline, and nurturance-cohesion. The most stable factor across generations was pressure to use tobacco and alcohol. Essentially the same factor was also identified in a preliminary conducted on a group of students at the University of Colorado, Boulder.

THE DENVER PILOT STUDY

We conducted a nongenetic pilot study in the metropolitan area of Denver in autumn 1976 in order 1) to learn as much as possible about the smoking behavior composite, 2) to investigate the relationship of smoking behavior to a number of ancillary traits that were suspected of having some effect on its development and maintenance, 3) to obtain practical experience in the logistics of obtaining a complex set of data from numerous subjects, 4) to test and refine the various questionnaires we intended to use in the main Swedish study, and 5) to obtain data that could be used in the development of our information-management system.

The goal of the pilot study was to collect data on 200 randomly sampled subjects from 20 to 65 years of age, of predominantly Caucasian ancestry and living in an urban or suburban area, in order to provide individuals as similar as possible to those in the main Swedish study.

The sample was selected by a stratified random sampling procedure in suburban Jefferson County, Colorado, using 1970 census data and a 1975 city directory, which referenced households geographically. A total of 226 index subjects randomly chosen at the rate of one per city block and 151 additional subjects (mostly spouses) were tested. When compared with the 1970 census statistics for suburban Jefferson County, the index sample contained higher proportions of persons who were over 35 years of age, high school graduates, and home owners. Home owners were easier to locate than renters (who move more often), and were probably older than renters. These factors could explain some of the sample differences. An analysis of 461 persons who declined to participate indicated that nonsmokers were somewhat underrepresented and ex-smokers somewhat overrepresented in the pilot study sample.

The pilot study data were collected from the battery of questionnaires discussed earlier, that was completed by the subjects under the supervision of project personnel. A typical session averaged 12 subjects and three supervisors. Most subjects were able to complete the test booklet within 2½ hours. Each subject signed an informed consent document before the session and was paid $10 (U.S.) at the end of the session.

THE SWEDISH TWIN STUDY: I. THE SAMPLE

Two information sources compiled and maintained by the Department of Environmental Hygiene at the Karolinska Institute were used to choose the sample of twin units. These are the "old" and "new" Swedish Twin Registers which contain data, respectively, on the following approximate numbers of like-sex twin pairs: 11,000 born from 1886 to 1925 and 13,000 from 1926 to 1958 [13, 43]. Both registers were surveyed in 1975 to determine the number of twin pairs that would be potential candidates for the study. The survey was limited to pairs living in the Stockholm, Göteborg, and Malmö metropolitan areas in order to reduce the subsequent costs of data collection. This "target" population may be described as a largely urban and suburban population of southern and central Sweden. The survey was restricted further to twin pairs born from 1911 to 1935, both of whom had spouses and at least one child of the marriage. The birth range was chosen to assure that most twins in 1976 would not be older than 60 years and would have at least one child 20 or more years of age. Even though the correlation analysis required only one child per twin unit, we surveyed for at least one child in each of the two families of each unit because the additional progeny would considerably enhance the data base. A small feasibility

study was also performed, which provided reasonable assurance that we could obtain our desired minimum sample of 30 twin units for each of the four types. Letters describing the study and the criteria for participation were then sent in June 1977 to all potentially useful twins and their spouses and, eventually, to their children. Once it was realized that we could obtain our minimum sample from the Stockholm and Göteborg areas, we decided not to use Malmö subjects.

The sample on which data were collected is described in Table 1. The minimum goal of 30 twin units was exceeded for each of the four types. Also important is the fact that 102 of the 138 units contained at least one child of the desired age in both families. The average number of children per unit was 1.9. The total sample of 909 subjects consisted of 548 from the twin generation (twins and their spouses) and 361 from the progeny generation (adult children of twins). One individual in the progeny generation was later dropped from the sample when blood group analyses indicated that his "father" was very likely not his biological parent. The numbers of twin units in Stockholm, Göteborg, or divided between the two localities were, respectively, 90, 43, and 5. There were 610 subjects tested in Stockholm and 299 tested in Göteborg.

During the sampling procedure, prospective twin subjects were classified as (1) acceptable or nonacceptable with respect to our criteria of having adult children by current spouse and being currently married; (2) declining to participate; or (3) failing to respond to our letters soliciting information about family characteristics and willingness to participate. Twins in the latter two categories may be referred to as "not-interested." The "not-interested" rate was approximately 35% for twin pairs (ie, cases where either one or both twins of a pair were not interested) and 30% for individuals. Using information already in the Swedish Twin Registries, we compared concordance rates for the "acceptable," "nonacceptable," and "not-interested" pairs of twins. The comparisons used variables concerned with demography, personality, medical history, education, employment, and use of various drugs, including tobacco. The results have given no indication of different concordance rates among the groups for these variables. Further analyses to determine whether the sample tested in this study was different from cohort twins and the registry as a whole are under way.

THE SWEDISH TWIN STUDY: II. TESTING PROCEDURES

On the basis of results obtained in the Denver pilot study, the basic questionnaire battery was refined, shortened, and translated from English into Swedish. A few questionnaires, such as the Teenage, were added. Perfection of the translation involved several rounds of forward and back translation, including a "blind" back translation into English by a professional linguist fluent in both languages. Special care was taken to preserve the meaning of the English idioms that occurred in such places as the personality and smoking motives questionnaires. A few items that could not be adequately translated or had no real applicability to the Swedish culture were deleted.

Almost all of the test sessions in Stockholm and Göteborg were held in autumn 1977. A standard protocol was used to minimize procedural differences between the testing locations. Each session required from 1½ to 2½ hours for completion, and averaged 14 and 10 subjects, respectively, in the two locations. In order to minimize spurious within-family correlations, relatives were not seated together.

In addition to completing the questionnaire battery, each subject was asked to provide 60 ml of blood (70 ml for subjects belonging to MZ twin units) and 10 ml of whole saliva, to lie down for a resting blood pressure test, and to sit for a facial photograph. Blood sam-

TABLE 1. Distribution of Twin Units in the Swedish Sample†*

	Number of twin units				
	MZ male	MZ female	DZ male	DZ female	Total
With at least one adult child in both families	25	30	18	29	102
With at least one adult child in one family and none in the other family	14	6	14	2[a]	36
With at least one adult child in at least one of the two families (sum of first two distributions)	39	36	32	31	138

*A twin unit is a pair of like-sex twins, their spouses, and one or more adult progeny (ie, children ⩾20 years of age) in at least one of the two families.
†A sample of twins from the Swedish Twin Registry.
[a]One of these units consisted only of one twin, her spouse, and her child.

ples and pressure were obtained by registered nurses. The photographs are being used in analysis of assortative mating for physical characteristics and in a study of the relationship between physical characteristics and smoking behavior. Each subject was paid the equivalent of $22 (U.S.) in Swedish crowns for the blood sample.

Each blood sample included 20 ml of clotted blood, 20 ml of ACD blood, and 20 ml of heparinized blood. An additional 10 ml of heparinized blood was taken from MZ twin-unit subjects for drug-binding studies not directly related to the Swedish twin study. All samples were stored at 4°C for no more than 24 hours before processing. Serum and plasma were separated, the red cells glycerolyzed [6], and all samples stored in 0.3 and 2.0 ml aliquots. White cells were prepared by the ficoll/hypaque method [7]. All samples were processed in Stockholm or Göteborg and stored there at −40°C until shipment to our biochemical laboratory at the University of Colorado Medical Center in Denver. Samples were shipped every two weeks. Time in transit did not exceed 72 hours, and all but one shipment were received frozen. Each subject's sample was divided into two parts and shipped separately to avoid the total loss of any individual's sample.

ZYGOSITY OF SWEDISH TWINS AND PATERNITY OF THEIR PROGENY

Although the striking physical and behavioral similarity of MZ twins may permit a reasonably clear-cut separation of MZ and DZ pairs, use of these criteria could result in as much as 10% misclassification [14, 29, 46]. It is essential to minimize this problem in our study because each misclassification invalidates the correlation equations involving an average of 6.5 relatives. Fortunately, misclassification can be greatly reduced by comparison of blood groups, serum proteins, and red cell enzymes in the twin pairs. If the a priori diagnosis of a twin pair was DZ, we accepted discordance for at least one marker locus as proof of dizygosity. If the a priori diagnosis was MZ, then we required more than one discordant locus to change the diagnosis from MZ to DZ. To accept an a priori diagnosis of MZ, we required a probability of dizygosity less than 0.05. To change an a priori diagnosis of DZ to MZ, we required a probability of dizygosity less than 0.01.

The probability of dizygosity, P_{DZ}, was calculated in the usual way. For example, with a two-allele, codominant system, the relative chance (or odds) that two twins are DZ, given that both are A_1A_1, is $\frac{1}{4}(1 + p)^2$, where p is the frequency of the A_1 allele. Formulas and

https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

tables giving odds for several common genetic markers have been published by Smith and Penrose [60], using English gene frequencies. Table 2 shows similar odds of dizygosity calculated for the Swedish sample and includes marker systems for which odds have not been previously calculated. Odds for the heterozygotes of dominant systems are also shown. In many instances, the exact parental genotype could be derived from the phenotypes of the children of the twin pair (eg, if both twins are type A_1 for the ABO system and one of the children from each is type O, then the genotype of both twins is A_1O). If C_i is the relative chance of dizygosity at the ith locus tested, then the total relative chance of dizygosity is $C_{DZ} = (2.3333)\,(½)\,(\prod_{i=1}^{n} C_i)$, where 2.3333 is the prior relative odds of dizygosity [49], (½) accounts for the likeness in sex, and the product of C_i's is taken over n concordant loci. The probability of dizygosity is, therefore, $P_{DZ} = C_{DZ}/(1 + C_{DZ})$, and the probability of monozygosity is $P_{MZ} = 1 - P_{DZ}$.

The results of zygosity determinations for 136 of the 137 twin pairs in the Swedish sample are shown in Table 3. None of the 71 originally diagnosed as MZ was found to be DZ, and four of the 65 originally diagnosed as DZ were found to be MZ at a probability level of $\leqslant 0.01$. This is a misclassification rate of 0% and 6.15% for MZ and DZ pairs, re-

*TABLE 2. Odds (Relative Chance) of Dizygosity for Biochemical Marker Phenotypes/Genotypes**

Locus	Type	Odds	Locus	Type	Odds	Locus	Type	Odds
ABO:	A_1	0.6586	MNS:	MSMS	0.3854	Kell:	k^bk^b	0.4880
	A_1A_2	0.3350		MSMs	0.4108		K^bk^b	0.5113
	A_1O	0.4930		MSNS	0.3331		k^ak^b	0.4880
	A_2	0.4703		MSNs	0.4228	Fy:	aa	0.5072
	A_2O	0.4282		MsMs	0.4378		ab	0.6222
	B	0.4822		MsNS	0.3550		bb	0.6204
	BO	0.4332		MsNs	0.4506	Jk:	aa	0.4764
	A_1B	0.3389		NSNS	0.2683		ab	0.6178
	A_2B	0.2944		NSNs	0.3654		bb	0.6557
	O	0.6300		NsNs	0.4638	Lu:	aa	0.2680
P:	P_1	0.8509		MNS	0.4455		ab	0.5171
	P_1P_2	0.6250	AcP_1:	AA	0.4801		bb	0.9649
	P_2P_2	0.5671		AB	0.5417	Gm:	a	0.3842
Rh:	dccee	0.4780		BB	0.6113		ax	0.4049
	dccEe	0.3478		AC	0.3639		ab	0.5063
	dccEE	0.2525		BC	0.4107		axb	0.4603
	dCcee	0.3473		CC	0.2759		bb	0.6673
	dCCee	0.2519	6PGD:	11	0.9777	Km:	a	0.5268
	dCcEe	0.2522		12	0.5109		ab	0.4341
	DccEE	0.3482		22	0.2613		bb	0.9277
	DccEe	0.4252	AK:	11	0.9618	Hp:	11	0.4647
	DCCee	0.5057		12	0.5186		12	0.6157
	DCcee	0.5320		22	0.2697		22	0.6696
	DCcEe	0.4386	EsD:	11	0.8891	Gc:	11	0.7591
	DCcEE	0.2955		12	0.5506		12	0.5956
	DCCEe	0.3561		22	0.3106		22	0.3953
	DCCEE	0.2506	PGM_1:	11	0.8246	ADA:	11	0.9384
	Dccee	0.3478		12	0.5750		12	0.5293
				22	0.3504		22	0.2822

*Based on gene frequencies calculated from the Swedish sample tested in the present study.

https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

*TABLE 3. Zygosity Determinations of Swedish Twin Pairs**

Present classification[a]		Original classification[b] MZ	DZ	Total
MZ	P > 0.99	67	4	75
	P > 0.95	4	0	
DZ	⩾ 2 loci	0	57	61
	1 locus	0	3+(1)	
		71	65	136[c]

Rate of MZ to DZ error 0%
Rate of DZ to MZ error 6.15%
Total 2.94%

*A sample of twins from the Swedish Twin Registry.
[a]Diagnosis based on analysis of 20 marker loci.
[b]A priori diagnosis using questionnaire data from the Swedish Twin Registry.
[c]Of the 137 complete twin units, 136 were tested for zygosity. One pair, in which one member had only a finger prick of blood, was classified by twin registry questionnaire data as being DZ.

spectively, when based solely on external morphology and questionnaire data. These data imply correct "questionnaire" classification rates of 100% and 94% for MZ and DZ pairs, respectively, which are probably satisfactory for a number of epidemiological purposes. However, the DZ value is not adequate for the correlation analysis of partially expressed traits – eg, smoking behavior – which is very sensitive to changes in zygosity.

The odds of paternity for a particular locus were obtained from the ratio of two probabilities – viz, the probability that the alleged father of a child is the true biological father, given his phenotype and the phenotypes of the mother-child pair – versus the probability that the child resulted from any random male Swedish gamete. The product of the odds of paternity for each locus taken over all loci is the odds of paternity. This is easily converted into a probability of paternity that will vary from family to family and from child to child within each family.

We have been able to demonstrate a probability of paternity over 0.90 in 341 of the 351 children (98%) tested, with the analysis of 20 marker loci completed. Only one child has definitely been shown to be nonpaternal, giving a rate of nonpaternity of 0.28%. Apparent nonpaternity for a second child has been demonstrated only for the PGM_1 locus. The parental phenotypes are 22 (father) and 21 (mother), while the child's phenotype is 11. Since no other systems indicate nonpaternity for this child, the family will be investigated for the presence of a null allele.

CHARACTERISTICS OF THE SAMPLE OF TWINS AND THEIR RELATIVES DERIVED FROM THE SWEDISH TWIN REGISTRY

We have described a twin methodology for the comprehensive study of genetic and environmental influences on human behavior. Theoretical concepts and analytical methods have been related to a "real-life" situation – viz, the complex trait of human smoking behavior. The composite phenotype of this trait has been described in considerable detail involving

many components, and a wide range of potentially important environmental effects has been identified. The methodological background has, therefore, been provided for future reports of extensive genetic and environmental analyses that are being conducted. The material presented in this section completes the background information. It includes data on age structure, certain aspects of smoking behavior, and the qualitative (marker gene) population genetics of our Swedish sample of twins and their relatives. Twin pair concordances for two classifications of smoking status (current smokers vs ex-smokers vs nonsmokers, and "never smoked" vs "ever smoked") are also presented as preliminary evidence that is consistent with at least some genetic control of variation in smoking behavior.

Table 4 gives the age structure of the sample. The modal ages of the twin and progeny generations are 57 and 22, respectively. The proportions of current smokers, ex-smokers, and nonsmokers are listed in Table 5. A large increase in the proportion of current smokers and a concomitant decrease in proportion of nonsmokers is seen in women of the progeny generation as compared with those of the twin generation. This difference in female smokers will have to be accounted for in subsequent analyses of the genetic and environmental control of smoking status in order to avoid serious parameter bias. Data for current usage of tobacco products, which differed little from the data for lifetime usage, are presented in Table 6. Approximately 8% to 12% of men currently smoke cigars and pipes in addition to, or in place of, cigarettes, with pipes being the main alternative. The comparable value for men in the Denver pilot study was 13%, with cigars rather than pipes being the main alternative. Only 9% of the Swedish sample currently smoke more than 20 cigarettes per day (Table 7), as compared with 47% in the Denver sample. We do not know at present how much of this large difference is due, at least in part, to factors such as differences in price or in nicotine and tar content of tobacco products in Sweden and the United States.

Gene frequency estimates for the 21 marker loci that have been analyzed so far in the Swedish sample are presented in Table 8. Only genes from biologically unrelated subjects were included, and the effective population size on which the estimates were based was between 400 and 450 persons. Comparisons are made to earlier Swedish estimates reported by other investigators. The data show that the present sample is quite representative of the Swedish population with regard to biochemical marker genes. All loci were found to be in Hardy-Weinberg-Castle equilibrium, with no evidence of assortative mating. Hence, these data provide no indication of any stratification or subdivision in the Swedish sample that would bias the parameter estimates to be obtained by correlation analysis.

A comparison of concordance-discordance frequencies for smoking status of twin pairs in the Swedish sample, in terms of whether they are current smokers, ex-smokers, or nonsmokers, is given in Table 9. MZ female twin pairs are significantly more concordant than DZ female twin pairs. The difference between the two types of male twin pairs is in the same direction, but not statistically significant. When combined over sexes, MZ twin pairs are seen to be significantly more concordant than DZ pairs. Table 10 presents the same types of comparisons for smoking status in terms of whether the twins have ever smoked or never smoked. The ever-smoked class represents the pooled classes of current smokers and ex-smokers from the analysis of Table 9. The comparative differences are very similar to those in Table 9, with MZ concordances being uniformly greater than DZ concordances. The differences are again significant for females and the combination over sexes. The concordance values in Tables 9 and 10 were tested for agreement with those based on a slightly younger group of all twins from the Swedish registers who were born between 1926 and 1935. No significant differences were found in the two types of concordances for any of

TABLE 4. Age Structure of the Swedish Sample (Tabular values in percent)*

Age (years)	Twin–male[a]	Twin–female[b]	Progeny–male[c]	Progeny–female[d]	Twin generation (combined over sexes)	Progeny generation (combined over sexes)	Overall (combined over sexes and generations)
⩽19	...	...	...	0.6	...	0.3	0.1
20–24	...	...	38.0	39.2	...	38.8	15.4
25–29	...	...	33.5	28.7	...	31.0	12.3
30–34	...	...	20.1	22.1	...	21.1	8.4
35–39	...	0.4	7.8	7.7	0.2	7.8	3.2
40–44	1.5	3.3	...	1.7	2.4	0.8	1.8
45–49	13.5	18.2	...	...	15.9	...	9.6
50–54	20.1	25.5	0.6	...	22.8	0.3	13.9
55–59	33.6	29.2	...	...	31.4	...	18.9
60–64	19.7	17.5	...	...	18.6	...	11.2
⩾65	11.7	5.8	...	...	8.8	...	5.3
Number of subjects	274	274	179	181	548	360	908

*A sample of twins from the Swedish twin registry, their spouses, and their progeny.
[a]Male subjects of the twin generation (twins and their spouses).
[b]Female subjects of the twin generation (twins and their spouses).
[c]Male subjects of the progeny generation (children).
[d]Female subjects of the progeny generation (children).

*TABLE 5. Proportions (%) of Current Smokers, Ex-Smokers and Nonsmokers in the Swedish Sample**

	Sex by generation				Overall (combined over sexes and generations)
	Twin–male	Progeny–male	Twin–female	Progeny–female	
Current smokers	33.6	43.0	29.9	47.0	37.0
Ex-smokers	33.6	22.9	14.2	17.1	22.4
Nonsmokers	32.8	34.1	55.8	35.9	40.6
Number of subjects	274	179	274	181	908

*A sample of twins from the Swedish Twin Registry, their spouses, and their progeny.

*TABLE 6. Current Usage of Tobacco Products in the Swedish Sample**

		Tobacco products (Tabular values in percent)						
Sex by generation	None	Pipe only	Cigars only	Cigars + pipe	Cigarettes only	Cigarettes + pipe	Cigarettes + cigars	Number of subjects
Twin–male	66.4	3.6	1.8	0.7	21.5	4.7	1.1	274
Progeny–male	57.5	2.2	1.1	...	34.6	2.8	1.7	179
Twin–female	69.7	...	...	...	30.3	...	...	274
Progeny–female	53.0	...	...	...	45.8	0.6	0.6	181
Number of subjects	572	14	7	2	287	19	7	908

*A sample of twins from the Swedish Twin Registry, their spouses, and their progeny.

*TABLE 7. Number of Cigarettes Currently Smoked per Day in the Swedish Sample**

	(Tabular values in percent)				
Number of cigarettes smoked per day[a]	Twin–male	Twin–female	Progeny–male	Progeny–female	Combined (over sexes and generations)
1–5	10.8	19.0	16.1	13.5	14.6
6–10	25.7	29.3	16.1	23.0	23.5
11–15	23.0	31.0	25.8	25.7	26.1
16–20	30.0	17.2	22.6	33.8	26.5
21+	10.8	3.5	19.4	4.0	9.3
Number of subjects	74	58	63	74	269

*A sample of twins from the Swedish Twin Registry, their spouses, and their progeny.
[a]Not including very light current smokers and ex-smokers.

TABLE 8. Comparison of Gene Frequency Estimates for 21 Biochemical Marker Loci in the Swedish Sample with Those Obtained in Previous Studies*

Locus	Allele	Present sample	Other Swedish data[a]	Reference
ABO:	A_1	0.2421	0.2217	Beckman, 1959 [2]
	A_2	0.0789	0.0856	
	B	0.0915	0.0768	
	O	0.5875	0.6159	
Rh:	R_1	0.4222	0.4149	Beckman, 1959 [2]
	R_2	0.1802	0.1696	

(continued next page)

https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

TABLE 8. (cont'd)

Locus	Allele	Present sample	Other Swedish data[a]	Reference
	R_o	0.0049	0.0176	
	R_z	0.0012		
	r	0.3827	0.3805	
	r′	0.0037	0.0045	
	r″	0.0049	0.0129	
MNS:	MS	0.2416	0.570	Beckman, 1959 [2]
	Ms	0.3233		
	NS	0.0730	0.430	
	Ns	0.3621		
Kell:	K^b	0.0504	0.036	Heiken, 1962 [30]
	k^a	0.0024	0.964	
	k^b	0.9472		
Fy:	a	0.4247	[0.430	Reinskou, 1973 [52]
	b	0.5753	0.570]	
Jk:	a	0.3805	[0.506	Lundevall, 1956 [36]
	b	0.6194	0.494]	
Lu:	a	0.0354	...	
	b	0.9648	...	
P:	P_1	0.4938	0.4603	Beckman, 1959 [2]
	P_2	0.5061	0.5397	
Sec:	Se	0.5548	...	
	se	0.4445	...	
Gm:	b	0.6338	0.655	Broman et al, 1963 [10]
	ax	0.1266	0.345	
	a	0.2396		
Km:	a	0.0737	...	
	b	0.9263	...	
Hp:	1	0.3634	0.3821	Höglund et al, 1970 [31]
	2	0.6366	0.6179	
Tnf:	C	0.9848	0.995	Beckman et al, 1961 [5]
	B	0.0076	0.005	
	D	0.0076		
Gc:	1	0.7425	0.743	Monn et al, 1971 [44]
	2	0.2575	0.257	
Pi:	M	0.9593	[0.946	Fagerhol et al, 1969 [23]
	S	0.0253	0.023	
	Z	0.0154	0.016	
	F		0.013]	
AcP_1:	A	0.3858	0.372	Broman et al, 1971 [9]
	B	0.5637	0.558	
	C	0.0505	0.070	
6PGD:	A	0.9776	0.981	Beckman et al, 1971 [4]
	C	0.0224	0.019	
AK:	1	0.9614	0.965	Skude & Jakobsson, 1970 [59]
	2	0.0386	0.035	
EsD:	1	0.8858	0.9150	Beckman & Beckman, 1977 [3]
	2	0.1142	0.0850	
PGM_1:	1	0.8161	0.772	Beckman et al, 1971 [4]
	2	0.1839	0.228	
ADA:	1	0.9375	[0.950	Camoens et al, 1972 [11]
	2	0.0625	0.050]	

*A sample of twins from the Swedish Twin Registry, their spouses, and their progeny.

[a]Frequencies in brackets in this column were not available for Sweden, so Norwegian frequencies are substituted.

https://doi.org/10.1017/S0001566000009041 Published online by Cambridge University Press

*TABLE 9. Comparison of Observed Numbers of Concordant-Discordant Twin Pairs for Smoking Status (Current Smoker, Ex-Smoker, or Nonsmoker) in the Swedish Twin Sample**

	Concordant	Discordant	Total	χ^2	P
MZ males	25	14	39		
	(22.0)	(17.0)		2.12	0.15
DZ males	15	17	32		
	(18.0)	(14.0)			
MZ females	24	12	36		
	(18.6)	(17.4)		7.27	0.005
DZ females	10	20	30		
	(15.4)	(14.6)			
MZ twins	49	26	75		
	(40.5)	(34.5)		8.55	0.005
DZ twins	25	37	62		
	(33.5)	(28.5)			

*A sample of twins from the Swedish Twin Registry. Values are from the current study, not from previous information in the registry. Numbers in parentheses are expected counts within sexes under the hypothesis of no differences in twin types (first four rows) and then by the expected count for combined sexes (final two rows).

*TABLE 10. Comparison of Observed Numbers of Concordant-Discordant Twin Pairs For Smoking Status (Ever Smoked or Never Smoked) in the Swedish Twin Sample**

	Concordant	Discordant	Total	χ^2	P
MZ males	32	7	39		
	(30.2)	(8.8)		1.043	0.35
DZ males	23	9	32		
	(24.8)	(7.2)			
MZ females	27	9	36		
	(22.9)	(13.1)		4.42	0.03
DZ females	15	15	30		
	(19.1)	(10.9)			
MZ twins	59	16	75		
	(53.1)	(21.9)		4.96	0.03
DZ twins	38	24	62		
	(43.9)	(18.1)			

*A sample of twins from the Swedish Twin Registry. Values are from the current study, not from previous information in the registry. Numbers in parentheses are expected counts within sexes under the hypothesis of no differences in twin types (first four rows) and then by the expected count for combined sexes (final two rows).

the four kinds of comparisons – ie, MZ males, DZ males, MZ females, and DZ females. Therefore, the concordance values in our Swedish sample of twins appear to be similar to those for all Swedish twins of approximately the same age.

The results in Tables 9 and 10 are consistent with the hypothesis of a certain amount of genetic control over variation in smoking behavior. Verification of this hypothesis, accompanied by estimates of the degree of genetic control, its manifestation in the different components of smoking behavior, and its relation to the biochemical polymorphisms and ancillary variables, will depend on the outcome of more detailed analyses based on the methods presented in this report.

REFERENCES

1. Ablin RJ (1971): Association of alpha$_1$-antitrypsin deficiency and primary pulmonary emphysema as determined by the method of radial immunodiffusion. Acta Biol Med Germ 27:667–670.
2. Beckman L (1959): A contribution to the physical anthropology and population genetics of Sweden. Hereditas 45:1–189.
3. Beckman G, Beckman L (1977): Population studies in northern Sweden. VII. Ethnic heterogeneity and prenatal selection in the esterase D polymorphism. Hum Hered 27:403–407.
4. Beckman G, Beckman L, Cedergren B (1971): Population studies in northern Sweden. II. Red cell enzyme polymorphism in the Swedish Lapps. Hereditas 69:243–248.
5. Beckman L, Holmgren G, Martensson EH (1961): Transferrin variants in the Swedish population. Proc 2nd Int Cong Hum Genet 737.
6. Boorman KE, Dodd BE (1970): "Introduction to Blood Group Serology." London: Churchill Press.
7. Böyum A (1968): Separation of leukocytes from blood and bone marrow. Scand J Clin Lab Invest 21:9–109 (Suppl 97).
8. Briscoe WA, Kueppers F, Davis AL (1966): A case of inherited deficiency of serum alpha$_1$-antitrypsin associated pulmonary emphysema. Am Rev Respir Dis 95:529–539.
9. Broman G, Grundin R, Lins PE (1971): The red cell acid phosphatase polymorphism in Sweden. Gene frequencies and application to disputed paternity. Acta Genet Med Gemmellol 20:77–81.
10. Broman B, Heiken A, Hirschfeld J (1963): Gm(a) frequencies in Sweden. Acta Genet (Basel) 13: 132–139.
11. Camoens H, Monn E, Berg K (1972): Genetic marker systems in the Arctic populations. III. Polymorphism of red cell adenosine deaminase (ADA) in Norwegian Lapps. Hum Hered 22:561–565.
12. Cavalli-Sforza LL, Feldman MW (1973): Cultural versus biological inheritance: Phenotypic transmission from parents to children (A theory of the effect of parental phenotypes on children's phenotypes). Am J Hum Genet 25:618–637.
13. Cederlöf R (1966): The twin method in epidemiological studies on chronic disease. Department of Sociology, University of Stockholm.
14. Cederlöf R, Friberg L, Jonson E, Kaij L (1961): Studies in similarity diagnosis in twins, with the aid of mailed questionnaires. Acta Genet 11:338–362.
15. Cederlöf R, Friberg L, Lundman T (1977): The interactions of smoking, environment, and heredity and their implications for disease etiology. Acta Med Scand Suppl 612.
16. Comrey AL (1970): "Comrey Personality Scales, Manual." San Diego: Educational and Industrial Testing Service.
17. Conterio F, Chiarelli B (1962): Study of the inheritance of some daily life habits. Heredity 17:347–359.
18. Dies R, Honeyman M, Reznikoff M, White C (1969): Personality and smoking patterns in a twin population. J Projective Tech Pers Assess 33:457–463.
19. Elston RC (1973): Reply to "Message from a referee on the Elston method." Behav Genet 3:319–320.
20. Elston RC, Gottesman II (1968): The analysis of quantitative inheritance simultaneously from twin and family data. Am J Hum Genet 20:512–521.
21. Eysenck HJ (1973): Personality and the maintenance of the smoking habit. In Dunn WL (ed): "Smoking Behavior: Motives and Incentives." New York: Wiley, pp 113–146.
22. Eysenck HJ, Eysenck SBG (1975): "Manual of the Eysenck Personality Questionnaire." London: Hodder and Stoughton.
23. Fagerhol MK, Eriksson AW, Monn E (1969): Serum Pi types in some Lappish and Finnish populations. Hum Hered 19:360–364.
24. Fisher RA (1958): Cancer and smoking. Nature 182:596.
25. Giblett ER, Anderson JE, Cohen F (1972): Adenosine-deaminase deficiency in two patients with severly impaired cellular immunity. Lancet 4:1067–1069.
26. Giblett ER, Anderson JE, Chen SH, Teng YS, Cohen F (1974): Uridine monophosphate kinase: A new genetic polymorphism with possible clinical implications. Am J Hum Genet 26:627–635.
27. Glass GBJ, Ishimori A, Buckwalter JA (1962): ABO(H) blood group substances of the gastric juice in peptic ulcer, cancer of the stomach and atropic lesions of the gastric mucosa. Gastroenterology 42:443–454.

28. Guirgis HA, Townley RG, Schanfield MS, Fudenberg H, Burke K (1975): Gm allotypes and mecholyl inhalation in asthmatic and normal families. Am Rev Respir Dis 111:917.
29. Hauge M, Harvald B, Fischer M, Gotlieb-Jensen K, Juel-Nielson N, Raebild I, Shapiro R, Videbech T (1968): The Danish twin register. Acta Genet Med Gemellol 17:315–331.
30. Heiken A (1962): Distribution of the Kell blood group factor K in the Swedish population. Acta Genet (Basel) 12:352–358.
31. Höglund C, Heiken A, Hauge M (1970): Studies on the haptoglobin system. I. The distribution of Hp groups in the Swedish population. Hum Hered 20:549–556.
32. Hopkinson DA, Cook PJL, Harris H (1969): Further data on the adenosine deaminase (ADA) polymorphism and a report of a new phenotype. Ann Hum Genet 32:361–367.
33. Ikard FF, Green DE, Horn D (1969): A scale to differentiate between types of smoking as related to the management of affect. Int J Addict 4:649–659.
34. Jones MC, Thomas GO (1971): Alpha$_1$-antitrypsin deficiency and pulmonary emphysema. Thorax 26:652–662.
35. Kueppers F (1967): Immunologic assay of alpha$_1$-antitrypsin in deficient subjects and their families. Humangenetik 5:54–58.
36. Lundevall J (1956): The Kidd blood group system investigated with anti-Jk^a. Acta Pathol Microbiol Scand 38:39–42.
37. Mather K, Jinks JL (1971): "Biometrical Genetics." Ithaca, New York: Cornell University Press.
38. Mausner B, Platt ES (1971): "Smoking: A Behavioural Analysis." Oxford: Pergamon.
39. McClearn GE, DeFries JC (1973): "Introduction to Behavioral Genetics." San Francisco: Freeman.
40. McKennell AC (1970): Smoking motivation factors. Br J Soc Clin Psychol 9:8–22.
41. McNeil C, Trentelman EF, Ladle JN, Plank HP (1964): A relation between metastasis of cancer and secretion of blood group substances. Am J Clin Pathol 42:498–501.
42. McNeil C, Trentelman EF, Ladle JN, Helmich WM, Plank HP (1965): Blood group secretion factors in bronchogenic carcinoma. Nature 208:209.
43. Medlund P, Cederlöf R, Floderus-Myrhed B, Friberg L, Sörensen S (1976): A new Swedish twin registry. Acta Med Scand Suppl 660.
44. Monn E, Berg K, Reinksou T, Teisborg P (1971): Serum protein polymorphisms among Norwegian Lapps: Studies on the Lp, Ag, Gc, and transferrin systems. Hum Hered 21:134–139.
45. Nance WE, Corey LA (1976): Genetic models for the analysis of data from the families of identical twins. Genetics 83:811–826.
46. Nichols RC, Bilbro WC (1966): The diagnosis of twin zygosity. Acta Genet 16:265–275.
47. Ott J (1974): Estimation of recombination fraction in human pedigrees. Am J Hum Genet 26: 588–594.
48. Raaschou-Nielsen E (1960): Smoking habits in twins. Dan Med Bull 7:82.
49. Race RR, Sanger R (1975): "Blood Groups in Man." Oxford: Blackwell, Ed 6.
50. Rao DC, Morton NE, Yee S (1974): Analysis of family resemblance. II. A linear model for familial correlation. Am J Hum Genet 26:331–359.
51. Raven JC (1952): Guide to using Progressive Matrices (1938): London: H. K. Lewis & Co.
52. Reinskou T, cited by Berg K (1973): Studies of polymorphic traits for the characterization of populations. In Ramat B (ed): "Genetic Polymorphisms and Diseases in Man." New York: Academic Press, pp 21–29.
53. Renwick JH (1973): Message from a referee on the Elston method. Behav Genet 3:317–318.
54. Rotter JB (1966): Generalized expectancies for internal versus external control of reinforcement. Psychol Monogr 80 (1):whole no. 609.
55. Russell MAH, Peto J, Patel UA (1974): The classification of smoking by factorial structure of motives. J R Stat Soc A 137:313–333.
56. Schaefer ES (1958): Development of a parental attitude research instrument. Child Dev 29:339–361.
57. Schaefer ES (1965): Children's reports on parental behavior: An inventory. Child Dev 36:413–424.
58. Shields J (1962): "Monozygotic Twins Brought Up Apart and Brought Up Together." London: Oxford University Press.
59. Skude G, Jakobsson A (1970): Determination of adenylate kinase phenotypes employing agar gel. Hum Hered 20:319–324.
60. Smith SM, Penrose LS (1955): Monozygotic and dizygotic twin diagnosis. Ann Hum Genet 19: 273–289.

61. Stevens PM, Hnilica VS, Johnson PC, Bell RL (1971): Pathophysiology of hereditary emphysema. Ann Intern Med 75:672–681.
62. Teng YS, Anderson JE, Giblett ER (1975): Cytidine deaminase: A new genetic polymorphism demonstrable in granulocytes. Am J Hum Genet 27:492–497.
63. Tomkins S (1968): A modified model of smoking behaviour. In Borgatta EF, Evans RR (eds): "Smoking, Health and Behavior." Chicago: Aldine Publishing Company, pp 165–186.
64. Vandenberg SG (1967): Hereditary factors in normal personality traits (as measured by inventories). In Wortis J (ed): "Recent Advances in Biological Psychiatry," Vol. 9. New York: Plenum, pp 65–104.
65. Veldman DJ, Brown OH (1969): Personality and performance characteristics associated with cigarette smoking among college freshman. J Conslt Clin Psychol 33:109–119.

Correspondence: Professor David W. Crumpacker, Department of EPO Biology, University of Colorado, Boulder, CO 80309.