Hum Genet (1996) 98 : 467–475

ORIGINAL INVESTIGATION

Anne Maria Herskind · Matthew McGue
Ivan A. Iachine · Niels Holm · Thorkild I. A. Sørensen
Bent Harvald · James W. Vaupel

Untangling genetic influences on smoking, body mass index and longevity: a multivariate study of 2464 Danish twins followed for 28 years

Received: 20 May 1996 / Revised: 21 May 1996

Abstract A multivariate twin study was conducted in order to evaluate to what extent smoking, BMI and longevity are influenced by common genetic factors. The study was based on a 28-year follow-up of a sample of 2464 Danish twins who were born in the period 1890–1920 and who answered a questionnaire, including requests for information on smoking status, height and weight, in 1966. By 1994, approximately 2/3 of the sample had died. To compensate for the right-censoring, age at death was imputed for twins who were still alive by using survival analysis; all living subjects were more than 73 years old (mean 80 years, SD 5) in 1994. Proportions of covariance resulting from genetic and environmental factors in common and unique to the three traits were estimated from covariance matrices using the structural equation model approach. The study found no evidence for a substantial impact of common genetic factors on smoking, BMI and longevity. This suggests that only a small fraction of the genetic influences on longevity is mediated via a genetic influence on smoking and BMI and, furthermore, that it is unlikely that the associations between smoking and mortality and between BMI and mortality are confounded by common genetic factors.

A. M. Herskind (✉) · I. A. Iachine · J. W. Vaupel
Center for Health and Social Policy,
Institute of Community Health, Odense University,
Winsløwparken 171, DK-5000 Odense C, Denmark
Tel.: +45 66158696 ext. 4938; Fax: +45 65 906938;
e-mail: a.herskind@chsmed.ou.dk

M. McGue
Center for Twin and Adoption Studies,
Department of Psychology, University of Minnesota,
Minneapolis, USA

N. Holm
Genetic Epidemiologic Research Unit,
Institute of Community Health, Odense University,
Odense, Denmark

T. I. A. Sørensen
Danish Epidemiology Science Center
at the Institute of Preventive Medicine,
University of Copenhagen, Copenhagen, Denmark

B. Harvald
Odense University Medical School, Odense, Denmark

Introduction

Twin studies, which are one of the most powerful methods for assessing genetic and environmental contributions to inter-individual variation, indicate that there is a heritable component to variation in human longevity (Herskind et al. 1996a; Hrubec and Neel 1981; Jarvik et al. 1960; McGue et al. 1993), a finding that is supported by family and adoption studies (Sørensen et al. 1988; Wyshak 1978). On the basis of an almost non-censored sample of 5810 Danish twins born 1870–1900, we have recently estimated the heritability of longevity, i.e. the proportion of the variation in longevity attributable to genetic factors, to be 0.25 (Herskind et al. 1996a). However, it is not only interesting to estimate the magnitude of genetic influences, but also to determine how these genetic influences on lifespan are mediated. The heritability of human longevity might reflect specific genetic effects, e.g. 'killer genes' or length of telomeres (repetitive sequences at the end of DNA strands; Harley et al. 1990; Schächter et al. 1993; Steel 1995), or might be mediated by genetic influences on life-shortening diseases, conditions or life-style factors.

Several long-term and prospective epidemiological studies have shown that factors such as smoking and body mass index (BMI = weight/height^2), have a major impact on longevity (Ben-Schlomo et al. 1994; Doll et al. 1994; Kushner 1993; Waaler 1984). Furthermore, quantitative genetic studies suggest that variation in both smoking and BMI have a strong genetic component (Bouchard 1994; Heath and Madden 1995; Herskind et al. 1996b). Therefore, the genetic influences on longevity might be accounted for, at least in part, by genetic influences on covariates such as smoking and BMI.

Erroneous inferences about causality in the relationship between the covariates (smoking and BMI) and longevity would be made if the covariates and longevity

were caused by the same latent genetic or environmental variables. This idea of "genetic confounding", which has been a subject of controversy (Clarke 1991; Donner 1991; Eysenck 1991; Greenland 1991; Stolley 1991a, b; Vandenbroucke 1991a, b), was originally proposed by R.A. Fisher, who suggested that the association between smoking and lung cancer was attributable to a genetic predisposition to smoking that was also linked to lung cancer (Fisher 1958a, b).

Both the question of how the genetic influences on longevity are mediated and the question of "genetic confounding" can be addressed in twin studies but only a few studies have been conducted so far. Hrubec et al. (1984) have found no reduction in the heritability of longevity when controlling for covariates (smoking, marital status and police registration for alcohol use), suggesting that the genetic influences on longevity are not mediated through genetic influences on the covariates. Case-control studies of smoking discordant monozygotic twins (MZ), who are genetically identical, do not support the hypothesis of "genetic confounding"' in the association between smoking and mortality (Floderus et al. 1988; Kaprio and Koskenvuo 1989, 1990). However, all of these previous studies are left-truncated and heavily right-censored.

In a sample of 2464 Danish twins born 1890–1920 and followed from 1966–1994, we have estimated, by path analysis, the extent to which the heritability of longevity is accounted for by genetic influences on covariates such as smoking and BMI. In addition, we have explored whether the association between either smoking and longevity or between BMI and longevity is caused by common latent genetic or environmental factors.

Materials and methods

Subjects

This study was based on The Danish Twin Register, which includes all twin pairs born in Denmark between 1870 and 1910 and all like-sexed pairs born between 1911 and 1930 in which both co-twins survived to the age of 6 years (Hauge 1981). Zygosity was determined by self-reported degree of similarity. A comparison of this method with laboratory methods (serological markers) showed that misclassification arising from self-report occurs in less than 5% of the cases (Hauge 1981).

In 1966, a questionnaire was mailed to non-emigrant like-sexed twins who had been born in the period 1890–1920 and who were traced and alive at the time. The questionnaire contained questions about health, social, behavioral and physical characteristics. A total of 3709 individuals answered the questionnaire yielding an individual response rate of 65%.

Excluded from the study were 813 twins whose co-twin did not participate in the survey, pairs of unknown zygosity (4 pairs) and pairs with incomplete information on height and weight for both co-twins (190 pairs). It was also decided to exclude pairs in which either twin had a BMI < 15 kg/m^2 (2 pairs) or a BMI > 45 kg/m^2 (20 pairs). Individuals with a BMI of 45–66 kg/m^2 constituted a distinct group of outliers (36 twins). Through hospital records and/or information from the co-twin, we were able to determine that 16 of these twins had given their weight in pounds, but we were not able to verify the values for the remaining 20 twins. Because of this uncertainty, we decided to exclude these cases. Among a 5% sample of twins with a BMI of 15–45 kg/m^2, none were found to have given their weight in pounds. Thus, 1232 twin pairs, viz. 213 MZ male, 322 DZ (dizygotic) male, 279 MZ female and 418 DZ female pairs, were included. In 1966, the age range of the sample was 46–76 years.

Death status was obtained from the Central Person Register, the Danish Cause-of-Death Register and various other public registers in Denmark. The last follow-up was completed on May 1, 1994. Some 968 individuals out of 2464 (39%) were still alive at the date of the last follow-up but all were older than 73 years (in 1994, the mean was 80.3 years with an SD of 5.0 and a range 73.3–95.3 years for males, and 80.6 years with an SD of 5.2 and a range 73.4–98.8 years for females). A total of 566 pairs were concordant for death, 364 discordant and 302 concordant alive. The status of the twins by sex and year of birth is presented in Fig. 1. Mean age at death for the non-censored part of the sample was 75.4 years (SD 8.2, range 53.8–97.0) for males and 77.2 years (SD 9.2, range 51.4–98.2) for females.

Smoking, BMI and mortality

In a multivariate twin analysis, standard model fitting procedures assume a linear relationship between the variables (Neale and Car-

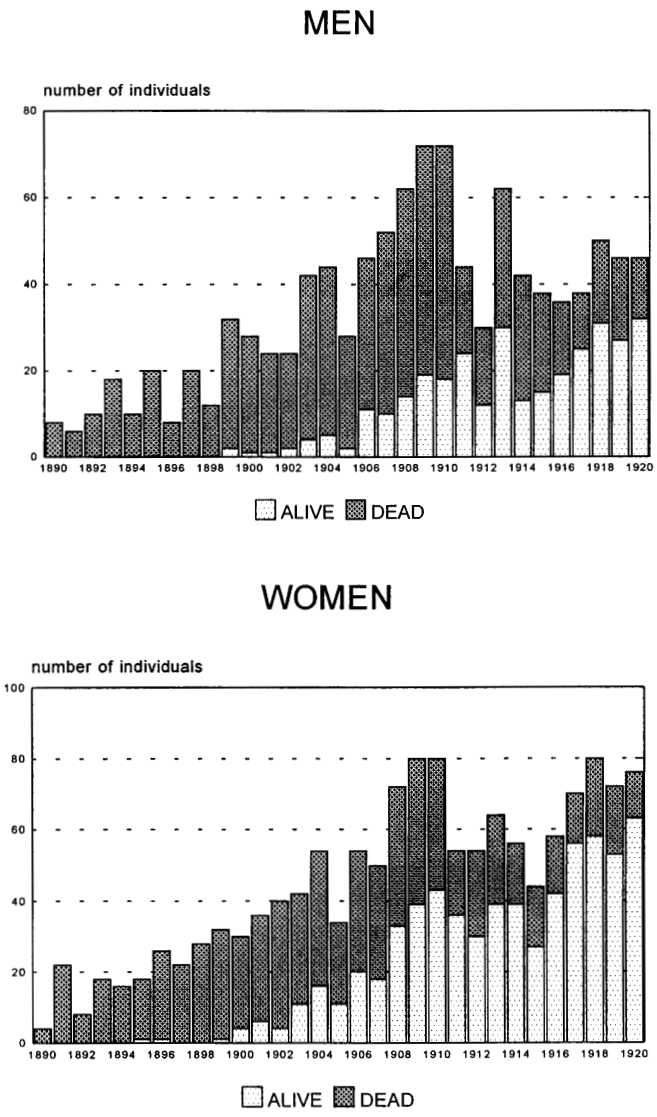

Fig. 1 Histogram illustrating absolute number of twins born and deceased (by May 1, 1994) by sex and birth cohort

Table 1 Number of individuals in a sample of 2464 Danish twins by sex, smoking and BMI

		Non-smokers	Others (ex-smokers, etc.)	Current pipe/cigar smokers	Current cigarette smokers	Total
Males	BMI < 22	11	12	45	33	101 (9.4%)
	BMI 22–28	81	142	338	211	772 (72.1%)
	BMI > 28	25	35	93	44	197 (18.4%)
	Total	117 (10.9%)	189 (17.7%)	476 (44.5%)	288 (26.9%)	
Females	BMI < 22	106	48	48	102	304 (21.8%)
	BMI 22–28	355	142	137	175	809 (58.0%)
	BMI > 28	159	41	37	44	281 (20.2%)
	Total	620 (44.5%)	231 (16.6%)	222 (15.9%)	321 (23.0%)	

don 1992). Smoking status was divided into four groups of increasing risk: non-smokers, ex-smokers, current cigar or pipe smokers and current cigarette smokers. Non-smokers were defined as individuals who had smoked less than 100 cigarettes, 50 cigars or cigarillos, or 5 packs of tobacco in their lifetime. Persons who gave no information on life-time smoking status (117 individuals) or current smoking status (52 individuals) were pooled with the ex-smokers (251 individuals). Current cigarette smokers who also smoked other kinds of tobacco were included in the group of current cigarette smokers. Because the association between BMI and mortality is J- or U-shaped rather than linear, BMI was divided into three groups based on increasing risk of mortality: 22–28 kg/m^2, < 22 kg/m^2 and > 28 kg/m^2. The distribution of smoking and BMI by sex is given in Table 1. Mean age in 1966 was not significantly different in the various smoking and BMI subgroups.

As previously mentioned, despite a long follow-up, the data set was partly right-censored. Because there is no standard method available to handle right-censored twin data, censored observation times were replaced by imputed uncensored survival times by using survival analysis taking all available survival and covariate information for the twin pair in question into account (see Appendix). To test the reliability of the imputation, 10 different data sets were constructed.

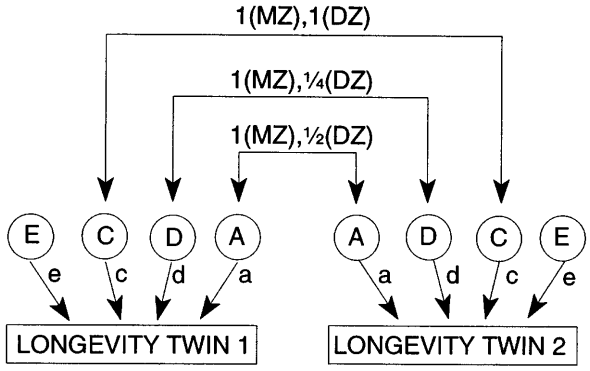

Fig.2 Path diagram illustrating the univariate case. Latent variables are in *circles*; observed variables are depicted by *squares*. *Double-headed arrows* represent correlations, *single-headed arrows* link the observed variables to the latent variables, *A* additive genetic factors, *D* genetic factors attributable to dominance, *C* shared environmental factors, *E* non-shared environmental factors, *MZ* monozygotic twins, *DZ* dizygotic twins

Biometrical models

Covariance between twins can provide useful information about genetic and environmental contributions to variation within individuals. Structural equation (SE) modelling, which has been described in detail elsewhere (Neale and Cardon 1992), is one of the most powerful approaches for separating the sources of phenotypic variation. The path diagram in Fig. 2 illustrates the univariate model for decomposing variance in longevity. The total phenotypic variance can be decomposed into two genetic and two environmental components. Additive genetic factors (A) are the effects of genes taken singly and added over multiple loci, whereas genetic dominance (D) represents genetic interaction (within loci). Shared environmental effects (C) are the environmental effects that are shared by family members, and non-shared environmental effects (E) are the environmental influences that are not shared by family members. The diagram indicates how each type of factor contributes to the covariance within an MZ or DZ twin pair. Additive genetic factors and genetic dominance are perfectly correlated in MZ twins, whereas DZ twins, like ordinary siblings, share only half of the additive genetic effects and one quarter of the genetic dominance effects. Shared environmental effects are perfectly correlated in both MZ and DZ twins. Lower case letters (path coefficients) represent the genetic and environmental loadings on the trait. The model assumes negligible effects of assortative mating, epistasis, genotype-environment correlation and/or interaction. It is also assumed that shared environmental influences are similar for MZ and DZ twins (see Neale and Cardon 1992 for an analysis of these assumptions).

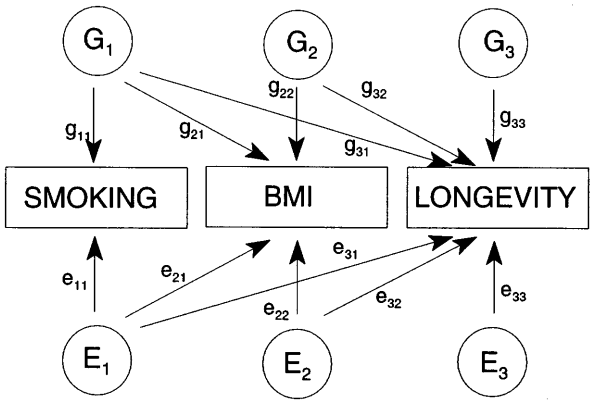

Fig. 3 Cholesky path diagram for the multivariate case depicting common and unique factors for genetic and environmental sources of variance and covariance for smoking, body mass index and longevity. The figure illustrates only one twin in a pair. *G* Genetic factors, *g* genetic loadings, *E* environmental factors, *e* environmental loadings

In a manner similar to the decomposition of phenotypic variance, the genetic and environmental contributions to the covariance between traits can be assessed. A Cholesky model illustrating the multivariate case for smoking, BMI and longevity (Neale and

Table 2 Correlation matrix for smoking, body mass index and longevity in Danish MZ and DZ *male* twins (example from data set no. 1). Correlation: Pearson's product moment, polychoric or polyserial correlations depending on whether both variables are continuous, both categorical or one is categorical and the other continuous

			Twin 1			Twin 2		
			Smoking	BMI	Longevity	Smoking	BMI	Longevity
MZ	Twin 1	Smoking	1.000					
		BMI	0.083	1.000				
		Longevity	–0.315	0.025	1.000			
	Twin 2	Smoking	0.507	–0.052	–0.103	1.000		
		BMI	0.049	0.482	–0.021	–0.055	1.000	
		Longevity	–0.217	0.016	0.437	–0.213	–0.094	1.000
DZ	Twin 1	Smoking	1.000					
		BMI	–0.025	1.000				
		Longevity	–0.065	0.022	1.000			
	Twin 2	Smoking	0.214	–0.092	–0.121	1.000		
		BMI	–0.009	0.157	–0.074	–0.091	1.000	
		Longevity	–0.067	0.082	0.189	–0.169	–0.042	1.000

Table 3 Correlation matrix for smoking, body mass index and longevity in Danish MZ and DZ *female* twins (example from data set no. 1). Correlation: Pearson's product moment, polychoric or polyserial correlations depending on whether both variables are continuous, both categorical or one is categorical and the other continuous

			Twin 1			Twin 2		
			Smoking	BMI	Longevity	Smoking	BMI	Longevity
MZ	Twin 1	Smoking	1.000					
		BMI	–0.085	1.000				
		Longevity	–0.115	–0.072	1.000			
	Twin 2	Smoking	0.585	–0.017	–0.090	1.000		
		BMI	–0.018	0.647	–0.077	0.031	1.000	
		Longevity	–0.004	–0.151	0.353	–0.137	–0.192	1.000
DZ	Twin 1	Smoking	1.000					
		BMI	–0.075	1.000				
		Longevity	–0.224	–0.131	1.000			
	Twin 2	Smoking	0.345	0.043	–0.106	1.000		
		BMI	–0.116	0.136	0.018	–0.092	1.000	
		Longevity	–0.072	–0.057	0.098	–0.129	–0.088	1.000

Cardon 1992) is given in Fig. 3. In this model, the proportion of total variance in longevity caused by genetic effects, i.e. heritability (H^2), is the sum of the squared standardized path coefficients:

$$H^2_{longevity} = g_{31}^2 + g_{32}^2 + g_{33}^2$$

The heritability of longevity is thus accounted for by genetic factors in common with smoking and BMI ($g_{31}^2 + g_{32}^2$) and genetic factors that are specific to longevity (g_{33}^2).

The proportion of total variance caused by environmental effects, i.e. environmentality (E^2) is similarly decomposed:

$$E^2_{longevity} = e_{31}^2 + e_{32}^2 + e_{33}^2$$

The genetic correlation measures the extent to which two phenotypes share genetic effects. The genetic correlation between smoking and longevity is:

$$r_{g\ smoking,longevity} = g_{11} * g_{31}/(g_{11}^2 * (g_{31}^2 + g_{32}^2 + g_{33}^2))^{1/2}$$

and between BMI and longevity is:

$$r_{g\ BMI,longevity} = g_{22} * g_{32}/((g_{21}^2 + g_{22}^2) * (g_{31}^2 + g_{32}^2 + g_{33}^2))^{1/2}$$

The environmental correlations are computed in the same way:

$$r_{e\ smoking,longevity} = e_{11} * e_{31}/(e_{11}^2 * (e_{31}^2 + e_{32}^2 + e_{33}^2))^{1/2}$$

and

$$r_{e\ BMI,longevity} = e_{22} * e_{32}/((e_{21}^2 + e_{22}^2) * (e_{31}^2 + e_{32}^2 + e_{33}^2))^{1/2}$$

Data on BMI and smoking used in the present analysis were categorical. To fit structural equation models to these data, an underlying normally distributed "liability" for BMI and smoking, respectively, was assumed. Correlations between the three phenotypes, which are equal to the covariances in the present case, for each sex and zygosity group were computed using the programme PRELIS 2.01 (Jöreskog and Sörbom 1993).

A Cholesky model including additive genetic effects (A) and non-shared environmental factors (E) was fitted to the correlations. This AE model was chosen because it gave a good fit in each of the three univariate cases. Shared environmental factors seem to have no effect on BMI (Herskind et al. 1996b; Vogler et al. 1995) or longevity (Herskind et al. 1996a; Sørensen et al. 1988), whereas dominance effects were found to be non-significant for both smoking (Heath and Madden 1995) and BMI (Herskind et al. 1996b). Asymptotic weighted least squares estimates for the genetic and environmental parameters were obtained by using the statistical software Mx 2.1 (Neale 1994).

Results

Correlation matrices for the three phenotypes are shown in Tables 2 and 3. Correlations for longevity between twin 1 and twin 2 in a pair were significantly different from

Table 4 Standardized path coefficients derived from a multivariate analysis of 2464 Danish twins (example from data set no. 1). *Brackets* refer to path-coefficients in Fig. 3

		G_1	G_2	G_3	E_1	E_2	E_3
Males	Smoking	0.8699 (g_{11})			0.4933 (e_{11})		
	BMI	0.0589 (g_{21})	0.6592 (g_{22})		–0.0048 (e_{21})	0.7496 (e_{22})	
	Longevity	0.0070 (g_{31})	0.1001 (g_{32})	0.6440 (g_{33})	–0.1193 (e_{31})	–0.0439 (e_{32})	0.7477 (e_{33})
Females	Smoking	0.7798 (g_{11})			0.6261 (e_{11})		
	BMI	0.0125 (g_{21})	0.7896 (g_{22})		–0.0340 (e_{21})	0.6125 (e_{22})	
	Longevity	–0.1130 (g_{31})	–0.1100 (g_{32})	0.5541 (g_{33})	–0.1151 (e_{31})	–0.0447 (e_{32})	0.8080 (e_{33})

Table 5 Heritability estimates for longevity with and without control for covariates, together with genetic and environmental correlations for smoking, BMI and longevity (mean, range and standard error of the mean for 10 data sets) in a sample of 2464 Danish twins (AE model)

		Males			Females		
		Mean	Range	SE	Means	Range	SE
Heritability of longevity		0.42	0.37–0.47	0.03	0.33	0.30–0.34	0.02
Heritability attributable to smoking and BMI		0.02	0.01–0.04	0.01	0.03	0.02–0.05	0.01
Heritability residual		0.40	0.37–0.45	0.03	0.29	0.27–0.32	0.02
Smoking-longevity	Genetic correlation	0.07	(–0.01)–0.13	0.04	–0.23	(–0.16)–(–0.30)	0.02
	Environmental correlation	–0.24	(–0.16)–(–0.33)	0.06	–0.14	(–0.08)–(–0.18)	0.04
BMI-longevity	Genetic correlation	0.18	0.15–0.29	0.05	–0.20	(–0.14)–(–0.26)	0.06
	Environmental correlation	–0.09	(–0.06)–(–0.14)	0.03	–0.06	(–0.01)–(–0.11)	0.03

zero in all sex and zygosity groups. The data in Tables 2 and 3 further reveal that the correlation for longevity between twin 1 and twin 2 was higher for MZ than for DZ twins, suggesting genetic influences on longevity. The same pattern of twin correlation was observed for both smoking and BMI.

Smoking was consistently and significantly negatively associated with longevity within all groups (the association was non-significant only for twin 1 in the DZ male group). The association between BMI and longevity was significant or borderline-significant in 3 of the 4 female groups but in none of the male groups.

The model fitting procedure showed that the AE model provided a good fit to the data (males: $\chi_{21}^2 = 25.73$, $P = 0.259$; females: $\chi_{21}^2 = 17.46$, $P = 0.683$). The asymptotic weighted least squares estimates for the standardized path coefficients are presented in Table 4. Off-diagonal elements (g_{21}, g_{31}, g_{32}, e_{21}, e_{31}, e_{32}) were small, suggesting no substantial influence of common genetic or environmental factors on the three traits. All 10 replicate data sets showed consistent results; therefore, the model fit index, correlations and path coefficients are given for one data set only.

Estimates of the heritability of longevity in total and decomposed in terms of the portion in common with and independent of smoking and BMI are given in Table 5, which also gives the genetic and environmental correlations between the three traits for males or females (mean of 10 data sets). The range and standard error of the mean illustrate the small impact of imputation on the estimates. As can be seen from Table 5, only a small and non-significant fraction (5%–9%) of the heritability of longevity can be accounted for by genetic influences on smoking and BMI. The genetic and environmental correlations are relatively small and not significantly different from zero.

Discussion

The two major findings of this study are: (1) that only a small fraction of the genetic influences on longevity is accounted for by genetic influences on smoking and BMI; (2) that it is unlikely that the associations between smoking and longevity and between BMI and longevity are the result of common genetic factors. Our study is based on self-reported information on height, weight and smoking. Because smoking was socially acceptable in the 1960s, there is reason to believe that the self-reported information on smoking behaviour is reliable. Only a few studies have assessed the validity of self-reported BMI. Some studies find self-reported information on height and weight to depend on sex, age and body size (Kuskowska-Wolk et al. 1989), whereas others have found these measurements to be useful in large studies (Stunkard and Albaum 1981). The validity of the longevity data for the Danish twin panel of like-sexed cohorts born 1881–1920 has been investigated by comparing information on death status with the nation-wide Danish Cancer Registry. The two registers are ascertained independently but show 99% agreement on year of death (Holm 1983). Further data correction has increased this to almost 100%. Death rates

for twins and the general Danish population do not differ significantly after 6 years of age, except for a modest excess in female twin mortality in the 60–89 years age-range (1.14 times higher than in the standard population; Christensen et al. 1995). A substantial difference in death rate between twins and singletons would have indicated that results from twin studies might not be valid for the general population.

The present study is based on a single assessment of smoking status and BMI at one time point. The question therefore arises as to whether our assessment of these risk factors sufficiently represents life-long patterns to allow reliable prediction of death. The smoking information seems acceptable. People were 46–76 years old when they filled in the questionnaire, a time in life when it is very rare to start smoking. Breaking the smoking habit after filling in the questionnaire would tend to reduce the effect of smoking on mortality. Waaler (1984) has investigated the relationship between BMI and mortality in a sample of 1.7 million Norwegians aged 15 years and above. With one life-time measurement of BMI and by following the subjects for 10 years, he found an association between BMI and risk of mortality for all age-groups. Nevertheless, this effect might not be significant in smaller studies (Sjöström 1992). Individual relative risks for the 12 different sub-groups in the present study (Tables 6, 7) are in accordance with previous literature (Ben-Schlomo et al. 1994; Doll et al. 1994; Waaler 1984). Cigarette smokers had the highest risk of mortality, then pipe/cigar smokers, ex-smokers and non-smokers. The relative risk of mortality for the different BMI sub-groups showed a J-shaped tendency; however, the difference in relative risk between the BMI sub-groups were only significant for females. The relative risks given in Table 7 may seem large compared with the actual correlations in Tables 2 and 3 between smoking and longevity and between BMI and longevity, although it has been shown that even large differences in relative risk may not produce substantial correlations (Vaupel 1988).

Despite a 28-year follow-up, the study population was right-censored, with 39% being still alive in 1994. The sample was, however, less right-censored than in previous reports (Floderus et al. 1988; Hrubec et al. 1984; Kaprio and Koskenvuo 1989). Earlier studies have handled this right-censoring by analysing longevity as a categorical variable (dead/alive), an approach that deals with the "liability to death" (see Neale and Cardon 1992) instead of longevity (Hrubec and Neel 1981; Hrubec et al. 1984; Jarvik et al. 1960). We imputed age at death for the censored part of the sample by using survival analysis. Remaining life expectancy alone could have been used to estimate age at death for the fraction of twins still alive but we also took into account all available information on survival, smoking and BMI for the pair in question when age of death was imputed. There are several advantages of this approach compared with the liability method: We used the information that all twins who were still alive in 1994 had survived to at least 73 years (mean 80, SD 5), together with information on smoking behaviour and BMI.

The higher heritability for longevity found in the present study compared with the study conducted on non-censored data (5810 twins born 1870–1900 with only 0.6% still alive in 1994), which provides the most direct univariate estimate of the heritability of longevity, might have arisen because of the inclusion of different birth cohorts in the two studies (more recent birth cohorts might be influenced less by environmental chance), truncation (accidental deaths might be relatively more frequent in younger than older age groups) or imputation of a fraction of the data in the present study. However, in the non-censored study, no cohort differences were found and all 10 imputed data sets in the present study gave consistent results.

There has been much speculation about the mediation of genetic influences on longevity. Hypotheses of a direct genetic influence on lifespan have been suggested (Harley et al. 1990; Schächter et al. 1993; Steel 1995). Genetic factors could also influence lifespan more indirectly through, e.g., development of diseases such as cardiovascular diseases, cancer, infections or obesity, all of which reduce length of life (Sørensen et al. 1988). Mediation through behaviour such as smoking, alcohol intake or stress is also possible and could take place via a genetic predisposition to a personality type or to genetically influenced taste aversions or sensitivities to alcohol. Studies within behavioural genetics have found a genetic influence on behavioural traits that are injurious to health (Heath and Madden 1995).

Smoking and BMI have been established as being genetically influenced. Twin studies demonstrate a heritability of BMI in the range of 0.5 to 0.8 throughout adult life (Bouchard 1994; Herskind et al. 1996b), whereas the estimated heritability is significant, but lower, in both adoption and family studies (Bouchard 1994). The heritability of smoking has been repeatedly estimated as being in the range 0.35 to 0.75 (Heath and Madden 1995).

Both smoking and obesity significantly increase the risk of death. The relative risk for cigarette smokers and ex-cigarette smokers depends on the type of tobacco, duration and extent of smoking and years since giving up the habit. Ex-cigarette smokers have an intermediate mortality risk compared with individuals who have never smoked and current smokers, although whether former smokers ever reach the same level of mortality as people who have never smoked remains controversial. Both pipe and cigar smokers have a lower risk than cigarette smokers (Ben-Schlomo et al. 1994; Doll et al. 1994). Several large and long-term prospective studies have found a J- or an asymmetric U-shaped relationship between BMI and mortality with an approximately 1.5-fold to 2-fold increase in all-cause mortality at the endpoints of the curve. There is a wide optimum area for values between 22–28 kg/m^2 (Kushner 1993; Waaler 1984).

Multivariate twin studies provide an opportunity for testing whether the genetic influences on mortality are accounted for by genetic influences on certain covariates; however, only a few twin studies have reported an estimate of the heritability of longevity with a control for co-

variates. A study based on the Swedish Twin Register (Hrubec et al. 1984), including twins born 1886–1925, focused on the relationship between mortality and the covariates smoking, marital status and police registration for alcohol abuse. As in the present study, twins were only included if they survived to a certain year (1962), when the questionnaires were sent out. In addition, the study was heavily right-censored and, at the time of analysis (1980), only 1/4 of the twins had died. Using a Cox proportional hazard model, the Swedish data suggested that genetic influences on mortality were not accounted for by genetic influences on the covariates. Heritability of "liability to death" was 0.4–0.6 depending on cohort and was highest among the youngest and most right-censored.

In the late 1950s, R.A. Fisher started the lung cancer/smoking controversy, which was the subject of a long debate (e.g. Clarke 1991, Donner 1991, Eysenck 1991, Greenland 1991, Stolley 1991a, b, Vandenbroucke 1991a, b). Fisher suggested that the association between smoking and lung cancer was not due to smoking causing cancer but rather reflected that smoking and lung cancer were influenced by common genetic factors. This hypothesis can be extended to the relationship between smoking and all-cause mortality.

The Swedish study by Hrubec et al. (1984) found that, when genetic factors were controlled for, an effect of smoking on mortality was apparent. A twin control study, based on the Swedish Twin Register (Floderus et al. 1988), compared deaths in smoking MZ twins and their non-smoking co-twins, a study design that controlled for the effect of genes. A relative risk of 1.6 ($P = 0.056$) was found for male smokers for all causes of death, suggesting that smoking is a risk factor for mortality. Among females, the relative risk was 1.2 ($P = 0.315$). However, despite a 21-year follow-up, this study was also left-truncated and heavily right-censored (30% deceased). Similar results with a relative risk of all-cause mortality for smokers of 13.0 ($P < 0.01$) were obtained in a Finnish twin control study with a follow-up of 12 years (Kaprio and Koskenvuo 1989).

The present multivariate study on longevity in twins is less right-censored than the previously mentioned studies. Furthermore, the right-censoring has been compensated for by using imputed values for age at death for those who were still alive. Imputation was performed by using survival analysis. In contrast to the other twin studies, path analysis was used, a method that is well-suited for partitioning covariance between traits into genetic and environmental factors.

The study has shown that the genetic influences on longevity are not mediated to any substantial degree through genetic influences on smoking and BMI. Hence, longevity must either be influenced directly by genetic factors or be mediated via genetic influences on other behavioural patterns, diseases or conditions. The gene coding for ApoE has received considerable attention as a "longevity gene" (van Bockxmeer 1994) but a recent study has shown that only approximately 1% of the genetic variance in longevity is accounted for by genetic variance in ApoE (U. Gerdes, H. Olsen, K. Andersen-Ranberg, J. W. Vaupel, B. Jeune, unpublished). Many different genes are probably involved, each having a small effect on longevity. Furthermore, we have found no evidence that the associations between smoking and mortality and between BMI and mortality are "confounded" by genetic or environmental factors, indicating that Fisher's theory does not hold for the association between smoking and all-cause mortality.

Acknowledgements The authors wish to thank EDP-assistant Susanne Berntsen, secretary Gudrun Hauge, Axel Skytthe (for technical assistance) and Anatoli I. Yashin (for valuable discussions). This study was supported by the United States National Institute of Aging (Research grant P01-AG08761) and The Danish Research Council.

References

Ben-Schlomo Y, Smith GD, Shipley MJ, Marmot MG (1994) What determines mortality risk in male former cigarette smokers? Am J Public Health 84: 1235–1242

Bockxmeer FM van (1994) ApoE and ACE genes: impact on human longevity. Nat Genet 6: 4–5

Bouchard C (1994) Genetics of obesity: overview and research directions. In: Claude Bouchard (ed) The genetics of obesity, 1st edn. CRC Press, Boca Raton, pp 223–233

Clarke C (1991) Invited commentary on R.A. Fisher. Am J Epidemiol 134: 1371–1374

Christensen K, Vaupel JW, Holm NV, Yashin AI (1995) Mortality among twins after age 6: fetal origins hypothesis versus twin method. BMJ 310: 432–436

Doll R, Peto R, Wheatley K, Gray R, Sutherland I (1994) Mortality in relation to smoking: 40 years' observations on male British doctors. BMJ 309: 901–910

Donner A (1991) Editorial: placing R.A. Fisher in a broader perspective. Am J Epidemiol 134: 1369–1370

Eysenck HJ (1991) Were we really wrong? Am J Epidemiol 133: 429–433

Fisher RA (1958a) Cancer and smoking. Nature 182: 596

Fisher RA (1958b) Lung cancer and cigarettes. Nature 182: 108

Floderus B, Cederlof R, Friberg L (1988) Smoking and mortality: a 21-year follow-up based on the Swedish Twin Registry. Int J Epidemiol 17: 332–340

Greenland S (1991) Invited commentary: science versus public health action: those who were wrong are still wrong. Am J Epidemiol 133: 435–436

Harley CB, Futcher AB, Creider CW (1990) Telomeres shorten during ageing of human fibroblasts. Nature 345: 458–460

Hauge M (1981) The Danish Twin Register. In: Mednick SA, Baert AE, Bachmann BP (eds) Prospective longitudinal research: an empirical basis for the primary prevention of psychosocial disorders. Oxford University Press, Oxford, pp 217–221

Heath AC, Madden PAF (1995) Genetic influences on smoking behavior. In: Turner JR, Cardon LR, Hewitt JK (eds) Behavior: genetic approaches in behavioral medicine. Plenum Press, New York London, pp 45–63

Herskind AM, McGue M, Holm NV, Sørensen TIA, Harvald B, Vaupel JW (1996a) The heritability of human longevity. A population-based study of 2872 Danish twin pairs born 1870–1900. Hum Genet 97: 319–323

Herskind AM, McGue M, Sørensen TIA, Harvald B (1996b) Sex and age specific assessment of genetic and environmental influences on body mass index in twins. Int J Obes 20: 106–113

Holm NV (1983) The use of twin studies to investigate causes of diseases with complex etiology with a focus on cancer (in Danish). Odense University, Odense

Hrubec Z, Neel JV (1981) Familial factors in early deaths: twins followed 30 years to ages 51–61 in 1978. Hum Genet 59: 39–46
Hrubec Z, Floderus-Myrhed B, Faire U de, Sarna S (1984) Familial factors in mortality with control of epidemiological covariables. Swedish twins born 1886–1925. Acta Genet Med Gemellol (Roma) 33: 403–412
Jarvik L, Falek A, Kallmann FJ, Lorge I (1960) Survival trends in a senescent twin population. Am J Hum Genet 12: 170–179
Jöreskog KG, Sörbom D (1993) Windows Prelis 2.01. Scientific Software Int., 1525 East 53rd Street, Chicago, IL 60615, USA
Kaprio J, Koskenvuo M (1989) Twins, smoking and mortality: a 12-year prospective study of smoking-discordant twin pairs. Soc Sci Med 29: 1083–1089
Kaprio J, Koskenvuo M (1990) Cigarette smoking as a cause of lung cancer and coronary heart disease. A study of smoking-discordant twin pairs. Acta Genet Med Gemellol (Roma) 39: 25–34
Kushner RF (1993) Body weight and mortality. Nutr Rev 51: 127–136
Kuskowska-Wolk A, Karlsson P, Stolt M, Rössner S (1989) The predictive value of body mass index based on self-reported weight and height. Int J Obes 13: 441–453
McGue M, Vaupel JW, Holm N, Harvald B (1993) Longevity is moderately heritable in a sample of Danish twins born 1870–1880. J Gerontol 48: B237-B244
Neale MC (1994) Mx: statistical modelling, 2nd edn. Department of Psychiatry, Box 710 MCV, Richmond, VA 23298, USA
Neale MC, Cardon LR (1992) Methodology for genetic studies of twins and families, 1st edn. Kluwer, Dordrecht
Schächter F, Cohen D, Kirkwood T (1993) Prospects for the genetics of human longevity. Hum Genet 91: 519–526
Sjöström LV (1992) Mortality of severely obese subjects. Am J Clin Nutr 55: 516S-523S
Sørensen TIA, Nielsen GG, Andersen PK, Teasdale T (1988) Genetic and environmental influences on premature death in adult adoptees. N Engl J Med 318: 727–732
Steel M (1995) Telomerase that shapes our ends. Lancet 345: 935–936
Stolley PD (1991a) When genius errs: R.A. Fisher and the lung cancer controversy. Am J Epidemiol 133: 416–425
Stolley PD (1991b) Author's response to "How much retropsychology?" Am J Epidemiol 133: 428
Stunkard AJ, Albaum JM (1981) The accuracy of self-reported weights. Am J Clin Nutr 34: 1593–1599
Vandenbroucke JP (1991a) Invited commentary: how much retropsychology? Am J Epidemiol 133: 426–427
Vandenbrouke JP (1991b) Invited commentary: reply to Eysenck. Am J Epidemiol 133: 434
Vaupel JW (1988) Inherited frailty and longevity. Demography 25: 277–287
Vogler GP, Sørensen TIA, Stunkard AJ, Srinivasan MR, Rao DC (1995) Influences of genes and shared family environment on adult body mass index assessed in an adoption study by a comprehensive path model. Int J Obes 19: 40–45
Waaler HT (1984) Height, weight and mortality. The Norwegian experience. Acta Med Scand 679: 1–56
Wyshak G (1978) Fertility and longevity in twins, sibs, and parents of twins. Soc Biol 25: 315–330
Yashin AI, Iachine IA (1994) Mortality models with application to twin survival data. In: Halin J, Karplus W, Rimane R (eds) CIIS: First Joint Conference of International Simulation Societies Proceedings, Zurich, Switzerland
Yashin AI, Iachine IA (1995) How long can humans live? Lower bound for biological limit of human longevity calculated from Danish twin data using correlated frailty model. Mech Ageing Dev 80: 147–69

Appendix Imputation of survival data for twins using the correlated frailty model

The survival data used in the study are incomplete: some of the life-spans are censored. Linear structural equation models, traditionally used for genetic studies of twins, cannot handle censored survival times were replaced by imputed uncensored survival times using the correlated frailty model.

The correlated frailty model is a bivariate survival model which is based on the notion of an individual gamma distributed frailty (Yashin and Iachine 1994). This model was previously applied to the analysis of the Danish twins' survival data without covariates (Yashin and Iachine 1995). To use this model in our analysis we had to modify it in order to include the observed covariates.

Let X_i,Z_i,Y_i i = 1,2 be the survival time, individual frailty and observed covariate information for the i^{th} twin in the pair respectively. Under the proportional hazards assumption the individual hazard rate is given by:

$$\mu(x \mid Z_i = z_i, Y_i = y_i) = z_i R(y_i)\mu_0(x) \tag{1}$$

where $\mu_0(x)$ is the baseline hazard and $R(y)$ is the relative risk function of the form:

$$R(y) = \begin{cases} 1, \text{if } 22 \le BMI \le 28 \\ \alpha_1, \text{if } BMI < 22 \\ \alpha_2, \text{if } BMI > 28 \end{cases} \times \begin{cases} 1, \text{if never smoked} \\ \beta_1, \text{if currently smoke cigarettes} \\ \beta_2, \text{if currently cigar/pipe smokers} \\ \beta_3, \text{if everyone else} \end{cases} \tag{2}$$

It can be shown that, if Z_i, i = 1,2 has a variance of σ^2 and $Corr(Z_1,Z_2) = \rho$, the bivariate survival function for a twin pair has the following form:

$$\begin{aligned} S(x_1, x_2 | y_1, y_2) &= \left(R(y_1)\left[S_B(x_1)^{-\sigma^2} - 1\right] + 1 \right)^{-\frac{1-\rho}{\sigma^2}} \\ &\times \left(R(y_2)\left[S_B(x_2)^{-\sigma^2} - 1\right] + 1 \right)^{-\frac{1-\rho}{\sigma^2}} \\ &\times \left(R(y_1)\left[S_B(x_1)^{-\sigma^2} - 1\right] + R(y_2)\left[S_B(x_2)^{-\sigma^2} - 1\right] + 1 \right)^{-\frac{\rho}{\sigma^2}} \end{aligned} \tag{3}$$

where $S_B(x)$ is the survival function for the individuals with $R(y) = 1$. A gamma-Gompertz parametric form of $S_B(x)$ was used in the analysis:

$$S_B(x) = \left(1 + s^2 \frac{a}{b}\left(e^{bx} - 1\right)\right)^{-\frac{1}{s^2}} \tag{4}$$

It was verified, by comparing the empirical and the theoretical survival functions visually, that this parameterization fitted the data well.

The bivariate correlated frailty model was then fitted to the data (including the censored observations) using a maximum likelihood estimation procedure for each sex and zygosity separately and the parameters of the model where estimated. The estimates are presented in Table 6.

Table 6 Parameter estimates (SE) and model fit index for the bivariate correlated frailty model by sex and zygosity

Sex	Zyogsity	a	b	s^2	σ^2	ρ_{mz}	ρ_{dz}	α_1	α_2	β_1	β_2	β_3	LogLik
Males	MZ	9.2E-07 (1.2E-06)	0.1 (0.0)	0.00 (0.00)	1.13 (0.35)	0.70 (0.18)		1.23 (0.45)	1.19 (0.33)	4.75 (2.18)	2.58 (1.13)	1.45 (0.63)	–1150.6681
	DZ	5.3E-06 (7.5E-06)	0.1 (0.0)	0.54 (0.25)	0.86 (1.75)		0.46 (0.16)	1.01 (0.28)	1.18 (0.51)	1.52 (0.78)	1.05 (0.30)	1.19 (0.84)	–1756.0938
Females	MZ	4.7E-06 (3.1E-06)	0.1 (0.0)	0.00 (0.00)	0.91 (0.57)	0.77 (0.54)		1.36 (0.37)	1.65 (0.48)	1.67 (0.39)	1.39 (0.39)	1.08 (0.31)	–1243.8988
	DZ	5.5E-07 (4.8E-07)	0.1 (0.0)	0.00 (0.00)	1.95 (0.34)		0.15 (0.09)	1.92 (0.70)	2.98 (1.09)	3.30 (1.29)	2.50 (1.13)	2.16 (0.79)	–1867.0738

Table 7 Individual relative risk of mortality by 1994 stratified by sex, smoking and BMI in a sample of 2464 Danish twins. Computation of relative risk was based on the model in which frailty *z* is equal to 1 (see Appendix for further details)

		Non-smokers	Others (ex-smokers, etc.)	Current pipe/ cigar smokers	Current cigarette smokers
Males	BMI < 22	1.08	1.36	1.52	2.40
	BMI 22–28	1.00	1.26	1.41	2.22
	BMI > 28	1.23	1.55	1.73	2.73
Females	BMI < 22	1.63	2.38	2.79	3.68
	BMI 22–28	1.00	1.46	1.71	2.26
	BMI > 28	2.26	3.30	3.86	5.11

For males there was no significant difference between the βs suggesting no increasing risk with BMI. Table 7 shows the relative risk for individuals in the 12 subgroups whose frailty is equal to 1. The relative risk at population level is probably less than the individual relative risk, because mean *z* decreases as the population grows older [the more frail (with a large *z*) tend to die first].

For each pair of twins, where at least one of the survival times was censored an appropriate conditional survival model given all the available survival and covariate information was constructed using standard probability calculus. A random sample of size 10 was drawn from this conditional model for each censored individual. The above-described procedure yielded 10 new data sets which were completely uncensored and where censored survival times were replaced by imputations. These data sets were subsequently used in the genetic analysis.