HOW MUCH CAN THE INCREASED MORTALITY IN SMOKERS BE EXPLAINED BY OTHER FACTORS?

PETER LEE[1], JOHN FRY[1] and JAN HAMLING[1]

[1]P.N.Lee Statistics and Computing Ltd., 17 Cedar Road, Sutton, Surrey, SM2 5DA, England

Telephone : (44) 181 642 8265

Fax : (44) 181 642 2135

E-mail : PeterLee@pnlee.demon.co.uk

Author for correspondence : Peter Lee

Sponsor : British-American Tobacco Company Ltd

Short title : EXPLAINING THE INCREASED MORTALITY IN SMOKERS

SUMMARY

Many lifestyle risk factors are commoner in smokers, but studies of smoking and mortality often adjust only for age. Based on 10-year follow-up of 4189 British adults aged 45+ (1137 deaths) we studied the joint association of smoking and 36 other risk factors with mortality. These risk factors seem to explain about 20% of the excess mortality seen in heavy and moderate current cigarette smokers and recent ex-smokers. For some of the risk factors, associations with mortality are largely explicable by other factors. Nevertheless about half the risk factors were independently associated with mortality. Although current smokers of 20+ cigarettes/day had higher adjusted mortality rates than never smokers (relative risk 2.31, 95% limits 1.70-3.15), the joint effect of other factors may be greater still. Never smokers with all of four risk factors strongly independently associated with mortality had 18.1 times the mortality (95% limits 5.28-61.9) of those with one or none.

1. INTRODUCTION

Numerous epidemiological studies have reported an increased overall risk of mortality in smokers[1]. Smokers are not randomly selected from the population, so the possibility has to be borne in mind that some of the increased mortality may result, not from a direct effect of smoking, but from differences between smokers and non-smokers in respect of other factors. For some diseases associated with smoking, the possibility of important confounding is strong - a recent review by Doll[2] lists cancers of the liver, cervix uteri and large bowel, cirrhosis of the liver, suicide and poisoning as such diseases. However, results from the major prospective studies linking overall mortality to smoking typically present relative risks adjusted only for age and not for any other potentially confounding factor[3]. Because of this, such attempts to estimate mortality from smoking that have considered the possibility of confounding at all have used essentially arbitrary corrections for it[4], not basing their findings on analyses in which formal statistical adjustment had been made for a wide range of relevant risk factors.

In 1994 we presented results of analyses from the Health and Lifestyle Survey[5], a representative sample of 9003 British adults in which extensive data were collected in 1984-1985. We compared the distribution of 33 lifestyle factors generally considered to be associated with adverse health between current smokers, ex-smokers, never smokers living with a smoker and other never smokers. 27 of the 33 risk factors showed a significantly higher prevalence in heavy smokers than in never smokers and only two showed a lower prevalence. For many risk factors, prevalence increased with amount smoked, decreased with time of smoking cessation and was increased in passive smokers.

Given various assumed values for the strength of the relationship of a risk factor to mortality, theoretical calculations were presented showing to what extent failure to take the risk factor into account in analysis might bias the smoking-mortality relationship. The possible confounding effect from multiple risk factors was also discussed.

Since the time that report was prepared, data resulting from mortality follow-up of the Health and Lifestyle Survey subjects to November 1995 have become available. Though there are too few deaths in never smokers to allow adequate analysis of possible confounding in the study of passive smoking, there are sufficient deaths to investigate the extent to which the association between active smoking and mortality is modified by statistical adjustment for the other factors studied. We report the results of this investigation.

2. METHODS

The Health and Lifestyle Survey was conducted in three stages; a questionnaire completed during an interview at the subject's home, a home visit by a nurse, and a self-completion questionnaire to assess personality and psychiatric status. Fuller details of the survey are given elsewhere[5,6].

Mortality follow-up was conducted through the National Health Service Central Register in Southport, death certificates being available for 1183 of the original 9003 subjects. Although there were 13 additional subjects who had been reported as dead but with no trace found at Southport or who had been registered as dead but with no death certificate found, and although there were 539 untraced subjects, our main analyses take no account of this, simply comparing the probability of dying with death certificate available by original risk factor status.

The relationship of risk of mortality to smoking and other risk factors was studied by logistic regression analyses in which the probability (p) of subject j dying was modelled by the formula:

$$Y_j = \log(p_j/(1-p_j)) = \beta_0 + \sum_{i=1}^{n} \beta_i x_{ij}$$

where x_{ij} (i=1,...n) are known values of n regressor variables and β_j are coefficients to be estimated. Where a regressor variable is a presence/absence variable, e^{β} estimates the relative risk of mortality from that variable.

Logistic regression analysis is facilitated if the variables considered do not have 0% (or 100%) mortality at any level. To avoid this problem subjects aged 18-44 at the start of the study were omitted from analysis. This resulted in a loss of only 46 (3.9%) of the total deaths in the study and should have little effect on the conclusions. All analyses were adjusted for sex and age at the start of study (using the age groups 45-59, 60-64, 65-69, 70-74 and 75+) by including a variable with 10 levels representing all combinations of sex and age group.

Following preliminary analysis to identify smoking groups with adequate numbers of deaths and differing mortality, smoking was introduced into the model as a six-level variable:- never smoked, ex-smoker given up 0-9 years ago, other ex-smoker, current smoker of 20+ cigarettes a day, current smoker of 10-19 cigarettes a day and other current smoker. Thirty-six other risk factors were included in the analyses. All the factors considered in our paper[5] were included, with the exception of vital status of the father, there being too few deaths in those aged 45+ with the father still alive to allow reliable estimation. Five additional risk factors were included based on preliminary univariate analyses of mortality, these being three dietary factors, indicating consumption of soft drinks, nuts and cream, and two psychosocial factors, bored and lonely. The levels of the 36 risk factors can be seen in Tables III and IV. Factors measured on a continuous or semi-continuous variable have been categorized into no more than six levels. A fuller description of most of the factors considered is given elsewhere[5]. For fried foods, fruit, vegetables, salads and sweet food consumption, scores were obtained by summing answers from a 5-point frequency scale (0 = never, 1 = less than once a week, 2 = once or twice a

week, 3 = most days (3-6), 4 = once a day, 5 = more than once a day) for appropriate questions (e.g. for fruit consumption the answers relating to fresh fruit in summer, fresh fruit in winter and fruit juice were summed).

Data on body mass index and cutting down on fatty foods (obtained by the nurse at home visits) and on neuroticism, extroversion and type A personality (obtained from the self-completion questionnaire) and also on household income had missing values relatively commonly, so "missing" was included as a level for these risk factors. For other risk factors (obtained from the original questionnaire), missing values were quite rare and it was decided (to avoid problems with zero death rates) to restrict attention to subjects with complete data on all of them. This slightly reduced the number of subjects aged 45+ considered, from 4391 to 4189, and the number of deaths, from 1137 to 1075, but still allowed adequate numbers of deaths for analysis.

The modelling process was performed in several ways to ensure that each variable was assessed for its effects on the model. A stepwise regression modelling procedure involved first incorporating the age/sex variable, then the smoking variable, then the risk factor found to make the most statistically significant difference to the model, then the risk factor making the next most significant difference, and so on until a model was reached for which none of the remaining risk factors would make a difference that was statistically significant at $p<0.1$. This is referred to as our 'stepwise model'. Other modelling procedures included the fitting of:

(Model A) age/sex plus each risk factor individually (to find the effect of each risk factor on the age-sex model),

(Model B) age/sex and smoking plus each risk factor individually (to find the effect of each risk factor on the age-sex-smoking model),

(Model C) age/sex, smoking, and each risk factor in our stepwise model, then the inclusion of each other risk factor individually,

(Model D) age/sex, smoking and our stepwise model, then the exclusion, in turn, of each risk factor in the stepwise model, and

(Model E) age/sex and smoking plus all the risk factors regardless of their statistical significance (the 'total' model).

Additional logistic regression analyses were carried out,

(a) omitting six risk factors where an association with mortality seemed particularly likely to arise because of the effects of illness on the risk factor rather than the effects of the risk factor on mortality. These risk factors were work status, body mass index, hours slept, bored, lonely and depression/nervous illness,

(b) omitting those 315 subjects, among whom 131 died, who reported their own health in general as poor, and

(c) omitting the six risk factors mentioned in (a) and the 315 subjects mentioned in (b).

When presenting results of the logistic regression analyses, relative risks, 95% confidence levels and the significance of the differences are given, comparing each exposure level with a base exposure level for the factor. Probability (p) values are

presented as significances with +++, - - - representing $p<0.001$; ++, - - $p<0.01$; +, - $p<0.05$; (+), (-) $p<0.1$ and N.S. representing not significant, with plus signs indicating an increased mortality and minus signs a decreased mortality. Also presented, as appropriate, are

(a) the deviances (differences in sum of squares) associated with introducing or removing factors from a model, the deviance being distributed approximately as chisquared on n-1 degrees of freedom, where n is the number of levels of the factor concerned, and

(b) the % excess risk explained by adjustment, as calculated by $100\ (R_2-R_1)/(R_2-1)$, where R_1 is the relative risk after adjustment and R_2 is the relative risk before adjustment.

3. RESULTS

Mortality by age and sex

Table I shows the number of subjects and deaths by age and sex. Overall there were 1075 deaths (25.7%) among the 4189 subjects. As expected, mortality clearly increased with increasing age and was higher in men than in women.

Logistic regression analyses involving smoking and 36 risk factors

Table II shows that, after adjustment for age and sex, there was a clearly increased mortality in smokers. This was most marked for current smokers of 20+ cigarettes a day, current smokers of 10-19 cigarettes a day and ex-smokers who had given up in the last 10 years. However, mortality was also significantly increased in longer term ex-smokers and in other current smokers (which included smokers of <10 cigarettes a day and smokers of pipes and/or cigars only).

17 of the 36 risk factors considered were included in the stepwise model (the model starting with age, sex and smoking only, then adding risk factors selected in the order of the most significant effect on the model, selection ending when no further significant effect was made by adding any remaining risk factor - see Methods). With the exception of "other ex-smokers", adjustment for these 17 factors reduced the relative risk associated with smoking, with between 24.9% and 28.7% of the excess risks explained by these factors.

Further including all the 36 risk factors in the regression analysis (Model E) did not explain more of the risk associated with smoking. In fact, it tended somewhat to increase the relative risks associated with current smoking as compared with those estimated in the stepwise model.

Table III presents information regarding the 17 risk factors included in the stepwise model showing, by level of each factor, the numbers of subjects and deaths, the relative risks and deviances adjusted for age and sex only (using Model A) and the relative risk and deviances from the stepwise model itself. Factors are shown in the order in which they were introduced into the stepwise model (based on significance). As can be seen, this order is not necessarily the same as that based on deleting the risk factor from the stepwise model.

Adjustment for smoking and the other risk factors in the model generally tended to reduce relative risks associated with specific risk factors. This was particularly evident for bored, no educational qualifications, social class, depression/nervous illness and salad consumption where the deviance explained by the factor in Model A (after adjustment for age and sex only) was over 4 times the deviance explained by the factor in Model D (after adjustment for age, sex and the other factors in the stepwise model). Exceptionally, risky occupation and ever tried to lose weight, not significant (at $p<0.1$) after adjustment for age and sex only, became significant when the other risk factors were included.

Table IV presents similar information for factors not included in the stepwise model. For many of these factors the analyses adjusted for age and sex showed statistically significant variation. This was particularly evident for household income, fruit consumption, nuts, cream, neuroticism and lonely, all of which had an age/sex adjusted deviance which was statistically significant at $p<0.001$ but which was not significant at $p<0.1$ when adjustment was made for the other factors in the stepwise model.

Risk factors and mortality in current smokers of 20+ cigarettes a day

Based on the results of our analyses, an attempt was made to determine which specific risk factors contributed most to the effect of adjustment on the smoking relative risks noted in Table II. We concentrated on the relative risk associated with current smoking of 20+ cigarettes/day (compared with never smoking) where the relative risk adjusted for age and sex only, of 2.818, was reduced to 2.315 when the 17 risk factors included in the stepwise model were adjusted for. Table V shows:

(I) for all 36 risk factors, the relative risk (of smoking 20+ cigarettes/day) after adjustment for age, sex and the risk factor (Model B), and the excess risk explained by adjustment for the specific factor (e.g. for bored, excess risk explained = 100*(2.818-2.597)/(2.818-1) = 12.2%).

(ii) for the 17 risk factors included in the stepwise model, the relative risk (of smoking 20+ cigarettes/day) after adjustment for age, sex and all factors except the specific factor (Model D), and the excess risk explained by adjustment for that factor (e.g. for bored, excess risk explained = 100*(2.362-2.315)/(2.362-1) = 3.5%).

(iii) for the 19 risk factors not included in the stepwise model, the relative risk (of smoking 20+ cigarettes/day) after adjustment for age, sex all factors in the stepwise model and the specific factor (Model C), and the excess risk explained by adjustment for the specific factor (e.g. for household size, excess risk explained = (2.315-2.320)/(2.315-1) = -0.4%).

From the results in Table V it can be seen that the effect of adjustment for specific risk factors on the age/sex adjusted risk for current smoking of 20+ cigarettes/day was nearly always to reduce it. Thus, of the 36 risk factors considered 26 explained 1% or more of the excess risk, 11 explained 5% or more of the excess risk and 3 (bored, salad and fruit) explained 10% or more. In contrast, for only 3 of the risk factors (alcohol, coffee and ever tried to lose weight) did adjustment *increase* the age/sex adjusted risk for current smoking of 20+ cigarettes/day, and then by a relatively small amount, with no more than -3.8% of the excess risk explained.

Were the effects of adjustment independent, it is clear that a very large part of the age/sex adjusted relative risk of 2.818 for current smoking of 20+ cigarettes/day could have been explained by the risk factors considered jointly. (Adding the excess risks explained by the 17 factors included in the stepwise regression model gives a figure as high as 67.4%.) However, the further results presented in Table V, and the fact that adjustment for the 17 factors only actually explained 27.7% of the excess risk, shows clearly that the effects of adjustment by the individual factors are not independent. For many of the risk factors, the effect of adjustment for that risk factor (on the relative risk for current smoking

of 20+ cigarettes/day) was lower when the other factors in the stepwise model were taken into account than when only age and sex were. For factors in the stepwise model, this was particularly evident for bored, no educational qualifications, work status, social class and salad consumption, for all of which the percentage excess risk explained reduced by 4% or more. The factors which most affected the relative risk (for current smoking of 20+ cigarettes/day) in spite of adjustment for the other factors were salad consumption and body mass index, adjustment for which reduced the relative risk by respectively 0.12 and 0.11, and coffee consumption, which increased the relative risk by 0.11.

Alternative logistic regression analyses omitting six risk factors which may be affected by illness

Some analyses were rerun omitting six risk factors (body mass, work status, hours slept, depression/nervous illness, bored and lonely) where the association with mortality may partly have arisen, not because the risk factor affects mortality, but because the illness leading to death affects the risk factor.

The new stepwise model now included 13 of the risk factors. Of the 17 in the original stepwise model, four were omitted risk factors and three (social class, never tried to lose weight and salad consumption) were not included due to lack of significance. There were three factors (household income, neuroticism and nuts) in the new model that were not in the original model.

Table VI shows the effect of adjustment on mortality associated with smoking when the six risk factors were omitted. Compared with the analyses in Table II (when the six factors were not omitted) it is evident that less of the excess risk is explained in the new stepwise model. The excess risks explained by all the factors considered, for the smoking categories with the largest relative risks (current 20+ and 10-19 cigarettes/day and ex-smoker <10 years ago), were about 20 to 25% lower in Table VI than in Table II. This suggests that some of the excess risk explained in Table II may be an artefact of over-adjustment. It should be noted however that, unlike Table VI, adjustment in Table II for those risk factors not selected by the stepwise process generally *reduced* the excess risk explained.

Alternative logistic regression analyses omitting subjects in poor health at time of interview

Analyses were also rerun omitting subjects who defined their health as poor. The stepwise model for the remaining subjects included 17 risk factors, 16 the same as in the original model, with neuroticism now included and social class excluded. As can be seen by comparing Table VII with Table II, the smoking relative risks adjusted for age and sex are similar to those when no such omission was made, but the excess risks explained by the stepwise model become slightly less (e.g. 23.8% vs 27. 7% for current smoking of 20+ cigarettes/day). The excess risks explained by all 36 factors were, however, not very different.

Analyses were also rerun both omitting the six risk factors which were thought to be possibly affected by ill-health and omitting the subjects who defined their health as poor. When all the other factors were included, adjustment explained some 15% or so of the excess risk associated with current smoking of 20+ or 10-19 cigarettes/day or with recent ex-smoking, 7% of the excess risk in other current smokers and none of the excess risk in other ex-smokers - see Table VIII.

Investigation of interactions

Although the analyses described have adjusted for the possible interaction between age and sex, no other interactions between the variables studied have so far been taken into account. It would not be feasible to study all possible interactions, but some analyses were conducted to investigate some of the more important ones.

First, we looked at the interactions of our six level smoking variable with age and with sex. The interaction with age was not at all significant (χ^2 = 18.40 on 20 d.f., p = 0.56) but there was a suggestion of a possible interaction with sex (χ^2 = 10.97 on 5 d.f., p = 0.052). Table IX shows the relative risks by smoking for the two sexes. As can be seen, the association with current smoking of 20+ cigarettes/day was very similar in the two sexes, the major cause of the interaction being the relatively high risk in males for current smokers of 10-19 cigarettes/day, not evident in females. In view of the lack of statistical significance of this interaction, its relatively small size compared to the main effects of smoking and sex, the likelihood that some of it was due to chance (it seems unlikely the risk in males is actually higher for 10-19 than 20+ cigarettes/day), and the smaller power of sex-specific analyses, it was decided not to rerun all our earlier analysis on a sex-specific basis.

Second, for various selected risk factors, we looked at the effect of introducing interactions between the risk factor and age, sex and smoking in models which included the main effects of age, sex, smoking and the risk factor and the interaction between age and sex. We chose for analysis the four risk factors most strongly associated with mortality

in Table III which were not those particularly likely to be affected by illness. For death of mother, alcohol consumption and type A personality there was no evidence ($p>0.1$) of any interaction with age, sex or smoking. For no educational qualifications, there was no indication of an interaction with sex or smoking, but there was a significant interaction with age ($\chi^2 = 10.59$ on 4 d.f., $p = 0.03$). As noted in Table V, adjustment for this variable reduced the relative risk associated with current smoking of 20+ cigarettes a day from 2.818 to 2.680, so explaining 7.6% of the excess risk. Further adjustment for its interaction with age reduced the relative risk a little more, to 2.645, so that 9.4% of the excess risk was now explained.

Joint effects of risk factors

Using the same four risk factors, and defining a score of one for those with levels indicating a markedly increased risk of mortality (based on the results in Table III), we estimated the relative risk of mortality, adjusted for age and sex, in relation to the number of risk factors present. As can be seen from Table X, in both never smokers and current smokers, risk increased steeply with the total number of risk factors present. Compared with those with 0 or 1 risk factors, those with all 4 risk factors had almost a 5-fold increased risk in current smokers (RR = 4.89, 95% CI = 2.57-9.31) and an even higher risk in never smokers (RR = 18.1, 95% CI = 5.28-61.9), though the latter estimate had a large standard error, being based on only 23 deaths.

4. DISCUSSION

In our previous analyses on the Health and Lifestyle Survey[5] we demonstrated that current smokers, compared to never smokers, had an increased prevalence of a wide variety of risk factors, with ex-smokers intermediate for many. We discussed in depth the likely implications of these differences in terms of the effect they may have in confounding relationships of smoking with mortality. However lack of available mortality data at that time limited the basis of our conclusion "that confounding by multiple risk factors may be an important issue in smoking studies where weak associations are observed." Now that mortality data are available from the first 10-11 years of follow-up of the population, we have attempted to come to more reliable conclusions.

In order to obtain adequate numbers of deaths, and to avoid problems of diagnostic error associated with determining cause of death from death certificates, we decided in the first place to limit attention to all cause mortality. Also, as the association with smoking was relatively similar in the two sexes, we decided to present our main results for the sexes combined (though always adjusted for sex). Deaths were relatively uncommon under age 45 so we restricted attention to those 4189 subjects of age 45+, in which a total of 1075 deaths occurred. Our main concern was to investigate possible confounding effects of a large number of risk factors, which entailed a large amount of computing, and not to obtain absolutely precise estimates of relative risk. We therefore decided to carry out logistic regression analyses based on probability of death over the period and not use more complex techniques which took into account when the deaths occurred in the period. Limited analyses adjusting for time of death, and experience in other studies where

analyses have been carried out both ways, suggests the conclusions are not likely to depend crucially on the choice of technique to determine relative risk.

Our analyses showed that, after adjusting for age and sex, there was a clearly increased mortality in current and ex-smokers as compared to that in never smokers. As expected, the increase was highest in current heavy (20+/day) cigarette smokers, where the relative risk was 2.82 (95% CI 2.13-3.73), was next highest in current moderate (10-19/day) cigarette smokers and in recent (<10 years ago) ex-smokers, and least in long-term (10+ years ago) ex-smokers and in current light (1-9/day) cigarette smokers and smokers of pipes and cigars. The increase was, however, statistically significant in each smoking group.

In attempting to estimate, from the data available in the Health and Lifestyle Survey, the extent to which confounding might explain the increased relative risk associated with smoking, a number of problems arise, not all of which have been addressed in our analyses. These problems are considered in turn below.

First, although data on over 1000 variables were collected in the study, information was not available on a number of factors known to be associated with health, many of which are also associated with smoking. These include exposure to specific chemicals and to sources of radiation; use of drugs; sexual activity; hormone-related factors such as menopausal status, use of oral contraceptives and hormone replacement therapy; region of birth; methods of heating and cooking; and intake of specific vitamins and antioxidants

(since only dietary frequency data were available). Clearly it is possible that omission of some of these factors may lead to underestimation of the extent to which confounding may explain the association between smoking and mortality. Additional analyses based on studies with an even wider range of risk factors would clearly be valuable here.

Second, the analyses conducted in this study take no account of any errors there may be in determining smoking habits or the risk factors considered. In principle, random errors in determining smoking are likely to underestimate the true association of mortality with smoking, while random errors in the other risk factors are likely to underestimate the true extent of possible confounding. Given that the effect of other risk factors is to reduce the associations with smoking (as we observed), the effect of these two types of error are likely to act in opposing directions. In the absence of data from the study on errors in the variables determined, and also bearing in mind the likelihood that some errors will not be random and will be associated with health status and the level of other variables, we decided at this stage to ignore this problem in analysis and rely on the data as recorded.

The third problem is choosing which risk factors to consider as potential confounding variables from the large number for which the Health and Lifestyle Survey provides information. Regression analyses using many hundreds of risk factors would not have been a feasible proposition. When selecting risk factors we used a number of criteria, amongst which were inclusion of variables commonly considered as risk factors in smoking-related diseases, inclusion of variables commonly considered to be part of a healthy lifestyle, avoidance of variables not generally considered to be risk factors for

smoking-related diseases and combination of related variables into single indices providing initial analysis had shown similar patterns of relationship to smoking. We also extended the list of factors somewhat from that used in our earlier paper[5], based on preliminary univariate analyses relating mortality to a longer list of risk factors. Using these selection criteria we ended up with a list of 36 risk factors, as can be seen in Tables III and IV. Clearly other choices of risk factors might have somewhat affected our estimates.

The fourth problem is whether to include only independently statistically significant risk factors in the model, or to include all the risk factors. While it is quite common practice to include in the model only those factors that are significant or near significant, statisticians often warn that confounding may occur from failure to adjust for relevant factors, even if they are not significant[7]. We decided to approach this problem in two ways; by conducting a stepwise logistic regression in which additional factors were included as long as they were significant (as judged by a likelihood-ratio test) at $p<0.1$, and by conducting a logistic regression including all the risk factors selected. Based on the 36 risk factors selected, 17 were included as independent risk factors in the stepwise logistic regression. The effect they had on most of the smoking relative risks was to reduce the excess risk by about 25%. Exceptionally, the relative risk for ex-smokers who had given up 10 or more years ago was little affected. This was unsurprising in view of our earlier analyses[5] showing that, in terms of risk factor prevalence, long-term ex-smokers were quite similar to never smokers. Adding in all the other 19 risk factors in the analysis did not, in practice, explain any more of the excess risk associated with smoking - in fact is explained slightly less.

A fifth problem is whether to include interactions in the analyses. We did not do so, partly as the number of interactions that could be studied is so large, and partly because limited analyses involving some of the risk factors that were most strongly associated with mortality showed that interactions of the risk factors with smoking were always non-significant and that interactions of the risk factors with age and sex were usually non-significant and had little or no effect on the smoking relative risks. However, we noted one exception, an interaction between having no educational qualifications and age, adjustment for which slightly reduced the estimated risks associated with smoking. Though there may be additional interactions in our data that are statistically significant, we feel that analyses using models involving only main effects are likely to demonstrate the major part of the effect of confounding.

A more serious problem lies in the difficulty in deciding whether or not particular risk factors should be considered as potential confounding variables in analysis. In our context, an association between a risk factor and mortality may, in principle, arise for one or more of three main reasons: (1) because smoking affects prevalence of the risk factor either directly, or indirectly via an effect on health; (2) because of an effect of the risk factor on mortality that is independent of smoking; and (3) because the risk factor may be associated with another risk factor (or factors) which has an effect on mortality that is independent of smoking (i.e. the risk factor of interest may be a marker for a true risk factor).

In the first case, adjustment for the factor in analysis should not be carried out as it will lead, by "over-adjustment", to understatement of the true association of smoking with mortality. In the second case, adjustment should be carried out to avoid confounding. In the third case, there is no problem if one includes the other risk factor in analysis. If the other factor with which it is associated (the true cause) is already in the analysis, including the marker risk factor should have no effect on the estimate of risk (unless the marker is measured much more accurately than the true risk factor). If it is not, including the marker risk factor will partly correct for the confounding effect of the true risk factor.

However a severe difficulty lies in the fact that one does not always know, for a particular risk factor, which of the three cases applies. Furthermore, the association between a risk factor and mortality may arise partly for one of the three reasons and partly for another. Thus, for example, among the healthy, taking no exercise may really have long-term adverse consequences on health, but equally part of the association of lack of exercise with increased mortality may arise because certain illnesses caused by smoking both limit the ability to take exercise and increase the risk of mortality.

While the results in Table II may indicate the proportion of smoking associated excess mortality that can statistically be explained by the risk factors in the model, it is likely that, to some extent, they may overstate the true confounding effect of these factors. In an attempt to avoid over-adjustment we carried out various alternative analyses, using one or both of two devices. One device was to exclude six risk factors for which at least some of the association between the risk factor and mortality seemed particularly likely to

arise because of the effects of smoking on the risk factor. These included body mass, where smoking is known to have an effect both directly and indirectly via effects on illness, and five factors, such as bored, which seemed likely to be affected by illness which is itself affected by smoking. The other device was to omit from analysis those who reported at interview that their health was poor, on the basis that such people are particularly likely to have had their risk factors modified by their health status.

The results of these analyses, presented in Tables VI to VIII, still showed that adjustment for the risk factors reduced the relative risks associated with smoking (particularly in heavier current smokers and in more recent ex-smokers), but to a lesser extent than seen in Table II. However, these analyses may well overcompensate for the problem of "over adjustment," inasmuch as some of the association of the excluded risk factors with mortality may arise independent of smoking.

The analyses presented also allow a number of conclusions to be drawn about the role of the various risk factors. One conclusion is that, although smoking is strongly associated with an increased risk of mortality, other risk factors are also important and their joint effect may multiply risk more in never smokers than does smoking in smokers. This can be seen from the results in Table X where, among never smokers, those subjects with all of four risk factors strongly associated with mortality (lack of education, mother dead, ex-drinkers (moderate or heavy) of alcohol, and not having a Type A personality) had an estimated 18.1-fold increase in risk of death (95% CI 5.28-61.9) compared to those with 0 or 1 risk factors. Among those with 0 or 1 of the four risk factors, there was a 5.2-fold

increase in the risk of death for current smokers compared with never smokers (95% CI 2.24-12.21).

A second conclusion is that the confounding effect of multiple risk factors on the smoking/mortality association is considerably less than would be expected from combining the confounding effects of the risk factors considered individually. One reason for this is that some of the associations between risk factors and mortality measure the same true underlying relationship. The joint effect of education, income and social class, for example, is much less than the sum of their individual effects. Another reason is that some of the risk factors considered proved not to be independent predictors of mortality at all, or to have adjusted associations that are very much weaker than the unadjusted associations. Some examples of these can be seen in the following brief comments on the results for specific risk factors:

Mother dead. This risk factor had an independent and strong association with mortality, but was not strongly associated with smoking so had little confounding effect.

No educational qualifications. A strong association with mortality was weakened but remained significant after adjustment for other risk factors. Once the other factors were in the model, however, it explained little of the increased mortality in current heavy smokers.

Marital status. The increased risk of mortality in those who had never been married was little changed by adjustment, but adjustment for this factor had little effect on the relationship of smoking with mortality.

Household size. This factor showed no clear relationship with mortality.

Work status This lost some of its significance when the other factors were introduced into the model. It explained little of the association of smoking with mortality.

Social class This was quite strongly related to mortality and retained some of its association with smoking even after adjustment for the other factors.

Risky occupation This became significant only after the other risk factors were introduced into the model.

Household income This was strongly associated with mortality in age and sex adjusted analyses but lost its significance in the model when the other factors were included.

Alcohol consumption Those who gave up drinking after being moderate or heavy drinkers were at increased risk, but this factor explained none of the smoking association.

Anything to keep healthy The increased mortality associated with doing nothing to keep healthy became non-significant when the other factors were included in the model.

Get enough exercise Although associated with mortality this factor was little associated with smoking.

Ever tried to lose weight This factor was significant only after other factors had been included in the model. It did not help to explain the association between smoking and mortality.

Depression/nervous illness This retained some significance after the other factors were included in the model and explained some of the association with smoking (5% of the increased mortality in current heavy smokers before adjustment and 1.5% after adjustment).

Hours slept The increased mortality in those who sleep for nine or more hours a day became non-significant after adjustment for other factors. It did not help to explain the increased mortality in smokers.

Time before first meal in day When other factors were adjusted for, the increased mortality associated with increasing time to first meal essentially disappeared.

Fried foods This had little association with mortality, especially after other factors were adjusted for.

Breakfast cereal consumption This was negatively associated with mortality and explained some of the association of mortality with smoking, even after adjustment for the other factors (it explained 7.8% of the increased mortality in heavy smokers before adjustment and 4.0% after adjustment).

Slices of bread (per day) High consumption levels were associated with some increase in mortality but this factor did not explain any of the increased mortality in smokers.

Fruit consumption In analyses adjusted only for age and sex, fruit consumption was strongly negatively associated with both mortality and smoking. However, when the other factors were introduced into the model, the negative association with mortality largely disappeared and the factor no longer confounded the relationship of smoking with mortality.

Vegetable consumption As for fruit consumption, adjustment for other risk factors essentially eliminated the negative association with mortality.

Salad consumption A strong protective effect was evident in analyses adjusted only for age and sex. This weakened but remained significant when the other factors were introduced. Because smokers eat salads much less often than non-smokers, this factor explained some of the association with smoking (16.0% of the excess risk in current heavy smokers before adjustment and 8.5% after adjustment).

Sugar in hot beverages (teaspoons) Weak associations of this factor with mortality disappeared when the other factors were adjusted for.

Cups of tea drunk Little association with mortality was evident.

Cups of coffee drunk Even after adjustment for other factors, increased coffee consumption was associated with significantly reduced mortality. As smokers drink more coffee than non-smokers, failure to adjust for coffee consumption leads, in these data, to underestimation of the relationship between smoking and mortality.

Sweet food consumption A reduced mortality in heavy consumers of sweet foods became non-significant when other factors were adjusted for.

Uses low fat/polyunsaturated spread on bread After adjustment for other factors, this factor also had little association with mortality.

Body mass index Being underweight was strongly associated with mortality even after adjustment for other factors, and adjustment for body mass index explained 7.9% of the increased mortality in current heavy smokers. However, underweight may to some extent reflect pre-existing ill health.

Cut down on fatty foods Not cutting down had no effect on mortality when the other factors were included in the model.

Neuroticism An increased mortality in those with higher neuroticism scores largely disappeared when the other factors were included in the model.

Extroversion This had no association with mortality and so no effect on the model.

Type A personality This was negatively associated with mortality but explained none of the smoking risk once the other factors were introduced.

Bored This was strongly associated with mortality and explained some of the association with smoking even when the other factors were introduced (12.2% of the excess risk in heavy smokers explained in the age-sex only model, and 3.5% explained when the other factors were introduced).

Lonely A markedly increased age and sex adjusted mortality in those who were often or always lonely largely disappeared after adjustment for the other factors.

Soft drinks An increased mortality in those who frequently took soft drinks was largely independent of the other factors. However, it had little confounding effect on the smoking/mortality association.

Nuts In age and sex adjusted analysis, consumption of nuts was negatively associated with mortality and explained some of the relationship between smoking and

mortality. However, these effects were substantially reduced when the other factors were introduced.

Cream The pattern of results was very similar to that for nuts, the lower mortality in frequent consumers of cream being mainly explained by adjustment for other factors.

Although the analyses presented here provide valuable information about the relationship of a variety of risk factors to mortality, the main question we sought to answer was "how much can the increased mortality in smokers be explained by other risk factors?". Our analyses in Tables II and VIII, taken together, suggest that allowance for the risk factors recorded in this study can explain perhaps 20% or so of the excess mortality associated with smoking in the smoking groups where risk is most elevated - current moderate and heavy cigarette smokers and recent ex-smokers. Bearing in mind the fact that the Health and Lifestyle Study did not record a number of relevant risk factors and the possibility of residual confounding due to errors in risk factors that were considered, this figure is probably an underestimate. However it is our opinion that the true figure is not greater than 30-40%.

Further insight into the question can be gained by larger studies looking at additional risk factors and analyses systematically investigating mortality by cause.

ACKNOWLEDGEMENTS

We thank the Health and Lifestyle Survey at the University of Cambridge and the ESRC Data Archive for permission to use the data from the Health and Lifestyle Study. We alone bear the responsibility for the further analysis and interpretation of these data. We also thank Mrs P J Wassell and Mrs D P Morris for their assistance in the typing of this manuscript. Financial support was provided by British-American Tobacco Company Ltd, to whom we are also grateful.

Table I. Number of subjects and deaths by age and sex

Sex	Age	Number of subjects	Number of deaths	% dead
Male	45-59	872	113	13.0
	60-64	310	75	24.2
	65-69	214	86	40.2
	70-74	216	122	56.5
	75+	221	162	73.3
	Total	1833	558	30.4
Female	45-59	1092	78	7.1
	60-64	396	70	17.7
	65-69	302	67	22.2
	70-74	254	107	42.1
	75+	312	195	62.5
	Total	2356	517	21.9

Table II. Effect of multivariate adjustment on mortality associated with smoking

			Relative risk (95% CI)			RR(95% CI)	
	No. of subjects	No. of deaths	Adjusted for age and sex	Also adjusted for 17 factors†	Excess risk* explained(%)	Adjusted for all 36 factors	Excess risk* explained (%)
Never smoked	1406	295	1.00 (base)	1.00 (base)		1.00 (base)	
Ex-smoker <10 years ago	433	119	1.95 (1.45-2.62)	1.71 (1.26-2.34)	24.9	1.71 (1.24-2.35)	25.7
Other ex-smoker	833	261	1.51 (1.19-1.92)	1.54 (1.20-1.98)	-5.7	1.50 (1.16-1.94)	1.7
Current smoker of 10-19 cigs/day	463	122	2.28 (1.71-3.04)	1.92 (1.41-2.61)	28.2	1.98 (1.44-2.74)	23.0
Current smoker of 20+ cigs/day	521	139	2.82 (2.13-3.73)	2.31 (1.70-3.15)	27.7	2.41 (1.73-3.35)	22.5
Other current smokers	533	139	1.41 (1.06-1.87)	1.29 (0.96-1.74)	28.7	1.35 (1.00-1.83)	13.9

* % of smoking-associated excess risk adjusted for age and sex explained by adjustment for risk factors.
† The 17 risk factors included in the stepwise model.
RR(95% CI) = Relative risk (95% confidence interval)

Table III. Effect of multivariate adjustment on risk factors (other than age, sex and smoking) included in the stepwise model

			Adjusted for age and sex			Also adjusted for smoking and the other 17 factors in the stepwise model		
	No. of subjects	No. of deaths	RR (95% CI)p	p	Dev 1 (d.f.)	RR (95% CI)	p	Dev 2 (d.f.)
Bored								
No	2086	498	1.00 (base)			1.00 (base)		
Sometimes	1664	402	1.17 (0.98-1.38)	(+)	45.1	1.02 (0.85-1.22)	NS	7.3
Often/always	4339	175	2.38 (1.85-3.06)	+++	(2)	1.45 (1.10-1.91)	++	(2)
No educational qualifications								
No	1361	226	1.00 (base)		35.8	1.00 (base)		8.8
Yes	2828	849	1.73 (1.44-2.08)	+++	(1)	1.37 (1.11-1.70)	++	(1)
Body mass index								
Acceptable	1294	297	1.00 (base)			1.00 (base)		
Underweight	154	75	3.05 (2.05-4.53)	+++		2.41 (1.59-3.65)	+++	
Overweight	1353	310	1.02 (0.83-1.25)	NS		1.04 (0.83-1.29)	NS	
Obese	525	120	1.24 (0.94-1.63)	NS	37.6	1.05 (0.78-1.43)	NS	17.9
Missing	863	273	1.38 (1.11-1.72)	++	(4)	1.00 (0.72-1.40)	NS	(4)
Work status								
Full time	1219	131	1.00 (base)			1.00 (base)		
Part time	452	50	0.96 (0.62-1.47)	NS		1.03 (0.66-1.60)	NS	
Retired	1879	781	1.36 (0.94-1.95)	(+)		1.23 (0.84-1.79)	NS	
Unemployed	112	20	1.91 (1.13-3.23)	+		1.26 (0.73-2.18)	NS	
Sick/disabled	148	54	3.54 (2.36-5.31)	+++	47.1	1.91 (1.23-2.97)	++	12.3
Keeping house	379	39	2.06 (1.27-3.36)	++	(5)	1.73 (1.04-2.87)	+	(5)
Mother dead								
No	859	68	1.00 (base)		17.6	1.00 (base)		9.3
Yes	3330	1007	1.80 (1.35-2.40)	+++	(1)	1.57 (1.17-2.12)	++	(1)
Type A personality								
No	1654	456	1.00 (base)			1.00 (base)		
Yes	1317	226	0.69 (0.56-0.84)	- - -	26.1	0.73 (0.59-0.91)	- -	11.1
Missing	1218	393	1.18 (0.98-1.42)	(+)	(2)	1.14 (0.86-1.52)	NS	(2)
Alcohol consumption								
None	1379	413	1.00 (base)			1.00 (base)		
Light	1520	353	0.83 (0.68-1.02)	(-)		0.89 (0.72-1.10)	NS	
Moderate/heavy	1012	202	0.82 (0.65-1.03)	(-)	25.0	0.84 (0.65-1.07)	NS	9.1
Ex (mod/heavy)	278	107	1.77 (1.29-2.43)	+++	(3)	1.37 (0.99-1.91)	(+)	(3)
Breakfast cereal								
Once/day +	1445	403	1.00 (base)			1.00 (base)		
Most days	462	105	0.89 (0.68-1.18)	NS		0.88 (0.66-1.19)	NS	
1 or 2/wk	501	85	0.71 (0.53-0.95)	-		0.68 (0.51-0.93)	-	
<1/wk	498	116	1.16 (0.89-1.52)	NS	23.0	1.21 (0.91-1.61)	NS	13.3
Never	1283	366	1.32 (1.09-1.60)	++	(4)	1.11 (0.90-1.37)	NS	(4)
Get enough exercise								
Yes	2541	654	1.00 (base)		9.2	1.00 (base)		4.7
No	1648	421	1.29 (1.09-1.52)	++	(1)	1.22 (1.02-1.47)	NS	(1)
Soft drinks								
None	1815	488	1.00 (base)			1.00 (base)		
<1/wk	932	195	0.84 (0.67-1.03)	(-)		0.89 (0.71-1.12)	NS	
1 or 2/wk	559	132	0.94 (0.73-1.21)	NS		0.99 (0.76-1.29)	NS	
Most days	378	99	1.17 (0.88-1.55)	NS	18.2	1.23 (0.91-1.65)	NS	12.3
1+/day	505	161	1.49 (1.17-1.91)	++	(4)	1.43 (1.11-1.85)	++	(4)

Table III. (cont) Effect of multivariate adjustment on risk factors (other than age, sex and smoking) included in the stepwise model

			Adjusted for age and sex			Also adjusted for smoking and the other 17 factors in the stepwise model		
	No. of subjects	No. of deaths	RR (95% CI)p	p	Dev 1 (d.f.)	RR (95% CI)	p	Dev 2 (d.f.)
Cups of coffee drunk								
None	1399	440	1.00 (base)			1.00 (base)		
1 or 2/day	1744	486	0.79 (0.66-0.94)	- -		1.00 (0.83-1.21)	NS	
3 or 4/day	641	109	0.66 (0.51-0.86)	- -	20.2	0.84 (0.64-1.12)	NS	8.3
5+/day	405	40	0.52 (0.36-0.75)	- - -	(3)	0.60 (0.41-0.89)	-	(3)
Risky occupation								
Missing/no	2700	638	1.00 (base)		1.1	1.00 (base)		8.1
Yes	1489	437	1.09 (0.92-1.29)	NS	(1)	0.75 (0.61-0.91)	- -	(1)
Social class								
I + II	1189	210	1.00 (base)			1.00 (base)		
III Non Manual	502	142	1.54 (1.17-2.03)	++		1.29 (0.96-1.73)	(+)	
III Manual	1502	405	1.54 (1.25-1.90)	+++	36.8	1.24 (0.96-1.60)	(+)	7.7
IV + V	996	318	1.98 (1.58-2.48)	+++	(3)	1.45 (1.11-1.91)	++	(3)
Marital status								
Married	3012	643	1.00 (base)			1.00 (base)		
Single	269	91	1.49 (1.10-2.03)	+	10.4	1.46 (1.05-2.02)	+	5.6
Divorced/Separated /Widowed	908	341	1.28 (1.05-1.56)	+	(2)	1.12 (0.91-1.38)	NS	(2)
Depression/nervous illness								
No	3337	840	1.00 (base)		16.2	1.00 (base)		2.9
Yes	852	235	1.50 (1.23-1.83)	+++	(1)	1.21 (0.97-1.49)	(+)	(1)
Ever tried to lose weight								
Yes	1446	300	1.00 (base)		0.4	1.00 (base)		3.7
No	2743	775	0.94 (0.79-1.12)	NS	(1)	0.82 (0.67-1.00)	(-)	(1)
Salad consumption (score)								
0-2	882	331	1.00 (base)			1.00 (base)		
3	632	185	0.89 (0.69-1.14)	NS		1.03 (0.79-1.34)	NS	
4	888	228	0.68 (0.54-0.86)	- -		0.81 (0.64-1.04)	(-)	
5	780	140	0.52 (0.40-0.68)	- - -		0.72 (0.55-0.95)	-	
6	562	119	0.62 (0.47-0.82)	- - -	41.2	0.86 (0.64-1.16)	NS	10.2
7-10	445	72	0.48 (0.35-0.66)	- - -	(5)	0.71 (0.50-1.01)	(-)	(5)

RR (95% CI) = Relative risk (95% Confidence interval); d.f. = degrees of freedom

Dev 1 is the difference in deviance based on comparison of the model including age, sex and the factor in question (Model A) with the model including only age and sex (for which the mean deviance is 0.925 on 4790 d.f.).

Dev 2 is the difference in deviance based on comparison of the stepwise model (for which the mean deviance is 0.866 on 4131 d.f.) with the stepwise model excluding the risk factor in question (Model D).

See text for codes for p values.

Table IV. Effect of multivariate adjustment on risk factors (other than age, sex and smoking) not included in the stepwise model

			Adjusted for age and sex			Also adjusted for smoking and the factors in the stepwise model		
	No. of subjects	No. of deaths	RR (95% CI)	p	Dev 1 (d.f.)	RR (95% CI)	p	Dev 2 (d.f.)
Household size (other than subject)								
0	794	311	1.00 (base)			1.00 (base)		
1	2040	585	0.90 (0.74-1.10)	NS		1.21 (0.88-1.66)	NS	
2	743	118	0.80 (0.60-1.06)	NS		1.07 (0.73-1.56)	NS	
3	411	38	0.51 (0.34-0.76)	- -	12.7	0.76 (0.47-1.23)	NS	7.2
4+	201	23	0.66 (0.40-1.08)	(-)	(4)	0.84 (0.47-1.50)	NS	(4)
Household income (£ per week)								
135+	1030	122	1.00 (base)			1.00 (base)		
97-134	523	102	1.44 (1.06-1.97)	+		1.03 (0.74-1.43)	NS	
79-96	431	120	2.00 (1.46-2.73)	+++		1.21 (0.86-1.71)	NS	
54-78	736	289	2.28 (1.73-3.01)	+++		1.29 (0.94-1.76)	NS	
<54	674	259	1.93 (1.45-2.58)	+++	41.4	0.96 (0.67-1.36)	NS	7.2
Missing	795	183	1.51 (1.14-2.01)	++	(5)	0.97 (0.71-1.33)	NS	(5)
Anything to keep healthy								
Yes	2603	652	1.00 (base)		8.2	1.00 (base)		0.7
No	1586	423	1.27 (1.08-1.50)	++	(1)	1.08 (0.90-1.28)	NS	(1)
Time before first meal in day								
<30 mins	2060	537	1.00 (base)			1.00 (base)		
30 mins - 1 hr	1000	281	1.03 (0.84-1.25)	NS		1.05 (0.85-1.28)	NS	
1-2 hrs	360	86	1.01 (0.75-1.36)	NS		0.91 (0.66-1.24)	NS	
2-3 hrs	225	51	1.28 (0.88-1.84)	NS		0.99 (0.67-1.47)	NS	
3-4 hrs	190	46	1.39 (0.94-2.05)	NS	7.3	0.97 (0.64-1.47)	NS	0.9
4+ hrs	354	74	1.38 (1.01-1.87)	+	(5)	0.94 (0.66-1.32)	NS	(5)
Fried foods (score)								
0-4	808	211	1.00 (base)			1.00 (base)		
5-6	1139	304	0.93 (0.73-1.17)	NS		0.97 (0.76-1.25)	NS	
7	647	152	0.97 (0.73-1.27)	NS		1.00 (0.75-1.34)	NS	
8-9	1137	286	0.98 (0.77-1.25)	NS	2.8	0.92 (0.71-1.19)	NS	0.6
10-16	458	122	1.18 (0.87-1.60)	NS	(4)	0.95 (0.68-1.31)	NS	(4)
Slices of bread (per day)								
0-2	1041	231	1.00 (base)			1.00 (base)		
3	814	216	1.16 (0.91-1.48)	NS		1.18 (0.91-1.53)	NS	
4	1064	283	1.02 (0.81-1.28)	NS		0.98 (0.77-1.25)	NS	
5	394	113	1.13 (0.84-1.53)	NS	6.0	1.11 (0.81-1.52)	NS	3.7
6+	876	232	1.30 (1.02-1.66)	+	(4)	1.17 (0.90-1.52)	NS	(4)
Fruit consumption (score)								
0-4	916	295	1.00 (base)			1.00 (base)		
5-7	849	246	0.92 (0.73-1.16)	NS		1.18 (0.92-1.52)	NS	
8	625	179	0.80 (0.62-1.04)	(-)		0.98 (0.75-1.29)	NS	
9-10	856	189	0.68 (0.54-0.87)	- -	21.3	1.04 (0.79-1.36)	NS	2.4
11-15	943	166	0.61 (0.48-0.78)	- - -	(4)	1.05 (0.79-1.41)	NS	(4)
Vegetable consumption (score)								
0-5	634	219	1.00 (base)			1.00 (base)		
6	741	195	0.83 (0.63-1.08)	NS		1.03 (0.78-1.36)	NS	
7	708	173	0.72 (0.55-0.95)	-		0.94 (0.70-1.25)	NS	
8	819	182	0.72 (0.55-0.94)	-		0.98 (0.74-1.30)	NS	
9	703	166	0.73 (0.56-0.96)	-	8.5	0.98 (0.73-1.32)	NS	2.1
10-15	584	140	0.81 (0.61-1.07)	NS	(5)	1.16 (0.85-1.58)	NS	(5)

Table IV. (cont) Effect of multivariate adjustment on risk factors (other than age, sex and smoking) not included in the stepwise model

			Adjusted for age and sex			Also adjusted for smoking and the factors in the stepwise model		
	No. of subjects	No. of deaths	RR (95% CI)	p	Dev 1 (d.f.)	RR (95% CI)	p	Dev 2 (d.f.)
Nuts								
None	1921	626	1.00 (base)			1.00 (base)		
<1/wk	1489	312	0.66 (0.55-0.79)	- - -		0.86 (0.71-1.04)	NS	
1 or 2/wk	475	74	0.51 (0.38-0.68)	- - -		0.72 (0.53-0.98)	-	
Most days	151	30	0.66 (0.42-1.05)	(-)	36.8	0.82 (0.50-1.34)	NS	5.7
One/Day+	153	33	0.62 (0.40-0.98)	-	(4)	0.84 (0.52-1.35)	NS	(4)
Cream								
None	1902	582	1.00 (base)			1.00 (base)		
<1/wk	1498	325	0.71 (0.60-0.85)	- - -		0.89 (0.73-1.08)	NS	
1 or 2/wk	666	134	0.62 (0.49-0.79)	- - -	23.5	0.81 (0.62-1.04)	NS	3.2
Most Days+	123	34	0.67 (0.42-1.06)	NS	(3)	0.93 (0.58-1.51)	NS	(3)
Sugar in hot beverages (teaspoons)								
0	1897	411	1.00 (base)			1.00 (base)		
1	760	207	1.02 (0.82-1.27)	NS		0.90 (0.71-1.14)	NS	
2	904	271	1.28 (1.04-1.57)	+		1.03 (0.82-1.30)	NS	
3	221	71	1.27 (0.90-1.80)	NS	7.0	0.92 (0.63-1.34)	NS	2.0
4-6	407	112	1.10 (0.82-1.46)	NS	(4)	0.87 (0.62-1.21)	NS	(4)
Cups of tea drunk								
None	241	32	0.71 (0.46-1.10)	NS		0.85 (0.52-1.38)	NS	
1 or 2/day	564	118	0.84 (0.64-1.11)	NS		0.95 (0.71-1.27)	NS	
3 or 4/day	1231	354	1.00 (base)			1.00 (base)		
5 or 6/day	988	278	0.93 (0.75-1.15)	NS	9.4	0.92 (0.74-1.15)	NS	1.1
>6/day	1165	293	1.16 (0.94-1.43)	NS	(4)	0.91 (0.73-1.14)	NS	(4)
Sweet food consumption (score)								
0-8	777	189	1.00 (base)			1.00 (base)		
9-11	727	173	0.89 (0.68-1.17)	NS		1.03 (0.78-1.37)	NS	
12-14	883	223	0.85 (0.66-1.10)	NS		1.08 (0.82-1.42)	NS	
15-17	870	240	0.87 (0.67-1.12)	NS	9.7	1.19 (0.90-1.57)	NS	3.9
18+	932	250	0.68 (0.53-0.88)	- -	(4)	0.94 (0.71-1.24)	NS	(1)
Uses low fat/polyunsaturated spread on bread								
Yes	1021	207	1.00 (base)		5.4	1.00 (base)		0.02
No	3168	868	1.25 (1.03-1.52)	+	(1)	0.99 (0.80-1.22)	NS	(1)
Cut down on fatty foods								
Yes	1803	400	1.00 (base)			1.00 (base)		
No	1495	393	1.06 (0.89-1.27)	NS	4.7	1.01 (0.83-1.23)	NS	0.02
Missing	891	282	1.25 (1.02-1.54)	+	(2)	1.04 (0.57-1.89)	NS	(2)
Neuroticism								
0-4	612	144	1.00 (base)			1.00 (base)		
5-7	602	142	1.20 (0.89-1.63)	NS		1.15 (0.84-1.58)	NS	
8-10	574	137	1.73 (1.27-2.36)	+++		1.57 (1.13-2.18)	++	
11-14	637	134	1.38 (1.02-1.88)	+		1.17 (0.83-1.63)	NS	
15-24	525	114	1.82 (1.32-2.52)	+++	34.3	1.30 (0.90-1.89)	NS	9.0
Missing	1239	404	1.93 (1.49-2.51)	+++	(5)	1.66 (0.93-2.96)	(+)	(5)

Table IV. (cont 2) Effect of multivariate adjustment on risk factors (other than age, sex and smoking) not included in the stepwise model

			Adjusted for age and sex			Also adjusted for smoking and the factors in the stepwise model		
	No. of subjects	No. of deaths	RR (95% CI)	p	Dev 1 (d.f.)	RR (95% CI)	p	Dev 2 (d.f.)
Extroversion								
0-7	797	197	1.00 (base)			1.00 (base)		
8-10	759	195	1.07 (0.82-1.38)	NS		1.02 (0.78-1.35)	NS	
11-12	501	105	0.94 (0.70-1.27)	NS		0.91 (0.66-1.25)	NS	
13-15	593	129	1.00 (0.75-1.33)	NS		0.99 (0.73-1.34)	NS	
16-24	300	46	0.81 (0.55-1.20)	NS	17.9	0.86 (0.58-1.30)	NS	2.8
Missing	1239	403	1.40 (1.11-1.76)	0	(5)	1.42 (0.79-2.57)	NS	(5)
Hours slept								
<6	603	180	1.19 (0.93-1.53)	NS		0.95 (0.73-1.24)	NS	
6-7	1126	272	1.03 (0.83-1.28)	NS		1.01 (0.81-1.26)	NS	
7-8	1347	284	1.00 (base)			1.00 (base)		
8-9	877	238	1.08 (0.86-1.35)	NS		1.00 (0.79-1.27)	NS	
9-10	148	63	1.59 (1.06-2.37)	+	11.1	1.32 (0.87-2.00)	NS	2.4
10+	88	38	1.89 (1.13-3.15)	+	(5)	1.13 (0.66-1.94)	NS	(5)
Lonely								
No	2838	667	1.00 (base)			1.00 (base)		
Sometimes	1000	264	1.19 (0.98-1.44)	(+)	19.7	1.08 (0.87-1.35)	NS	1.0
Often/Always	351	144	1.82 (1.39-2.38)	+++	(2)	1.15 (0.83-1.60)	NS	(2)

RR(95% CI) = Relative risk (95% Confidence interval)
Dev 1 is the difference in deviance based on comparison of the model including age, sex and the factor in question (Model A) with the model including only age and sex (for which the mean deviance is 0.925 on 4179 d.f.; d.f. = degrees of freedom).
Dev 2 is the difference in deviance based on comparison of the stepwise model (for which the mean deviance is 0.866 on 4131 d.f.) with the model including the same factors plus the risk factor in question (Model C).
See text for codes for p values.

Table V. Effect of adjustment on relative risk associated with current smoking of 20+ cigarettes/day

	Age and sex only included		All stepwise model factors included	
	Relative risk	Excess risk explained* (%)	Relative risk	Excess risk explained† (%)
No risk factor	2.818		2.315	
Factors in the stepwise model				
Bored	2.597	12.2	2.362	3.5
No educational qualifications	2.680	7.6	2.324	0.7
Body mass index	2.698	6.6	2.427	7.9
Work status	2.725	5.1	2.322	0.6
Mother dead	2.791	1.5	2.335	1.5
Type A personality	2.769	2.7	2.313	-0.1
Alcohol consumption	2.859	-2.3	2.293	-1.7
Breakfast cereal	2.676	7.8	2.370	4.0
Get enough exercise	2.813	0.3	2.326	0.9
Soft drinks	2.810	0.4	2.290	-1.9
Cups of coffee drunk	2.887	-3.8	2.207	-8.9
Risky occupation	2.810	0.4	2.284	-2.4
Social class	2.651	9.2	2.338	1.7
Marital status	2.799	1.0	2.300	-1.1
Depression/nervous illness	2.727	5.0	2.334	1.5
Ever tried to lose weight	2.867	-2.7	2.281	-2.6
Salad consumption	2.527	16.0	2.437	8.5
Factors not in the stepwise model				
Household size	2.798	1.1	2.320	-0.4
Household income	2.667	8.3	2.321	-0.5
Anything to keep healthy	2.739	4.4	2.296	1.4
Time before first meal in day	2.741	4.2	2.343	-2.1
Fried foods	2.809	0.5	2.327	-0.9
Slices of bread	2.817	0.0	2.326	-0.9
Fruit consumption	2.584	12.8	2.310	0.3
Vegetable consumption	2.765	2.9	2.309	0.4
Nuts	2.658	8.8	2.301	1.0
Cream	2.683	7.4	2.292	1.7
Sugar in hot beverages	2.777	2.3	2.334	-1.5
Cups of tea drunk	2.769	2.7	2.346	-0.2
Sweet foods	2.743	4.1	2.325	-0.8

Table V. (cont) Effect of adjustment on relative risk associated with current smoking of 20+ cigarettes/day

	Age and sex only included		All the stepwise model factors included	
	Relative risk	Excess risk explained* (%)	Relative risk	Excess risk explained† (%)
Factors not in stepwise regression (cond)				
Uses low fat/PU spread	2.756	3.4	2.319	-0.3
Cut down on fatty foods	2.818	0.0	2.315	-0.1
Neuroticism	2.725	5.1	2.320	-0.4
Extroversion	2.805	0.7	2.312	0.2
Hours slept	2.810	1.3	2.317	-0.2
Lonely	2.736	4.5	2.316	-0.1

* % of excess risk associated with current smoking of 20+ cigs/day, adjusted for age and sex, explained by adjustment for risk factors stated.

† For factors in the stepwise model, % of excess risk adjusted for all factors in model except factor stated, explained by further adjustment for that risk factor; for factors not in model, % of excess risk, adjusted for all factors in model, explained by further adjustment for risk factor stated.

Table VI. Effect of multivariate adjustment on mortality associated with smoking (based on analyses omitting 6 of the 36 risk factors - see text)

	Relative risk (95% CI)		Excess risk* explained (%)	RR (95% CI)	Excess risk* explained (%)
	Adjusted for age and sex	Also adjusted for 13 factors†		Adjusted for all 30 factors	
Never smoked	1.00 (base)	1.00 (base)		1.00 (base)	
Ex-smoker <10 years ago	1.95 (1.45-2.62)	1.85 (1.36-2.51)	10.8	1.75 (1.28-2.40)	21.0
Other ex-smoker	1.51 (1.19-1.92)	1.54 (1.20-1.98)	-6.0	1.49 (1.15-1.92)	3.6
Current smoker of 10-19 cigs/day	2.28 (1.71-3.04)	2.07 (1.53-2.81)	16.1	2.05 (1.50-2.82)	17.5
Current smoker of 20+cigs/day	2.82 (2.13-3.73)	2.56 (1.90-3.46)	14.0	2.51 (1.82-3.47)	16.7
Other current smokers	1.41 (1.06-1.87)	1.40 (1.04-1.88)	2.3	1.38 (1.02-1.87)	6.8

• % of smoking-associated excess risk adjusted for age and sex explained by adjustment for risk factors.

† The 13 risk factors included in the new stepwise model.

RR(95% CI) = Relative risk (95% Confidence interval)

Table VII. Effect of multivariate adjustment on mortality associated with smoking (omitting subjects who assessed their health as poor)

	Relative risk (95% CI)		Excess risk* explained (%)	RR (95% CI)	Excess risk* explained (%)
	Adjusted for age and sex	Also adjusted for 17 factors†		Adjusted for all 36 factors	
Never smoked	1.00 (base)	1.00 (base)		1.00 (base)	
Ex-smoker <10 years ago	1.85 (1.35-2.52)	1.66 (1.20-2.31)	21.7	1.67 (1.19-2.33)	21.5
Other ex-smoker	1.53 (1.19-1.96)	1.55 (1.20-2.02)	-5.0	1.54 (1.17-2.01)	-1.4
Current smoker of 10-19 cigs/day	2.14 (1.57-2.91)	1.89 (1.36-2.62)	21.9	1.87 (1.32-2.64)	23.7
Current smoker of 20+cigs/day	3.09 (2.30-4.15)	2.59 (1.87-3.58)	23.8	2.66 (1.88-3.77)	20.3
Other current smokers	1.43 (1.06-1.92)	1.33 (0.98-1.81)	22.4	1.36 (0.99-1.87)	15.3

* % of smoking-associated excess risk adjusted for age and sex explained by adjustment for risk factors.
† The 17 risk factors included in the stepwise model.
RR(95% CI) = Relative risk (95% Confidence interval).

Table VIII. Effect of multivariate adjustment on mortality associated with smoking (omitting subjects who assessed their health as poor; based on analyses omitting 6 of the 36 risk factors)

	Relative risk (95% CI)		Excess risk* explained (%)	RR (95% CI)	Excess risk* explained (%)
	Adjusted for age and sex	Also adjusted for 13 factors†		Adjusted for all 30 factors	
Never smoked	1.00 (base)	1.00 (base)		1.00 (base)	
Ex-smoker <10 years ago	1.85 (1.35-2.52)	1.74 (1.26-2.41)	12.4	1.70 (1.22-2.37)	17.1
Other ex-smoker	1.53 (1.19-1.96)	1.56 (1.20-2.03)	-6.2	1.53 (1.17-2.00)	-0.2
Current smoker of 10-19 cigs/day	2.14 (1.57-2.91)	2.01 (1.45-2.78)	11.3	1.95 (1.39-2.73)	16.6
Current smoker of 20+cigs/day	3.09 (2.30-4.15)	2.82 (2.05-3.87)	13.0	2.78 (1.98-3.91)	14.6
Other current smokers	1.43 (1.06-1.92)	1.40 (1.03-1.90)	6.9	1.40 (1.02-1.92)	7.0

* % of smoking-associated excess risk adjusted for age and sex explained by adjustment for risk factors.
† The 13 risk factors included in the new stepwise model.
RR(95% CI) = Relative risk (95% Confidence interval).

Table IX Age adjusted mortality associated with smoking by sex

	Males			Females		
	No. of subjects	No. of deaths	Relative risk* (95% CI)	No. of subjects	No. of deaths	Relative risk* (95% CI)
Never smoked	288	51	1.00 (base)	1118	244	1.00 (base)†
Ex-smoker <10 years ago	228	74	2.45 (1.54-3.90)	205	45	1.78 (1.18-2.69)
Other ex-smoker	505	174	1.91 (1.28-2.84)	328	87	1.36 (0.98-1.89)
Current smoker of 10-19 cigs/day	187	77	3.88 (2.41-6.25)	276	45	1.55 (1.04-2.30)
Current smoker of 20+ cigs/day	285	84	3.11 (1.99-4.85)	236	55	3.09 (2.09-4.57)
Other current smokers	340	98	1.68 (1.09-2.58)	193	41	1.44 (0.93-2.21)

* Adjusted for age.
† The relative risk of females to males was 0.74 (95% CI 0.46-1.19) among never smokers.
95% CI = 95% Confidence interval

Table X. Association between four risk factors and age and sex adjusted mortality in never smokers and current smokers

	Never smokers			Current smokers		
	No. of subjects	No. of deaths	Relative risk* (95% CI)	No. of subjects	No. of deaths	Relative risk* (95% CI)
0 or 1 risk factors†	320	11	1.00 (base)	311	38	1.00 (base)‡
2	467	88	3.34 (1.68-6.65)	509	112	1.41 (0.92-2.16)
3	600	184	3.82 (1.94-7.52)	630	215	2.11 (1.39-3.18)
4	19	12	18.1 (5.28-61.9)	67	35	4.89 (2.57-9.31)

* Adjusted for age and sex.

† Subjects were considered to have a risk factor present if they had no educational qualifications, had a mother who was dead, were an ex-drinker (moderate or heavy) of alcohol or did not have a Type A personality.

‡ The relative risk of current smokers to never smokers was 5.23 (2.24-12.21) among those with 0 or 1 risk factors.

95% CI = 95% Confidence interval

REFERENCES

1. US Surgeon General. *Reducing the health consequences of smoking. 25 years of progress. A report of the Surgeon General*. US Public Health Service, Rockville, Maryland: 1989. DHHS (CDC) 89-8411.

2. Doll R. Uncovering the effects of smoking: historical perspective. *Statistical Methods in Medical Research* 1998;**7**:87-117.

3. US Surgeon General. *Smoking and health, a report of the Surgeon General*. Office on Smoking and Health: Rockville, Maryland, 1979. (PHS)79-50066.

4. Peto R, Lopez AD, Boreham J, Thun M, Heath C, Jr. *Mortality from smoking in developed countries 1950-2000; Indirect estimates from national vital statistics*. Oxford University Press: Oxford, New York, Tokyo, 1994.

5. Thornton A, Lee P, Fry J. Differences between smokers, ex-smokers, passive smokers and non-smokers. *Journal of Clinical Epidemiology* 1994;**47**:1143-1162.

6. Cox BD, Blaxter M, Buckle ALJ, Fenner NP, Golding JF, Gore M, Huppert FA, Nickson J, Roth M, Stark J, Wadsworth MEJ, Whichelow M. *The health and lifestyle survey. Preliminary report of a nationwide survey of the physical and mental health, attitudes and lifestyle of a random sample of 9,003 British adults*. Health Promotion Research Trust: London, 1987.

7. Breslow NE, Day NE. *The analysis of case-control studies,* Volume 1. IARC: Lyon, 1980. (Statistical methods in cancer research.) IARC Scientific Publication No. 32.