DETERMINANTS OF TRENDS IN US LUNG CANCER MORTALITY RATES

AN EVALUATION AND EXTENSION OF THE WORK OF SWARTZ

Authors: P.N. Lee and Mrs B.A. Forey

Date : 30 June 1994

Contents

Contents (cont'd)

Contents (cont'd)

1. Introduction

Based on US smoking prevalence data published by Harris (1980), Swartz (1992) used a mathematical model to construct detailed smoking histories of the US white male population by age and cohort. Using functions derived by Whittemore (1988) from the multistage model of carcinogenesis to relate lung cancer risk to these smoking histories, Swartz predicted that, among the age group 42-70, there should have been a 12% decline in lung cancer over the period 1970 to 1985. In contrast, he noted that the actual total rate of lung cancer increased by 26% over this period. Taking into account the decline in average tar content of cigarettes over this period (not taken into account in the prediction), and the relatively constant dose rate among smokers (the prediction assumed smokers smoke a constant amount), Swartz considered that "these results strongly suggest that the recent increase in lung cancer among white males in the USA is due entirely or in large part to factors other than cigarette smoking".

The suggestion that factors other than cigarette smoking may be a major determinant of lung cancer trends is an important one that demands further attention. The major purpose of this document is to try to gain insight into the reliability of Swartz's conclusions by determining how dependent they are on the particular way in which the analyses were conducted. Specifically we wished to investigate how contingent his conclusions were on various circumstances of his analysis, namely:

(i) the source of data used for estimating smoking prevalence;

(ii) the method used for estimating smoking histories from the prevalence data;

(iii) the use of the multistage model of carcinogenesis for estimating risk of lung cancer from smoking;

(iv) the specific form of multistage model used;

(v) the particular age group, period and sex used for contrasting observed and and predicted lung cancer rates; and

(vi) various aspects of smoking not taken into account in the model which might affect the comparison.

We also felt it useful to summarize available data on trends over time in lung cancer risk in nonsmokers, as this might cast separate light on the hypothesis that risk due to factors other than cigarette smoking is increasing.

2. Reproducing Swartz's results

A first step in the process was to attempt to reproduce Swartz's published findings. There were a number of problems in doing this.

2.1 Harris data not in numeric form

The source paper by Harris (1980) gives smoking prevalence data only in graphical and not in numerical form, and Swartz (1992) only cites (in his Table II) selected data. We wrote to Swartz (a copy of all our correspondence with Swartz is attached as Appendix A) asking him to supply a copy of the full data he had used. Unfortunately, he appears not now to have these data and accordingly

we derived our own estimates from the graphs. As shown in Table 1, which reproduces our estimates and compares them where possible with Swartz's tabulated figures, there is very little difference between the two sets of data. Accordingly we decided to use the data we had derived in all subsequent analyses.

2.2 Lung cancer data for whites or for the whole population

Swartz's paper refers to lung cancer rates for US white males. However the logic in restricting to whites is unclear given that the Harris smoking prevalence data relates to the whole US population, and the main mathematical prediction model used is based on a fit by Whittemore (1988) to the British Doctor's data of Doll and Peto (1976,1978), British Doctors being not all white (although of course the ethnic mix is different from that in the US). As we had readily available WHO lung cancer data for the US as a whole, and did not have data available for whites, (Swartz's Table III referred to a non-existent reference 30 as source), we decided to restrict our attention to overall US data for all subsequent analyses. The main purpose of attempting to reproduce Swartz's results was in any case to see whether we could reproduce his smoking-based predictions, not his estimate of the rise in age-standardized risk of lung cancer (which is a trivial calculation).

2.3 Possible errors in Swartz's lung cancer mortality function

Formula (1) of Swartz (1992) states that the parameter C is the smoking rate in packs per day. Having produced lung cancer estimates that were ridiculously low, and having looked in detail at

Whittemore (1988) from which the formula was derived, we realized that C was actually cigarettes per day. Swartz confirmed this in correspondence.

We also realized that Swartz's formula (1) could not be derived from the multistage model. Correspondence with Whittemore (see Appendix B) revealed that though she had used the correct formula in her 1988 fits to the British Doctors, US Veterans and New Mexico data, she had inadvertently published an incorrect formula and Swartz had used this without realizing it. The formulae, given in Appendix B, do not differ for continuous smokers, but they do differ for ex-smokers. Despite this, we were unable to reproduce exactly Whittemore's predictions for British doctors (Whittemore (1988) Table 1), although our results were very close. A possible explanation for the discrepancy may be a different level of accuracy in the parameters supplied. While trying to reproduce Swartz's results, we kept to his incorrect formula in the first place. Later, when trying the effect of alternative predictor functions, we used correctly derived multistage functions.

2.4 Other possible sources of difference

One possible source of difference lies in the handling of the Harris prevalence data. These data are presented for cohorts covering a 10-year spread of dates of birth, and we have followed Swartz in assuming that, as given, the data apply to the mid-point year of birth, and in using linear interpolation to estimate values for the intermediate years of birth. Swartz's description of this is brief and we may not have used precisely the same method. Where the

same age data were available in two successive mid-year cohorts, then linear interpolation was used within each individual age. In addition, linear extrapolation, based on the last five available ages within each individual cohort, was used to extend the data for the intermediate cohorts up to the final year (1980). The need for this stage was not mentioned by Swartz.

Another possible source of difference lies in the mortality and population data used for age-standardization. Swartz describes this as "age adjusted to the 1970 US population" without a specific reference, and does not state the width of age group used - mortality and population data are typically published in five-year age groups (..40-44, 45-49 ..), which are not directly applicable to his age range of 42-70. We have used the WHO data for the whole US population, using the simple population estimate of individual ages as one-fifth of the five-year age group, and a smoothing of rates based on linear interpolation between successive 5-year age groups.

2.5 Comparison of Swartz's reported findings and those we derived

Table 2 compares data on observed and predicted relative rates as presented by Swartz in his Table III and as derived by us. It can be seen that though our calculations agree with Swartz in predicting a declining rate when the rate actually increases, the magnitude of the increase and the decline are not the same. While the difference in actual rates may be explicable in terms of our using overall US data and Swartz using data for Whites, and while some of the differences mentioned in section 2.4 may have had some

effect, it is not at all apparent why we should end up with differing predictions. We hope to resolve this in further correspondence with Swartz.

2.6 Adequacy of the predictions

It is notable that Swartz only presents rates relative to 1970. Formula (1) of his paper was apparently intended to give a prediction of absolute risk but no data were presented to show how well it actually predicted. Actually the fit was not very good. At year 1970, for example, the actual lung cancer rate according to WHO was 1338.6 per million, but the model only predicted a figure of 781.5 per million. By 1985, the actual rate was 1501.7 per million and the predicted rate 742.6 million.

2.7 Sensitivity analyses

Based on the same data (US males aged 42-70), Swartz noted that the proportional decline in predicted relative lung cancer rate (12% for the main model - see Table 2) varied when some of the model assumptions were relaxed or varied. In particular he noted that the decline:

(i) remained at 12% if 0.5% drift was allowed in his smoking submodel,

(ii) remained at 12% if smokers were assumed to smoke 2 packs per day rather than one,

(iii) remained at 12% if smokers were assumed to start smoking at age 18 rather than 21 years,

(iv) reduced to 5% if Whittemore's pack-years function (his formula

2) was used, and

(v) reduced to 8% if a multistage model was used with five stages with only the fourth affected by smoking.

Compared with our estimates of Table 2 of a 5.2% decline we found that variations (i), (ii), (iii) and (iv) produced estimates respectively of a decline of 5.3%, 5.1% and 4.5%, and an increase of 1.6%. Thus we agreed that the first three variants made very little difference to the predictions, and that the predictions from the alternative pack-year function were closer to, although still lower than, the observed rates. We were unable to attempt to reproduce Swartz's fifth estimate, as he gave no details of the constants he had used for his predictions.

3. A more general test of the claim that observed lung cancer rates have risen faster than predicted lung cancer rates - methods

3.1 Age, sex and period

Rather than use a single period and sex and the rather odd age group 42-70 we decided to test the claim using each combination of:

Sex Male and female

Age 45-54, 55-64 and 65-74

Period 1956-1965, 1966-1975 and 1976-1985

Neither changes in diagnostic standards (Royal College of Physicians 1977) nor changes in the ICD Revision (Lee *et al* 1990) are likely to have had much effect on changes in observed lung cancer rates over this period.

Exceptionally we did not consider the oldest age group and the earliest period in combination as this involved people born around

1890 where the smoking data were clearly at their most unreliable. Note that all observed and predicted lung cancer rates are standardized to the age distribution of the 1970 US population of the sex being considered.

3.2 Smoking history submodel

We considered three submodels to construct smoking histories from smoking prevalence data by cohort:

Swartz without drift In this model, when smoking prevalence at one year exceeds that in a previous year, an appropriate number of subjects are moved from the never smoking category to the current smoking category. When the prevalence declines, an appropriate number of subjects are moved from the current smoking category to the former smoking category, the proportion moving in each age of starting group being the same. Subjects are not allowed to restart smoking, and thus only have one period of smoking at most.

Swartz with drift The "Swartz without drift" submodel assumes that within a cohort at any given age, some subjecte may start smoking or some give up smoking, but not both. The submodel with drift allows for both to occur at the same time by moving at each year an additional number of subjects, equal to 0.5% of the current smokers, from never smoked to current smoker, and an identical number from current smoker to former smoker.

Townsend By disallowing subjects to restart smoking once they had stopped, Swartz effectively minimizes the number of long term smokers. A contrasting algorithm which maximizes the number of long term smokers was used by Townsend (1978). Here subjects are

considered to be ranked in order of "desire to smoke". When prevalence decreases, the subjects with the lowest "desire to smoke" who are smoking at the time are assumed to give up. When it increases, the subjects with the highest "desire to smoke" who are not smoking at the time are assumed to start. Here there is no restriction on a subject having two or more periods of smoking. The "desire to smoke" is assessed as equivalent to the "duration of smoking". Swartz had avoided such models as he thought the multistage functions for predicting risk to be too complex. However, they are not in fact difficult to program.

Appendix C gives an example of how, for each of the three smoking submodels, prevalences of smoking are converted into numbers of subjects starting or stopping smoking at different ages. This output may be useful for checking the different predictions reached by Swartz and ourselves. Constructing the smoking history is one of the more complex parts of the calculation and may have been the source of the discrepancy.

In practice we found that the different treatment of prevalence increases between the Swartz and Townsend models did not make a substantial impact, since most prevalence increases occurred together at the younger ages for each cohort. Thus relatively few smokers restarted under the Townsend model. However the treatment of prevalence decreases had a greater effect, with Swartz ex-smokers being drawn from all available ages while under Townsend the later starters gave up soonest.

3.3 Predictors of absolute risk

Four functions were used to predict absolute lung cancer risk.

3.3.1 Swartz 1 British Doctors

In this model, risk at age t is given by

$$M(f) = 2.01\text{x}10^{-12}[(t-5)^{4.5} + pc(1+2pc)(t_1-t_0)^{4.5} + 2pc(t_1^{4.5}-t_0^{4.5})]$$

where t_0 is age at starting and t_1 is time of giving up. t-5 replaces t_1 for current smokers and when $t_1 \geq t-5$. p is a constant, 0.207, a value reported by Whittemore (1988) as her best fit to the British Doctor's data. c is the number of cigarettes per day taken as 20 by Swartz. $2.01\text{x}10^{-12}(t-5)^{4.5}$, the predicted risk in nonsmokers (the background rate) comes from a fit by Whittemore to age specific data on lung cancer risk in male nonsmokers in the American Cancer Society Cancer Prevention Study I.

3.3.2 Swartz 1 US Veterans

The formula is identical to that in Swartz 1 British Doctors except that the value of p used is 0.128, the value which Whittemore found to fit best to data for US Veterans.

3.3.3 Swartz 2 British Doctors

Here risk at age t is given by

$$M(t) = 2.01\text{x}10^{-12}(t-5)^{4.5}(1 + au)$$

where u is cumulative packs smoked and a is constant. For the British Doctors data the value of a fitted by Whittemore was 1.13×10^{-3}.

3.3.4 Swartz 2 US Veterans

The formula is identical to that in Swartz 2 British Doctors except that the value of a was 0.59×10^{-3}.

3.3.5 Swartz smoking submodels for predictors of absolute risk

The Townsend smoking submodel was not used with the predictors of absolute risk, only with the predictors proportional to excess risk (*vide infra*). There were two reasons for this. First, the Swartz 1 predictors are undefined for the Townsend submodel where multiple smoking periods may occur. Second, the Swartz 2 predictors, which can be calculated directly from the Harris prevalence data, are unaffected by the smoking model.

3.4 Predictors proportional to excess risk

Consider the formula

$$L(t) = B(t) + E(t)$$

where L(t) is the observed total absolute risk of lung cancer at year t, B(t) is the "background risk" (associated with factors other than smoking) and E(t) is the "excess risk" (associated with smoking). Swartz's main conclusions depended on comparison of the ratio $L(t_a)/L(t_b)$ of the risks observed at two time points t_a and t_b with the corresponding ratio $P(t_a)/P(t_b)$ of predicted risks. Since the null hypothesis is that the background risk is invariant of

time, and since the formulae used by Swartz only took account of variation in smoking over time, an equally valid test of the null hypothesis would clearly have been to compare the ratio $E(t_a)/E(t_b)$ of excess risks with the corresponding ratio of predicted excess risks. Furthermore, since one is considering a ratio, one only needs a function that is proportional to the excess risk. Thus, for example, if one postulates that excess risk is proportional to pack years smoked, one does not need to know the constant of proportionality to conduct the analysis. This simplifies the calculations as no model fitting is involved.

3.4.1 Predictors based on the multistage model

What appropriate predictors proportional to excess risk might one use? As a first step, a review of the evidence supporting a multistage model was carried out (Appendix D). This concluded that the multistage model had a lot going for it - it is flexible, reasonably tractable and in broad terms its predictors fit in with a number of observed facts. These include:

(i) the approximate power law relationship of incidence with duration of exposure when exposure is continuous;

(ii) the evidence that age *per se* does not affect incidence of many cancers;

(iii) the direct evidence from initiation/promotion studies that some cancers require multiple exposures in a specific order for cancer to arise;

(iv) the observation that tumour incidence may be increased as a result of exposure that has long since ceased;

(v) the evidence of a quadratic dose-response relationship for some carcinogens; and

(vi) the evidence that the joint effect of two carcinogens is often multiplicative, or at least markedly super-additive.

It also describes reasonably well patterns of incidence following cessation of exposure.

Accordingly it was decided to include a number of functions based on a multistage model with k stages.

<u>Multistage 1:0</u> First stage only affected

<u>Multistage 5:1</u> First and penultimate stages affected, first stage five times as strongly

<u>Multistage 1:1</u> First and penultimate stages affected equally

<u>Multistage 1:2</u> First and penultimate stages affected, penultimate stage twice as strongly
(This is equivalent to the model Whittemore found to fit best)

<u>Multistage 1:2E</u> As 1:2 but including the formula error that Swartz incorporated.

<u>Multistage 1:5</u> First and penultimate stages effected, penultimate stage five times as strongly

<u>Multistage 0:1</u> Penultimate stage only affected.

Formulae for all these models can be obtained from Appendix D. In all these models it is assumed that other stages are not affected. The evidence that the first and penultimate stages are affected is discussed in Appendix D. It is clear that a model in which only the first stage is affected will not adequately explain the decline in relative risk on cessation of smoking, and that a model in which

only the penultimate stage is affected will not adequately explain the strong relationship of risk to age of starting to smoke given age. These models are included only for completeness. Most model-fitting work to specific data sets has concluded that both stages are affected with the effects on the two stages not very different. The evidence in favour of the penultimate stage being twice as affected as the first came from an analyses by Brown and Chu (1987), a conclusion used by Whittemore (1988) in her model-fitting work.

Another function related to the multistage model is

<u>Duration^{k-1}</u> Here risk is assumed to be proportional to a power of how long smoking has occurred for.

3.4.2 <u>Predictors based on simple smoking statistics</u>

At the time of writing, the intended detailed review of models other than the multistage has not yet taken place. When this has been carried out, some additional functions may be included in further work. For the moment it was decided to include five other simple statistics which might be thought to be indicators proportional (in at least some circumstances) to excess risk. If t_0 is the assumed age of starting, t is current age and L is the "lag period" (number of years before t considered irrelevant to risk), then these can be defined as follows:

<u>Av % smokers</u> The average percentage of smokers for the period $(t_0, t-L)$

<u>Av % first 10 years</u> The average percentage of smokers during the period (t_0, t_0+9)

Av % last 10 years	The average percentage of smokers during the period (t-L-9, t-L)
% 20 years ago	The percentage of smokers at year t-L-20
% dur 30+ years	The percentage of smokers of at least 30 years duration at year t-L

3.5 Sensitivity analyses

3.5.1 The basic models

For each of the predictors of absolute risk (using the Swartz submodel) and for each of the predictors proportional to excess risk (using the Swartz and Townsend submodels) a "basic" model was listed. This basic model made various assumptions.

1. Age of start of smoking = 15 (more plausible than the value of 21 used by Swartz). N.B. Age of start of smoking is the earliest age at which smoking is allowed to occur; not all subjects will start at that time
2. Number of cigarettes per day smoked by smokers = 20
3. Lag = 5 years
4. k-1 = 4.5 (k is the number of stages in the cancer process)

A number of variants from the basic model were tested by changing one of the assumptions at a time.

3.5.2 Variants to the basic model used for predictors of absolute risk

Age of start of smoking	= 18
Age of start of smoking	= 21
Number of cigarettes a day	= 30
Number of cigarettes a day	= 40

Drift (see section 3.2) = 0.5%

The variation in drift only applies to the Swartz smoking submodel.

3.5.3 Variants to the basic model used for predictors proportional to excess risk

For the multistage based predictors

Age of start of smoking	= 18
Age of start of smoking	= 21
k-1	= 3
k-1	= 6
Lag	= 0
Drift (Swartz submodel) only	= 0.5%

For the simple smoking statistic predictors the same variants were used except variations in k-1 did not apply, and variations in drift were only relevant to the duration statistic.

3.5.4 Full output

A detailed analysis was run giving the observed and predicted absolute and relative lung cancer rates for all combinations of ages, sexes, periods, smoking submodels, predictors and variants. This output is too extensive to present, but Appendix E summarizes the details of observed and predicted percentage changes over the 10 year periods. The main conclusions to be drawn these analyses are discussed in sections 4 and 5 below, principal results being shown in Tables 3-7.

4. A more general test of the claim that observed lung cancer rates have risen faster than predicted lung cancer rates - results for predictors of absolute risk

4.1 Basic model

Table 3 compares observed 10 year percentage changes in lung cancer risk by age, sex and period with those predicted using the four predictors described in section 3.3 Two clear conclusions emerge from these results.

Firstly, with only a small number of exceptions the observed changes exceed those predicted by any of the four predictors. In many cases the observed changes are substantially greater. Exceptions are for males aged 45-54 for the period 1976-85 where the decline in risk is of the same order as that predicted by the Swartz 2 predictors, and for females aged 55-64 for the period 1956-65 where the predicted rises based on British Doctors data are somewhat greater than the observed rise.

Secondly, the variation in percentage change predicted by the four predictors is usually relatively small compared to the difference between observation and prediction, i.e. the conclusions are not strongly dependent on the precise predictor used.

4.2 Variants

Table 4 compares predictions for Swartz 1 British Doctors for the basic model and the five variants considered. The effect of including drift at 0.5% was very small. Increasing the assumed minimum age of starting to smoke tended to decrease the predicted 10 year percentage changes a little, and increasing the assumed number of cigarettes smoked per smoker tended to increase the predicted 10

year percentage changes, particularly for females, but generally conclusions were unaffected. Similar trends were seen for the other three predictors of absolute risk (results not shown but included in Appendix E).

4.3 Conclusions

By generalizing the results to a variety of age, sex and period combinations, Swartz's hypothesis that lung cancer rates have risen faster than predicted on the basis of smoking habits has been given considerable support. However, the limited number of smoking models tested, the fact that they do not necessarily actually predict absolute lung cancer rates well, and the fact that no allowance has been made for any variation in values of the various fitted constants in the models according to variation in the assumed values of age of starting to smoke or number of cigarettes smoked, limit the conclusions that can be drawn. The wider range of predictors considered in the next section, and the avoidance of the problem of fitting constants (by using predictors of relative excess risk rather than of absolute risk), should mean that the results considered in section 5 are a more valid test of the hypothesis.

5. A more general test of the claim that observed lung cancer rates have risen faster than predicted lung cancer rates - results for predictors of excess risk

5.1 Adjusting rates for background

Table 5 compares percentage changes in actual lung cancer rates and in lung cancer rates adjusted for background (estimated as described in section 3.3.1). It also shows lung cancer rates

adjusted for half the background as well as giving actual values of the lung cancer rate and background at the beginning of each period considered.

For males, for all time periods and age groups the estimated background rate is a relatively small part of the total rate. As a consequence there is relatively little difference between the estimated percentage changes over 10 years in actual rates with the corresponding estimated percentage changes in rates adjusted for background. When comparing with changes in the smoking based predictors it is clear that the correctness of the background adjustment is not crucial. One can generally make similar inferences comparing with unadjusted rates, with rates adjusted for background, or even with rates adjusted for twice the background rate assumed (results not shown).

For females, the estimated percentage changes are much more strongly dependent on the assumed background rate, particularly for the earlier periods, when the estimated background forms a large proportion of the total. It is arguable that background rates derived by a formula based on male data may overestimate background rates for females. For that reason, Table 5 includes for illustrative purposes, percentage changes for rates adjusted for half the assumed background rate. The variation in percentage change for the female data for 1956-65 and 1966-75 between the full and half background adjustment underlines the sensitivity of the female percentage changes on the background rates assumed.

5.2 Basic smoking model

Table 6 compares observed 10 year changes in lung cancer risk, adjusted and unadjusted for background, by age, sex and period with those predicted using five of the predictors described in section 3.4. A number of conclusions can be drawn from these results.

First, the predicted 10 year percentage increases are always greatest for the predictors that depend heavily on smoking early in life (Av% first 10 years and multistage 1:0) and are always least for the predictors that depend heavily on smoking late in life (Av% last 10 years and multistage 0:1). Results for other multistage predictors 1:2 (results shown) and 5:1, 1:1, 1:5 (results not shown but included in Appendix E) always predict intermediate increases, with the greater the ratio of early to penultimate stage affected the greater the increase. Only in very rare circumstances did any predictor for which results are not shown in Table 6 predict an increase or decrease outside the range for the predictors for which results are shown. The most notable exception was for:

dur 30+ F 45-54 1956-65 % change = 279.4

but here the index is very unreliable due to considerable uncertainty over the number of women smoking early in life in the 1920s. The other exception was:

dur 30+ F 45-54 1966-75 % change = 36.7

but this did not affect the overall conclusions.

Second, it was generally true that, using arguably the most appropriate predictor (Multistage 1:2), the 10 year percentage change in predicted excess rates was always less than the corresponding change in observed-background rates. In two cases (F,

55-64, 1956-65 and F, 65-74, 1966-75) where observed rates were low, the difference from background had been estimated to be negative (betraying inaccuracies in our formula for background risk) but this did not affect this overall conclusion. If one, implausibly from the available evidence, assumed that smoking in the first 10 years of life completely (Av % first 10 years) or virtually completely (Multistage 1:0) determined excess lung cancer risk, some of the predicted 10 year percentage changes become closer to the observed 10 year percentage changes in actual-background rates, but even then they were nearly all lower, the only exceptions being M, 45-54, 1976-85 and M, 65-74, 1976-85.

Third, although all the smoking-based predictors tended to underestimate the percentage rise in lung cancer rates, it was clear that they did predict them to a considerable extent. Consider, for example, the 8 male estimates for actual-background and the 8 corresponding predictions for multistage 1:2. Ranking them in order of the predicted percentage change and putting the observed percentage change alongside, we have:

Predicted:	-14.3	-6.8	-2.5	1.9	6.2	9.9	16.8	20.1
Observed :	-10.1	7.7	24.7	10.1	21.5	30.6	33.4	35.3

There is quite a strong rank correlation ($r^2 = 0.93$, $p<0.001$). The correlation is also strong for females.

Predicted:	0.3	8.2	14.6	47.0	56.0	66.0	103.6	112.2
Observed :	29.1	72.6	141.3	160.1	272.5	385.3	*	*

(*Background prediction greater than observed rate.)

5.3 Variants

Table 7S (Swartz smoking submodel) and Table 7T (Townsend) show

the effect of variants considered on the ratio

$$R\% = \frac{100 \text{ Predicted excess risk of end of period/ predicted excess risk of beginning}}{\text{Observed excess risk at end of period/ observed excess risk at beginning}}$$

For example, considering the data in the first column of table 6 (M; 45-54; 1956-65; Multistage 1:2), we have

$$R\% = \frac{100 \times 109.9}{130.6} = 84.2$$

It can be seen from these tables that the shortfall of predictions compared to observation was not materially affected by:

(i) the smoking submodel,
(ii) the assumed age of starting to smoke,
(iii) the assumed value of k-1,
(iv) the assumed lag time, or
(v) the assumed amount of drift.

The same conclusion could be reached using other predictors than multistage 1:2 (see Appendix E).

It is notable, looking at Table 7S (or 7T) how relatively consistent the shortfalls are in males, with the ratios averaging about 87% for the 8 age/period combinations considered. Inverting this (1/0.87 = 1.15) implies that in each 10 year period, the rate increases about 15% more than predicted by the multistage 1:2 model. Although this percentage depends to some extent on the smoking model considered, this would seem to imply that every year lung cancer risk rises by about 1-2% more than would be explained by smoking, as taken into account in the models used.

5.4 Conclusions

The analyses described so far strongly support Swartz's

hypothesis that observed rises in lung cancer have exceeded those expected based on trends in smoking habits. By considering a variety of combinations of age group, sex and period, and a variety of different predictors of risk, these analyses help to rule out the possibility that Swartz's conclusions are some sort of artefact of the particular choice of age, sex, period, or smoking based predictor used. Three further lines of approach seem worthy of attention. One is an examination of the possibility that the Harris data may have been seriously inadequate, and that alternative sources of US data may give different conclusions. This is considered in section 7. Another is to see whether the conclusions apply to other countries where adequate smoking and mortality data are available. Some preliminary results are given in section 9, and will be extended in a later report. A third is to consider whether there are any aspects of smoking, not taken into account in our analysis, that may have biassed our conclusions. This is considered briefly in section 6.

6. Are the smoking models adequate?

6.1 Aspects of smoking other than prevalence

The smoking-based predictors used are all dependent on data solely on the estimated age and sex specific percentage of smokers at different years. They do not take into account possible trends over time in amount smoked per smoker and tar delivery per cigarette, and only partially take age of starting to smoke into account. Nor do they consider smoking of pipes or cigars.

6.1.1 Tar delivery per cigarette

Table 8 gives data on the sales-weighted average tar level of brands smoked in the US from 1957 to 19~~78~~85. Over that period the average tar level has declined almost 3-fold. It is clear from the epidemiological evidence (Lee, 1992) that tar reduction is associated with a reduced risk of lung cancer, even though smokers may "compensate" to some extent for the reduced tar by increased inhalation. Clearly had our comparisons (and those of Swartz) taken into account the tar reduction (which would be difficult to do as there is no good evidence on effects of long-term reduction) this would have only served to strengthen the hypothesis, increasing the discrepancy between observed and predicted rates.

6.1.2 Amount smoked per smoker

If there had been a marked tendency over time for number of cigarettes smoked per smoker to have increased, this might have decreased the discrepancy.

Most of the available data in the literature on this statistic was originally presented as a distribution of the percentage of smokers smoking amounts in various different categories. In International Smoking Statistics (IntSS) (Nicolaides-Bouman et al (1993)) a standard method was used to convert these to "average cigarettes per smoker", allowing easier comparison. The resulting figures are summarized in Table 9, together with some results for earlier years taken from Harris (1980).

The data from the Milwaukee studies suggest that smoking levels were low during the 1920s and 1930s - 13 for men and 7 for women in

1934. When grossed up by the US population, these figures overstate national sales by about 30%. This might suggest that the true smoking level is even lower; however these studies were not representative of whole population, being based in one urban area. Prevalence data from the first nationally representative study in 1935 (Harris 1980, quoting Fortune Magazine 1935) show overall prevalences lower and a substantial urban/rural difference. Using the Fortune prevalences in the calculation reduces the overstatement level to about 10%.

With the exception of one non-representative study in 1947 which overstated by 15%, post war studies shown in Table 9 all understate national sales by around 30-40%.

Harris (1980) (using percentage distributions for 1 survey in 1965 and 5 surveys in the 1970s, all but one of which are included in Table 9) concluded that there had been a continuing rise in smoking level. Using the IntSS results shows that the increase between about 1955 and 1980 was from about 20 to 23 cigarettes per day for males (15% increase) and from about 15 to 20 for females (30% increase). It is difficult to be certain of this due to the methodological differences between surveys. Taking into account the 60% drop in tar levels (and assuming there are no substantial differences in tar delivery for cigarettes smoked by the two sexes), this increase would in fact represent a <u>decrease</u> in total tar exposure per smoker of about 55% for men and 45% for women.

Combining together all these disparate sources, the tar-corrected consumption (35 mg tar cigarettes per smoker per day) can be estimated as approximately:

	Male	Female
1924	10	(no data)
1934	13	7
1955	20	15
1980	9	8

From these estimates lifetime average tar-corrected consumption has been calculated and is shown in Table 9B. Two alternative methods of estimating the earlier and intermediate years were used, method 1 having higher early consumption than method 2. The results show that the lifetime average consumption rose over the first 10-year period for both sexes and all ages considered, but rose only slightly or fell in the later periods. It seems that the increase over time in numbers of cigarettes smoked per smoker is unlikely to be an explanation of the discrepancy observed by Swartz and confirmed by us. However, some more work is needed to clarify this further.

6.1.3 Age of starting to smoke

Trends in age of starting to smoke over time, if they had occurred, might in theory have had a moderately strong effect on trends in lung cancer rates. If, for example, smokers aged 60 in 1975 had started to smoke on average at age 15, and smokers aged 60 in 1985 had started to smoke on average at age 14, the risk in current smokers (based on a multistage model) would, all other things being equal, have increased by a factor of $(46)^{4.5}/(45)^{4.5}$ = 1.10. Figures given by Harris (Table 10) show the mean age of starting to smoke decreasing by an average of 0.7 years per 10 calendar years of birth, and by 2.5 years for women. Particularly

for women, the rate of decrease has slowed over recent cohorts. Figures by Haenszel (1956) are similar, although the decrease is slower than Harris for men and faster for women.

In theory, the process of building up the smoking sub-model from cohort based prevalences would automatically take age of starting smoking into account, as the prevalence increases with nonsmokers gradually switching to smokers. However, there is a problem arising from the way in which the Harris data is presented.

Harris's method produced data by cohort of respondents born in successive 10 year periods. However the prevalence estimates were calculated as relevant to single years, not at a fixed age and are thus averages over persons in a 10-year wide age range. For instance the 1901-10 cohort estimate for 1930 is based on persons aged 20-29. We have followed Swartz in interpreting the 10-year cohort data as being applicable to the single-year cohort born at the mid-year, in this example the estimate is taken as applying to 25 year olds born in 1905. This seems reasonable once the whole of a cohort are adult, but is more difficult to justify at younger ages. For instance, our estimate for 15 year olds born in 1905 is Harris's average of the 1901-1910 birth cohort, in 1920, when their ages range from 10-19; it seems clear that this would not be a homogeneous group on which to base the estimate.

It is a matter of judgement as to how low an age the Harris data should be used in the smoking submodel. Swartz used age 21 (with a variant model of 18) but gave no indication of the reason for this choice (indeed he may not even have considered this aspect of the problem). As already discussed (section 3) we have used 15 as

our basic model with variants of 18 and 21 and the effects of this were discussed in sections (4.2).

Thus any changes in the smoking pattern below age 15 are ignored in the smoking model. Although the age of starting to smoke is decreasing the average nevertheless remains above 15, and therefore any bias would be considerably less than in the theoretical example cited earlier in this section. Table 10 also shows the average age of starting derived from the Swartz smoking model. (Results for the Townsend smoking model, not shown, are virtually identical.) These results confirm that, when using Harris data from age 15, the smoking model gives average starting ages only slightly higher than the Harris originals for males. Curiously, the values for females are slightly lower, for which there seems no theoretical explanation.

6.1.4 Other tobacco products

Sales data for tobacco products other than manufactured cigarettes exist spasmodically from 1900 and then annually from 1920, although they are difficult to interpret as pipe, hand-rolled (cigarette) tobacco and chewing tobacco are only available as a combined group until 1949. However it is clear that they have become progressively less important compared to cigarettes. In 1900 the number of cigars sold was more than twice the number of manufactured cigarettes sold. This ratio had fallen to about one fifth by 1920 and has been less than one fiftieth since 1950.

Assuming that the proportions of tobacco used for pipes, hand-rolled cigarettes and chewing tobacco were the same as in 1949,

the consumption of pipes has also fallen steadily. In 1900 the weight consumed of pipe tobacco was more than 10 times that of manufactured cigarettes. This ratio had fallen to about 1½ in 1920, to one tenth in 1950 and to one fiftieth in 1985.

It is clear that taking into account consumption of pipes and cigars (which predominantly occurs in men and in older age groups - see IntSS) would only serve to increase the discrepancy between observed and smoking-predicted trends in lung cancer rates, not to explain it.

6.1.5 Conclusion

Overall it can be concluded that aspects of smoking other than prevalence cannot explain the tendency for the observed trend in lung cancer to have risen faster than that predicted by the smoking models we have used. There has been a substantial decrease over time in age at starting to smoke, but this has essentially been taken account of in our comparisons. The apparent early increase in number of cigarettes smoked per smoker has eventually been compensated for by the later large decline in average tar levels. Thus this does not seem to be a potential explanation for the difference between observed and predicted trends for the later periods studied, particularly not for the youngest age group whose smoking careers would only have started around 1950. However, it may have affected results for the earlier periods/older age groups studied. For males, it would also have been offset by the higher levels of smoking of other products in early years. Further work could be conducted to try to account for tar levels and number

smoked in the predictions, although it is clear already that for some age groups/periods this would only serve to enlarge the difference between observation and prediction.

6.2 Do the smoking models give plausible results?

Some detailed tables on the working of the smoking models have already been given (section 3.2, Appendix C). Clearly any such model will be a simplification of the true picture and cannot reflect such aspects as occasional smoking (either by young people before starting, or by ex-smokers) or short periods of quitting smoking.

Analysis by Cummings (1984) suggests that discontinuous smoking periods (not allowed in the Swartz model) are common in reality. Based on the 1978 NHIS, he reported that about 60% of current smokers had made at least one serious attempt to quit smoking in the past. About 30% of smokers make a serious attempt to quit smoking each year, but only about 20% of these succeed. Similarly, the Adult Use of Tobacco Survey in 1970 (USDHEW 1973) found that 49% of current smokers and 44% of former smokers had made at least one (unsuccessful) attempt to quit in the previous 5 years, with 29% and 17% respectively trying more than once. Although we are not aware of any data on the length of "quit periods", it seems likely from the high frequency of quit attempts that periods of a year or more are not negligible and should therefore feature in the model. Although possible under the Townsend model, "quit periods" occur only rarely

in practice. This is because they are caused in the model by the smoking prevalence falling then rising, which rarely happens in the fairly smooth patterns of the Harris data.

A feature of the Townsend model is that the percentage of the cohort who have ever smoked is constrained to be the same as the maximum percentage smoking at any one time. In a cohort-based analysis of smoking in Norway (similar to Harris for the US) Rønneberg et al (1994) gave data on ever smokers for cohorts born 1890-1939. For the female cohorts born up to 1919, the maximum percentage did equal the percentage of ever smokers, but in all male cohorts and in the later female cohorts it was between 5 and 10 percentage points lower. This implies that the model should involve some element of "drift". However Swartz's drift model is implausible in that the drift continues at the same rate right through into old age, and the fact that the average age of starting predicted by this model (Table 10) is much higher than the original confirms this.

The two rules for selecting which smokers give up when prevalence drops are opposite extremes - with Townsend only those with shortest duration give up whereas with Swartz all smokers are equally likely to give up. The Swartz method is supported by Haenszel et al (1956) who reported from the 1955 CPS Survey that the percentage of former smokers did not vary greatly by age of starting to smoke.

A more radical approach is to consider whether a smoking model is necessary at all. Where prevalences have been derived from series of surveys carried out in successive years (as with the IntSS

or Tobacco Advisory Council (TAC) data discussed in sections 7.2 and 7.3) then it is certainly necessary. However where the prevalences have been derived from smoking histories, the smoking model is really only trying to recreate the original data. If access to the original data were possible, risk assessments could be made directly.

7. Are the Harris data adequate?

7.1 Bias due to differential mortality in smokers

The Harris data used in Swartz (1992) was based on smoking histories of respondents in the 1978-80 Health Interview Surveys. Thus only persons who survived to 1978/80 were available to give estimates of earlier consumption. Since cigarette smokers have higher mortality than nonsmokers, such estimates would theoretically understate past prevalences of the whole population. Harris presented a method of correcting this source of bias, based on standard life table methods. Results were given in his Text Figures 3, 4, for ages 35+. The main effect of correction for differential mortality is to increase the prevalence estimates for men born before 1910. However, Swartz chose to use the uncorrected data from Harris.

To investigate this possible bias further, we considered data provided by Hammond (1969) giving life tables for lifelong nonsmoking men and for current smokers of 20-39 cigarettes a day. Starting with a population which consisted of 50% of each of these two groups at various different ages, we estimated the percentage which would be observed at various different times later (Table 11).

It can be seen that the major determinant of the observed percentage is the age of the cohort at follow-up. When considering subjects aged less than 50 at follow-up, the bias in estimating the percentage earlier in life is very small (<1%). For subjects in the 50-60 range at follow-up it is of order about 2%, while for subjects in the 60-70 range at follow-up it is of order about 5%. Of course, these calculations are approximate (we really need life-table data comparing all current cigarette smokers with all non cigarette smokers including ex-smokers, but they give a fair idea of what is going on). Provided we limit attention to subjects aged up to 70 at survey, this bias should not be too important.

Swartz's subjects were born 1900-1943 (age 42-70 in 1970-85) and some were therefore over 70 at the time of survey, as were the earliest born groups in our analyses (55-64/1956-67 and 65-74/1966-75, born 1892-1910; and some of 45-54/1956-65, 55-64/1966-75, 65-74/1976-85, born 1902-1920).

7.2 Can past prevalences of cigarette smoking be estimated retrospectively?

Another potential problem with basing prevalence estimates on smoking histories is that such recall may be inadequate. To gain insight into the validity of this approach, we compared estimates of past percentages of smokers based on smoking histories given by respondents in the 1984/85 UK Health and Lifestyle Survey (HLS) with percentages of smokers reported in surveys carried out annually by Research Services for ITL from 1948 onwards. Appendix F describes the results of this comparison in detail.

The percentages of male smokers in recent years estimated from the two sources are quite close, but for earlier years (1970 and earlier) and for all years for females, the estimates based on HLS are generally lower, by up to 10%, than the TAC estimates. However, there is no clear time trend, and so no indication that the differences would become larger were TAC data available in yet earlier years. Overall, the magnitude of the differences seems not unacceptably large.

As another approach, we used Harris's prevalence data combined with the assumption that smokers smoke 20 cigarettes per day to estimate the total national consumption. Harris's data for ages 15 and above was used and the methods for estimating prevalences for intermediate cohorts were the same as described in section 2.4. The age range covered by Harris decreased progressively in earlier years (up to age 95 in 1980, 85 in 1970, etc.) and the weighting method developed in IntSS Appendix IV was used to extend prevalences to the full age range. It was also used to estimate prevalence at ages 12-14, and at the younger ages in recent years not covered by Harris.

The results (Table 12) show that Harris's data accounted for around 80-85% of sales in the 1950s, falling to 72-73% in the 1970s. This is broadly in line with the general finding that, when grossed up, survey data almost invariably understate total sales. In fact, these results are closer to 100% than most of the US surveys assessed in IntSS, where results were mostly around 60-70% (see IntSS Tables 22.6-8). Had a lower smoking level been assumed for females, (suggested by Table 9 and by the general findings in IntSS

pxxx) then these current results would also have been lower. The trend to more serious understatement over the last 20 years appears to fit in with smokers tending to smoke more cigarettes per day in recent years, as discussed already in section 6.1.2.

7.3 Using an alternative source of data

7.3.1 Data available in International Smoking Statistics

Unlike the situation in the UK, consistent nationally based series of smoking statistics did not begin in the US until the later 1950s. In IntSS, data were gathered together from several individual surveys from the 1930s to the 1950s, and a number of major sources since. A method was developed (see IntSS Appendix IV and Supplement) which enabled estimation of the prevalence of smoking in standard 5-year age groups, for 5-year periods. For the US, the estimates (IntSS Suppl. Table 10 (TC/MC) p.56) start with the period 1931-35 and are therefore a sufficiently long series to be an alternative source of data for the smoking model.

However, it should be noted that the IntSS estimates are on a fairly weak basis in the early years. The following surveys contribute to the 1930s and 1940s estimates:

1935 Fortune, age bands 20-39, 40+.

1944 Gallup, all ages 18+ combined

1947 Hamtoft and Lindhard, ages 20-29, 30-39 ... 60-69, 70+, whites only in Columbus, Ohio

1949 Gallup, all ages 18+ combined.

The early estimates are heavily dependent on the age structure of the weighting system used to generate them (this having been derived

from available surveys in a number of countries, as described in IntSS Supplement). Differing methodology of the various surveys has not been taken into account, and is, in any case, unknown for the early surveys.

7.3.2 Comparison of Harris and International Smoking Statistics

The IntSS data are based on prevalences for 5-year periods by 5 year age groups. To convert this to a cohort basis, we have simply taken entries from the diagonals of the table so that, for instances, 15-19 year olds in 1931-35 comprise persons born 1912-1920 and are taken to represent the cohort born in the mid-year 1916. Overlapping of successive cohorts (e.g. 15-19 year olds in 1936-40 were born 1917-1925) has been ignored.

These data are shown in Table 13, together with differences from the nearest equivalent Harris data.

For males, the Harris prevalence estimates are consistently higher than the IntSS estimates, generally by 2-8 percentage points, but there are some larger differences, in the earlier cohorts compared (1915, 1925).

For females, the Harris prevalence estimates are consistently lower than IntSS estimates at younger ages, implying a slower take-up of smoking (older average age of starting to smoke). For the earliest cohort compared (1915), this difference persists into middle age, but for later cohorts, all Harris estimates over age 25 are 1-5 percentage points higher than the IntSS estimates, similar to but smaller than the results for males.

No comparison is possible with Harris's earliest cohorts (1885-1905).

It can be noted that, apart from the results for young women, these Harris/IntSS differences are in the opposite direction to the UK equivalent HLS/TAC differences (section 7.2).

7.3.3 Methods of using International Smoking Statistics data in smoking models

The weighting method described in IntSS Appendix IV was used to convert the estimates as presented in Table 13 for ages 15-19 into single years of age, and 20-24 into 21, 22-24. Other single year estimates were assumed equal to the estimate from the wider age group. Methods used were then the same as for the Harris data (section 2.4) except that, since the cohorts were 5 rather than 10 years apart, and since the data existed up to 1985 (instead of 1980), extrapolation was a less important feature of the method.

By extending back to the 1912 cohort, the data were sufficient to allow 3 of the original 8 age/period combinations to be studied:

Age	Period
45 - 54	1966-75
45 - 54	1976-85
55 - 64	1976-85

Two alternative methods were used for the 1912-1915 cohorts (not relevant to 45-54/1976-85):

a) Prevalence assumed to be the same as for the same age in the 1916 cohort

b) Prevalence estimated by linear extrapolation between 1916 and 1921 cohorts, within each individual age.

7.3.4 Results

Results are shown in Table 14 for the basic Swartz model using method a. (Method b and the Townsend model are included in Appendix E).

For males, for two of the three age/period combinations studied, the percentage changes over 10-year periods of almost all the indices studied are similar to those predicted using the Harris data, and therefore lower than the percentage changes in actual (or actual-background) rates. For age 45-54/1976-85, where with Harris there had been a fairly small difference between actual and predicted, that difference has generally disappeared with IntSS.

For females, the percentage changes are generally much lower than those predicted using Harris, and thus the differences between actual and predicted are even more substantial. However, there is greater variability between methods and models, which reduces confidence in the results.

Tables of rates (not shown) for females suggest that the cohort with peak predicted risk was born earlier (around 1927) according to the IntSS based analysis, than according to the Harris-based analysis (around 1935). This reflects the differences in uptake of smoking commented on in section 7.3.2

7.3.5 Future work using International Smoking Statistics data

In order to study a more useful range of ages/periods it would be necessary to extend back to earlier-born cohorts. The variability in results between the two simple methods used to extend

back by 3 single-year cohorts has demonstrated how this can have a substantial effect on results. It is planned to study more sophisticated methods, such as the Age-Cohort model used to extrapolate back smoking levels in the UK (Lee et al, 1990). However in view of the absence of comprehensive sales data before 1920, and the changes in population base associated with immigration and boundary changes, it is unlikely that satisfactory estimates could be made for many more years.

8. Trends in nonsmokers lung cancer rates

One of the most direct methods of obtaining evidence on whether factors other than smoking are playing an increasing role in the aetiology of lung cancer is to study trends over time in the risk of lung cancer among lifelong nonsmokers. Appendix G summarizes the evidence on this. The studies providing the most direct observations of trends in nonsmokers' lung cancer rates do not suggest that any obvious increase in risk has occurred since the second World War, although the possibility of a modest increase is not ruled out, especially in Japan. A number of papers have estimated trends indirectly, and have claimed large increases in risk in nonsmokers. However, most such studies tend to have obvious technical weaknesses and be difficult to interpret. A recent paper by Forastière et al (1993) is perhaps the most interesting of these papers, and will be considered in more detail when we come to investigate trends in Italian data. Overall it must be concluded that the evidence

considered in Appendix G does not provide any clear demonstration that lung cancer death rates in nonsmokers have actually increased in recent years.

9. Other countries

As we have available mortality and population data from WHO for a number of countries, our methods are readily applied elsewhere if suitable smoking data is available.

Cohort-based data have been published for Italy by La Vecchia et al (1986) and for Norway by Rønneberg et al (1994). Unlike Harris, where data were presented graphically for 10-year cohorts at each individual year, the Italian data are given for every 10th year and the Norwegian data are given for 5-year cohorts as averages over 5-year age groups. These have been transformed into single year estimates using the weighting method developed in Appendix IV of IntSS. However, the years involved are well outside the period originally considered in IntSS and this process requires more detailed consideration. Another problem is that in these smaller countries numbers of deaths are low and rates based on single years are not stable, particularly for younger women. Hence comparison with 10-year changes in actual rates may not be appropriate. The original data are given in Appendix H and preliminary results (using the basic Swartz model) are given in Table 15 (Italy) and Table 16 (Norway).

Results for males in both countries show a similar picture to the US results, with predicted 10-year percentage changes lower than actual (or actual-background) changes for nearly all indices in all

age/period combinations. Exceptions were

Norway, age 45-54, 1976-85

Italy, age 45-54, 1956-65 and 1976-85

where the predicted and actual changes were of similar magnitude.

For females in Norway the predicted changes were also lower than the actual changes for the later periods studied, but were higher in the first period (1956-65). For females in Italy, predicted changes were generally higher than actual changes, but with some exceptions in the latest period.

More work in this area is planned.

10. Discussion

10.1 Summary of main conclusions

Swartz (1992) observed that, in the US, male lung cancer rates, among the age group 42-70, had risen by 26% over the period 1970 to 1985. This rate contrasted with a 12% decline in lung cancer which he estimated should have occurred, based on trends in cigarette smoking habits. His findings suggested implicitly that the effect on lung cancer risk of trends over time in factors other than smoking may be of considerable importance.

In this report we have not attempted to study what factors other than smoking might have caused the discrepancy between the observed and smoking-predicted trends in lung cancer rates. Rather we have attempted to try to evaluate how reliable Swartz's conclusion of a discrepancy actually is, by investigating how much

it depends on various aspects of the analysis he undertook. As described below, our own analyses in the main strongly support Swartz's conclusions that there is an unexplained discrepancy.

We have shown clearly that the discrepancy exists over a wider time period (1956-1985) than used by Swartz, and also that it exists for females, not studied by Swartz. Furthermore the discrepancy is generally evident within 10 year age groups (over the range 45-74) and for successive 10 year time periods. Of 16 age/period/sex combinations studies, 14 showed this discrepancy, with only two (males aged 45-54 in 1976-85, and females aged 55-64 in 1956-65) showing a reasonable correspondence between observed and predicted trends.

It also seems clear that the discrepancy is not contingent on the exact form of the mathematical model used to relate smoking history to lung cancer risk, or the fact that Swartz had inadvertently used a function which did not actually correspond to that which Whittemore (1988) had recommended. We used a number of functions which might be expected to be reasonable indices of smoking-related lung cancer excess risk, some based on the multistage model (which we reviewed in detail finding considerable evidence in its support) and some based on simpler statistics. Although the discrepancy was weakened for statistics which gave much more importance to smoking early in life than to smoking later in life, it was in most analyses evident even then. For statistics which, more plausibly from the existing evidence, gave more comparable weight to smoking over the whole time period, the discrepancy was generally evident for all age/period/sex

combinations studied. Making plausible variations to various underlying parameters of the models used (e.g. number of stages of the multistage model assumed, minimum lag time between final exposure and onset of cancer) also did not affect our conclusions. Although, at this point in time, we have not yet reviewed in detail mathematical models of carcinogenesis other than the multistage, we feel it unlikely that alternative functions will provide different conclusions.

Given data on smoking prevalence at various ages, some assumptions have to be made to construct the distribution of the population starting and stopping smoking at various times. Swartz used one simple alternative which only allowed one smoking period per person, and tended to minimize the estimated number with a long duration of smoking. We investigated an alternative, based on the work of Townsend (1978), which allowed more than one smoking period per person, and tended to maximize the estimated number with a long duration. While it is evident that both alternatives are gross over-simplifications, the very fact that they are relatively extreme alternatives and gave very similar results tends to argue that this is not a reason for the discrepancy.

The adequacy of the actual smoking prevalence data derived by Harris and used by Swartz has been explored in a number of ways. These data were derived retrospectively from surveys conducted in 1978-80, and the estimates may be biased due to the differential mortality suffered by smokers and nonsmokers and by poor recall of past smoking habits. Using theoretical calculations based on the life-tables of smokers and nonsmokers we have demonstrated that

differential mortality is unlikely to be of any consequence except for those aged over 70 at the time of survey (i.e. born before 1910) and therefore cannot explain the discrepancy in the later-born groups studied. Moreover, for these later-born groups, the results have been confirmed by using the alternative data derived from IntSS based on contemporary surveys. The exception is the latest-born group of males (age 45-54/1976-85) who showed only a small difference between observed rates and Harris-based predictions, and even less with IntSS-based predictions. The use of contemporaneous surveys avoid the problem of recall bias. Comparisons between the Harris and IntSS data in the US, and between the HLS and TAC data in the UK, have both shown a reasonable level of consistency, and suggest that overstatement of past smoking habits at the expense of current smoking habits is not an explanation of the discrepancy pointed out by Swartz between observed and smoking-predicted lung cancer rates. More generally, though there may be weaknesses in the Harris data, they do not seem to provide any reason for this discrepancy.

We have considered the possibility that inadequate accounting for various aspects of the smoking habit other than smoking prevalence might have caused the discrepancy. Age of starting to smoke does not seem to be a problem in this respect since the Harris data, and the way we have incorporated them into our analyses, essentially already take into account the fact that, over the last century, US smokers have tended to have started smoking earlier. Following Swartz, we have not formally attempted to take into account the marked reduction in the tar level of cigarettes that

started in the 1950s. Had we done so, it is clear the discrepancy would have become greater not smaller. It also seems that the tendency over time for smokers to be more likely to smoke cigarettes and less likely to smoke pipes and cigars would, if taken into account in the analysis, have tended to increase rather than decrease the discrepancy. Number of cigarettes smoked per smoker is, however, a factor that might explain some of the discrepancy. The models used by Swartz assumed a constant smoking level, and though formulae based on the multistage model can be derived to take into account varying exposure, the ones used in this report have not done so. It is not straightforward to estimate what effect taking into account number of cigarettes per smoker might have. Since 1955 the increase has been quite small and has clearly been more than compensated by the reduction in tar level (even allowing for the fact that tar levels as measured under standard smoking conditions may not reflect tar intake by the smoker). Between the 1920s and 1950s, however, where tar levels have essentially been unchanged, there appears (though actual survey data are limited) to have been a substantial increase in the number of cigarettes smoked per smoker. The overall effect of the increase in tar per smoker up to about 1955, followed by a decrease, is complex and demands further attention. It seems unlikely, however, that it could explain the whole discrepancy observed, particularly as some of our analyses demonstrated the discrepancy to exist for populations where most smoking occurred after 1955.

10.2 Possible further work

10.2.1 USA

As noted in the previous paragraph, the smoking-based predictors we have used have not taken into account tar level and number of cigarettes smoked. Although historical data on both are somewhat limited, we intend to extend our work by studying some predictors that do take them into account.

Another area which seems worth pursuing is to extend the estimations of risk based on the IntSS data. Given the available smoking prevalence data and the earlier historical data on sales, it should be possible to construct smoking history estimates which are totally independent of the Harris data. Although this work may involve assumptions that are difficult to justify fully, so that early estimates of prevalence by age and sex may be open to criticism, they will avoid the problems of recall bias and differential mortality inherent in the Harris data. If the discrepancy remains evident using two sources of data, each with their own strengths and limitations, this will give further support to the hypothesis that Swartz put forward.

The Harris paper started with data from the Health Interview Surveys, consisting of smoking history information for each member of the population studied, and then converted it into estimates of smoking prevalence at different ages in different cohorts. Swartz took this prevalence data and, via certain assumptions, attempted to regenerate the smoking history information on an individual person basis in order to compute the lung cancer risk estimates. It would be technically far superior to use the original Health Interview

Survey smoking histories directly to compute the risk estimates. I understand, from an Office on Smoking and Health fact sheet (Appendix J), that these data are publicly available. There is an obvious case for trying to get hold of these data for further analysis.

10.2.2 UK

We have available on our computer data from the UK HLS and also from the TAC Alderson Hospital Case-Control Study giving detailed smoking data, each on a reasonably large population. The UK HLS is representative and provides data on age of starting, age at stopping (for ex-smokers), and number smoked. The controls from the Alderson study are less representative (10 areas in England and Wales) but have more detailed data, including changes in number smoked and brand smoked. One or both of these data sets could be used to produce smoking-based predictors of trends in risk which could be compared with trends in observed risk from national statistics.

Both the above studies would involve potential problems of recall bias and bias due to differential mortality. An alternative approach would be to use the TAC survey data for the UK published in IntSS. These survey data go back to 1948 and could be used directly to provide risk estimates for cohorts born from 1933. Backward extrapolation, using procedures analogous to those already developed to provide historical data on consumption per adult by age and sex (used in Lee et al (1990)), could be employed to provide risk estimates for earlier cohorts.

10.2.3 Other countries

Preliminary results for Norway and for Italy have been presented in this report, based on other authors' published estimates of smoking prevalence. More work is needed on these data, particularly for Norway where the small numbers of deaths in a year require the development of additional techniques to get a more reliable estimate of trends in observed rates.

We have not attempted at this stage to use IntSS data for these, or other, European countries. Preliminary work needs to investigate the best methods of obtaining historical smoking prevalence estimates.

10.2.4 Discussions with Swartz

As noted above, some details of Swartz's original paper still need resolution. It remains unexplained why we were unable to reproduce his results. Swartz has expressed interest in a possible collaboration. A first move might be to send this report to him for his comments. If this proves fruitful, a meeting might be advantageous.

11. References

Brown CC, Chu KC. Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. J Chronic Dis 1987;40(Suppl 2):171-9.

Cummings KM. Changes in the smoking habits of adults in the United States and recent trends in lung cancer mortality. Cancer Detect Prev 1984;7:125-34.

Doll R, Peto R. Mortality in relation to smoking: 20 years' observation on male British doctors. Br Med J 1976;ii:1529-36.

Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationship among regular smokers and lifelong nonsmokers. J Epidemiol Community Health 1978;32:303-13.

Forastière F, Perucci CA, Arca M, Axelson O. Indirect estimates of lung cancer death rates in Italy not attributable to active smoking. Epidemiology 1993;4:502-10.

Haenszel W, Shimkin MB, Miller HP. Tobacco smoking patterns in the United States. Public health monograph no 45. Washington DC: United States Government Printing Office, 1956.

Hammond. Life expectancy of American men in relation to their smoking habits. JNCI 1969;43:951.

Harris JE. Patterns of cigarette smoking. In: US Surgeon General, The health consequences of smoking for women. Washington DC: US DHEW, 1980:15-42.

Harris JE. Cigarette smoking among successive birth cohorts of men and women in the United States during 1900-80. JNCI 1983;71:473-9.

La Vecchia C, Decarli A, Pagano R. Prevalence of cigarette smoking among subsequent cohorts of Italian males and females. Preventive Medicine 1986;15:606-13.

Lee PN. Environmental tobacco smoke and mortality. Basle: Karger, 1992.

Lee PN, Fry JS, Forey BA. Trends in England and Wales Lung Cancer, Chronic Obstructive Pulmonary Disease and Emphysema Death Rates 1941-84 and Their Relation to Trends in Cigarette Smoking. Thorax 1990;45:657-65.

Milwaukee Journal. Consumer analysis of the greater Milwaukee market. Milwaukee Journal 1924-79.

Nicolaides-Bouman A, Wald N, Forey B, Lee P. International smoking statistics. A collection of historical data from twenty-two economically developed countries. London, Oxford, New York, Tokyo: Wolfson Institute of Preventive Medicine and OUP, 1993.

Rønneberg A, Lund KE, Hafstad A. Lifetime smoking habits among Norwegian men and women born between 1890 and 1974. Int J Epidemiol 1994;23:267-76.

Royal College of Physicians. Smoking <u>or</u> health. A report of the Royal College of Physicians, London, 1977.

Swartz JB. Use of a multistage model to predict time trends in smoking induced lung cancer. J Epidemiol Community Health 1992;46:311-5.

Townsend JL. Smoking and lung cancer. A cohort data study of men and women in England and Wales 1935-70. J R Stat Soc (Ser A) 1978;141:95-107.

USDHEW. Adult use of tobacco 1970. USDHEW, 1973.

U.S. 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;DHSS (CDC):89-8411.

Whittemore AS. Effect of cigarette smoking in epidemiological studies of lung cancer. Statistics in Medicine 1988;7:223-38.

World Health Statistics Annual. Geneva: World Health Organisation, Successive years.

TABLE 1
Estimates of prevalence of cigarette smoking in US from Harris

Male

	Cohort													
Age	1885	1895	1905	1905S	1915	1915S	1925	1925S	1935	1935S	1945	1945S	1955	1955S
15	5	10	19		22		22		22.5		20		18	
16	7	13	24		26.5		27.5		28		26		24	
17	11.5	16.5	29		32.5		35.5		33.5		31		29	
18	15	20	34		38		43		39		37		33.5	
19	17	23.5	36.5		44		49		54		41.5		37.5	
20	17	27.5	40		49		55		50		46		41	
21	20	32.5	45		55		60		55		50.5		42.5	
22	25	37	49		58		64		59		55		43.5	
23	26	41	53		62		66.5		61.5		57		43	
24	30	44	55		63.5		68		63		57.5		43	
25	31	45	57	57	65	65	69	69	63.5	63	58	57	45	43
26	31	46	58		66.5		69.5		64		57.5			
27	32	47	59		67.5		69.5		64		57			
28	31	47.5	59.5		68.5		70		63.5		56			
29	31	48.5	60		69		69.5		63		55			
30	32	49	60.5		69		69.5		61.5		54			
31	32	48.5	60.5		69		69		61		52			
32	32	48.5	61		69		69		60		50.5			
33	34	49	61.5		69		69		59		49			
34	34	49.5	61		69		68.5		57.5		47			
35	34	49	61.5	61	68	68	67	67	55	57	45.5	46		
36	34	48.5	61.		67.5		66.5		54.5					
37	34	48.5	61		67.5		66		54					
38	34	48.5	61		67.5		65		53					
39	34	48.5	60.5		67		64		52					
40	36	48.5	60		66		63		50.5					
41	36	48.5	60		65.5		62		49					
42	34	49.5	59.5		65		61		47.5					
43	34	49	59		65		60		46.5					
44	34	49	58.5		64		58.5		45					
45	36	47.5	57.5	57	62.5	63	56	56	45	45				
46	36	47	57		61		54							
47	34	47	57		60.5		53							
48	34	46.5	56.5		60		52							
49	34	46	56		59		51							
50	34	45	54.5		57		49.5							
51	34	45	53.5		56		48							
52	34	45	53.5		55		47							
53	34	45	53		54		45.5							
54	33	44	52		52		44							
55	33	43	49	49	49.5	50	43	43						
56	31	45	48		47.5									
57	31	41	47.5		46.5									
58	31	41	47		45.5									
59	31	40	46		44.5									
60	29	37.5	44		43									
61	29	37.5	43		41									
62	29	37.5	42		38.5									
63	29	37	41		37.5									
64	28	36	39		34.5									
65	26	34	37	37	32.5	33								
66	25	32	35											
67	25	315	33											
68	25	31	32											
69	25	29.5	30											
70	25	28	29											
71	25	26.5	27											
72	25	26.5	26											
73	25	26	25											
74	21	24	22											
75	17	21.5	21	21										

Note. Cohorts marked S are comparable data taken from Swartz Table II

TABLE 1 (cont)
Estimates of prevalence of cigarette smoking in US from Harris

Female

	Cohort							
Age	1885	1895	1905	1915	1925	1935	1945	1955
15	0	0.5	1	5	6	9	11	13
16	0	0.5	1.5	7	12	13	15	17
17	0	0.5	2	9	12	16.5	19	22
18	0	0.5	3	12.5	16	20.5	24	27
19	0	1	4	15	20	25	27	30
20	0	1.5	5	18	23.5	29	31	34
21	0	2	6	21	27	33	34.5	37
22	0	3	8	23.5	30	36	37.5	38
23	0	3.5	9	26	34	39	40	38
24	0	4	10	27.5	35	42	40	37
25	0	4	12	29	37	43	41.5	37
26	0.5	4.5	13	30	38	44	41	
27	0.5	5	14	31.5	39	44	41	
28	0.5	5	15	33	40	44.5	41	
29	2	5	16.5	34	41	44.5	40.5	
30	2	6	17	35	41.5	45	40	
31	2	6	18	35.5	42	44.5	39.5	
32	2	7	18.5	36	42.5	44	39	
33	2.5	7	19	36.5	43	44	39	
34	2.5	7.5	19.5	37	43	43.5	37	
35	2.5	8	20	37	43	42.5	35	
36	2.5	8	20	37	42.5	42		
37	2.5	8.5	21	37	42.5	42		
38	2.5	8.5	21.5	37.5	42.5	42		
39	2.5	8.5	22	37.5	42	41.5		
40	2.5	8.5	22	37.5	41.5	41		
41	2.5	9	22	38	41	40		
42	2.5	9	22.5	38	41	39		
43	2.5	9	22.5	38	40.5	38.5		
44	2	9.5	23	38	40	37		
45	2	9.5	23	38	39.5	36		
46	2	9	23	37.5	39			
47	2	9.5	23	37.5	38.5			
48	2	10	23	37.5	38			
49	2	10	23	37.5	37.5			
50	2	10	23	37	37			
51	2	10	23	36.5	36			
52	2	10	23	36.5	35.5			
53	2	9.5	23	36	35			
54	2.5	9.5	23	35	33.5			
55	2.5	9.5	22.5	34	33			
56	2.5	10	22	33.5				
57	2.5	10	22	33				
58	2.5	10	22	32.5				
59	2.5	10	22	32				
60	2.5	9.5	21.5	31.5				
61	2.5	9.5	21.5	29.5				
62	2.5	9.5	21	29				
63	2.5	9	20.5	28				
64	2.5	9	20.5	27				
65	2	9	19.5	27				
66	2	9	19					
67	2	9	18.5					
68	2	8.5	18					
69	2	8.5	17.5					
70	2	8.5	17					
71	2	8.5	16					
72	2	8.5	15.5					
73	2	8.5	15					
74	2	8.5	15.5					
75	2	8	17					

TABLE 2

Comparison of Swartz's observed and predicted lung cancer relative rates for US males with those that we derived

	Source			
	Swartz		Lee/Forey	
Year	Actual rate	Predicted	Actual rate	Predicted
1970	100	100	100	100
1971	103	100	100.2	100.6
1972	106	99	103.7	101.0
1973	107	99	104.6	101.3
1974	110	98	106.7	101.4
1975	112	97	106.9	101.4
1976	114	96	108.3	101.2
1977	116	95	109.5	100.9
1978	119	94	111.7	100.5
1979	120	93	112.2	100.1
1980	122	92	113.3	99.5
1981	122	91	113.4	98.8
1982	124	90	114.1	98.0
1983	124	89	112.2	97.0
1984	125	89	112.7	96.0
1985	126	88	112.2	94.8
Rise	+26%	-12%	+12.2%	-5.2%

Note: Rates normalized so that the 1970 rate equals 100. All rates age-adjusted to 1970 US age distribution. Predicted rates based on Swartz's formula (1).

TABLE 3

Predictors of absolute lung cancer risk
Comparison of observed and predicted 10 years percentage changes in risk for various age, sex and period combinations

Male

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4
Predicted								
Swartz 1 Brit Docs	9.3	-1.9	-13.3	19.5	5.7	-6.4	16.3	1.0
Swartz 1 US Vets	8.1	-2.0	-12.1	16.6	5.0	-6.0	13.8	0.8
Swartz 2 Brit Docs	10.1	0.5	-9.2	19.4	8.2	-2.3	17.6	6.0
Swartz 2 US Vets	9.0	0.4	-8.3	17.4	7.5	-2.1	16.1	5.6

Female

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Observed	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Predicted								
Swartz 1 Brit Docs	50.8	13.2	1.4	64.7	47.5	8.0	65.4	40.9
Swartz 1 US Vets	39.1	10.7	0.7	46.2	36.6	6.3	46.7	31.6
Swartz 2 Brit Docs	50.9	14.7	4.0	67.0	48.5	11.3	70.6	44.6
Swartz 2 US Vets	38.9	12.2	3.4	47.7	39.4	9.8	53.0	37.7

TABLE 4

Swartz 1 British Doctor's Model
Effect of varying assumptions
on predicted 10 year percentage change in risk

Male

Age	45-54			55-64			65-74	
Period	1956	1966	1976	1956	1966	1976	1966	1976
	1965	1975	1985	1965	1975	1985	1975	1985
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4
Predicted								
BASIC	9.3	-1.9	-13.3	19.5	5.7	-6.4	16.3	1.0
F18	9.1	-2.0	-13.3	17.7	5.5	-6.4	14.7	0.9
F21	8.4	-2.6	-13.3	16.0	5.0	-7.0	13.2	0.4
N30	10.1	-1.8	-14.0	21.6	6.2	-6.5	18.1	1.2
N40	10.6	-1.6	-14.3	23.0	6.5	-6.6	19.2	1.3
D005	9.2	-2.0	-13.2	19.0	5.6	-6.4	15.7	1.0

Female

Age	45-54			55-64			65-74	
Period	1956	1966	1976	1956	1966	1976	1966	1976
	1965	1975	1985	1965	1975	1985	1975	1985
Observed	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Predicted								
BASIC	50.8	13.2	1.4	64.7	47.5	8.0	65.4	40.9
F18	48.4	12.3	0.7	63.5	45.3	7.2	64.2	39.0
F21	45.0	11.4	-0.2	60.9	42.0	6.4	61.7	36.1
N30	61.2	15.2	2.2	82.3	56.9	9.4	82.7	48.6
N40	68.8	16.6	2.8	94.9	63.4	10.4	94.8	53.7
D005	49.8	12.9	1.2	63.3	46.1	7.7	63.3	39.3

Note: F = first year of smoking
N = number of cigarettes per day
D = drift

TABLE 5

Effect of adjustment for background on lung cancer rates and on 10 year percentage changes in lung cancer rates

Male

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Rate (per million per year) at beginning of period								
Observed	458	598	737	1215	1665	2031	2811	3720
Background	54	54	54	131	131	131	275	275
Percentage change over 10 years								
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1

Female

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Rate (per million per year) at beginning of period								
Observed	80	141	281	147	242	579	329	715
Background	54	54	54	132	132	132	278	278
Percentage change over 10 years								
Observed	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	385.3	141.3	29.1	---	272.5	72.6	---	160.1

TABLE 6

Predictors of excess lung cancer risk
Comparison of observed and predicted 10 year percentage changes in risk from basic model for various age, sex and period combinations

Male

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1
Av % first 10 yrs	14.9	6.0	-4.5	33.8	14.8	5.9	33.4	14.6
Multistage 1:0	15.2	5.9	-4.2	32.0	14.0	4.5	29.1	12.9
Multistage 1:2	9.9	-2.5	-14.3	20.1	6.2	-6.8	16.8	1.9
Multistage 0:1	9.3	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4
Av % last 10 yrs	7.2	-6.1	-18.1	15.2	1.6	-12.2	9.2	-5.4

Female

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Observed	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - Background	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Av % first 10 yrs	121.9	29.0	17.0	178.6	120.5	28.8	178.7	118.7
Multistage 1:0	129.2	30.4	16.6	158.8	107.6	26.6	146.9	93.7
Multistage 1:2	66.0	14.6	0.3	112.2	56.0	8.2	103.6	47.0
Multistage 0:1	61.1	13.1	-1.1	107.0	49.5	6.0	96.4	39.5
Av % last 10 yrs	50.7	9.0	-5.2	95.7	40.0	0.5	84.6	28.8

Note. --- indicates Observed - background was estimated to be negative for some age/year during the period.

TABLE 7S

Ratio (R%) change in predicted excess risk to change in observed excess risk - effects of variants to the model (Multistage 1:2, Swartz smoking submodel)

Male

Age	45-54			55-64			65-74	
Period	1956	1966	1976	1956	1966	1976	1966	1976
	1965	1975	1985	1965	1975	1985	1975	1985
BASIC	84.2	78.2	95.3	88.8	87.4	86.5	87.6	92.6
F18	84.1	78.1	95.0	88.0	87.3	86.4	86.8	92.4
F21	83.8	77.7	94.7	87.4	87.0	85.9	86.1	92.0
(K-1)3	84.7	79.2	96.9	89.6	88.4	88.1	88.5	94.3
(K-1)6	83.7	77.4	94.0	88.0	86.6	85.3	86.5	91.1
L0	82.9	76.3	93.0	87.6	85.8	85.3	86.0	90.6
D005	84.2	78.2	95.3	88.6	87.4	86.5	87.4	92.6

Female

Age	45-54			55-64			65-74	
Period	1956	1966	1976	1956	1966	1976	1966	1976
	1965	1975	1985	1965	1975	1985	1975	1985
BASIC	34.2	47.5	77.7	36.7	41.9	62.7	28.7	56.5
F18	34.0	47.3	77.4	36.6	41.6	62.4	28.6	56.1
F21	33.6	47.1	76.9	36.3	41.1	62.1	28.4	55.4
(K-1)3	35.3	48.4	79.6	37.5	43.1	64.1	29.3	58.1
(K-1)6	33.5	46.9	76.3	36.2	40.9	61.5	28.2	55.1
L0	33.2	46.1	75.4	35.9	40.7	61.5	28.0	55.6
D005	34.1	47.5	77.7	36.6	41.7	62.6	28.6	56.3

Note: F = first year of smoking
K-1 = power in multistage calculations
L = lag (years)
D = drift

TABLE 7T

Ratio (R%) change in predicted excess risk to change in observed excess risk - effects of variants to the model (Multistage 1:2, Townsend smoking submodel)

Male

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
BASIC	84.2	78.4	95.9	88.8	87.7	86.9	87.6	92.8
F18	84.2	78.3	95.3	88.1	87.4	86.5	86.7	92.5
F21	83.8	77.8	94.7	87.4	87.0	85.8	85.9	91.8
(K-1)3	84.8	79.5	97.6	89.6	88.6	88.4	88.5	94.4
(K-1)6	83.8	77.5	94.4	88.1	86.8	85.6	86.7	91.4
L0	83.1	76.6	93.7	87.7	86.0	85.7	86.0	0.0

Female

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
BASIC	34.3	47.6	78.0	36.6	42.0	63.1	28.6	56.9
F18	34.1	47.4	77.6	36.5	41.7	62.7	28.5	56.4
F21	33.7	47.2	77.0	36.3	41.2	62.3	28.3	55.7
(K-1)3	35.5	48.5	79.9	37.4	43.2	64.5	29.2	58.4
(K-1)6	33.6	46.9	76.5	36.2	41.0	61.9	28.2	55.5
L0	33.2	46.3	75.8	35.9	40.9	61.9	28.0	55.9

Note: F = first year of smoking
K-1 = power in multistage calculations
L = lag (years)

TABLE 8

Tar content of US cigarettes, sales-weighted average

Year	Tar (mgs/cig)
1957	35
1960	27
1965	23
1970	20
1975	18
1980	14
1985	13

Note: Selected years, taken from graph
Source: US Surgeon General (1989)

TABLE 9A

Number of cigarettes smoked per smoker per day; selected US surveys conducted 1947-80

Year	Survey[1]	Reprst[1] of US pop	Est from consumption categories[2]	Age range surveyed	Cigarettes per smoker Male	Female
1924	(a)Milwaukee	No[3]	-	-	10	-
1934	(a)Milwaukee	No[3]	-	-	13	7
1947	(10)Hamtoft and Lindhard	No[4]	2	20+	29[5]	21[5]
1955	(4) CPS	Yes	4	18+	18	13
1959	(9) ACS	No[6]	6	30+	21	15
1964	(3) AUT	Yes	8	20+	22	17
1965	(2) NHIS	Yes	3	20+	20	16
1966	(4) CPS	Yes	4	18+	19	16
1967	(4) CPS	Yes	4	17+	19	15
1968	(4) CPS	Yes	4	17+	19	16
1970	(3) AUT	Yes	No	20+	22	18
1975	(3) AUT	Yes	No	20+	23	19
1976	(2) NHIS	Yes	3	20+	21	18
1980	(2) NHIS	Yes	No	20+	23	20

- indicates not known

Notes

1 (a) From Harris (1980) quoting Milwaukee Journal (1924-1979)
Numbered sources taken from International Smoking Statistics (IntSS) Table 22.5, p463, full references and brief description for each survey pp470-474.

2 Number of categories in percentage distribution on which estimated mean cigarettes per smoker are based. See IntSS for full details of categories (Notes pp470-472) and method (Appendix III).

3 Greater Milwaukee area

4 Whites, in Columbus Ohio

5 Population weighted average of age-specific data

6 25 States, over-representative of white, married, better educated

Abbreviations:
CPS Current Population Surveys
ACS American Cancer Society Million Person Study
NHIS National Health Interview Surveys
AUT Adult Use of Tobacco Surveys

TABLE 9B

Estimated lifetime average tar corrected cigarette consumption per smoker per day

Method 1	Male			Female		
Age	50	60	70	50	60	70
Year						
1955	15.8	14.6	13.8	12.6	11.4	10.5
1965	17.4	16.0	14.9	13.5	12.6	11.8
1975	16.7	16.3	15.4	12.8	12.7	12.1
1985	13.8	15.0	15.1	10.7	11.7	11.8
Method 2						
1955	14.7	13.3	12.1	10.8	9.7	8.4
1965	16.4	15.1	13.9	12.1	11.4	10.5
1975	15.9	15.5	14.7	12.2	11.7	11.2
1985	13.7	14.5	14.4	10.7	11.3	11.0

Notes
Average taken from age started smoking by the relevant cohort (see first column of Table 10) up to age stated.
Consumption taken as
Males: 1924 10, 1934 13, 1955 20, 1980 23
Females: 1934 7, 1955 15, 1980 20.
Other years estimated as follows:
Method 1: Constant before 1924 (males) 1934 (females).
1945-1955 assumed constant, linear interpolation between 1934 and 1945, and between 1955 and 1980
Method 2: Males 1924-34 by linear interpolation. Same slope assumed for females, and for extrapolation before 1924.
Linear interpolation between subsequent date points.
Both methods: Tar corrected after 1957, see Table 8.

TABLE 10

Average age of starting to smoke
Comparison of survey based values and
values derived from smoking model

Cohort	Harris	Haenszel[2]	Cohort[3]	Swartz smoking model variants[1] Basic	F18	F21	D005
Males							
1881-90	21	19.3[4]	1885	22.9	23.5	24.8	27.0
1891-00	19	18.6	1895	21.9	22.6	23.9	26.0
1901-10	18	18.4	1905	19.1	20.3	22.1	23.6
1911-20	18	18.2	1915	18.6	19.8	21.7	22.5
1921-30	18	17.9	1925	18.0	19.2	21.3	20.9
1931-40	17		1935	17.9	19.1	21.3	19.6
1941-50	17						
1951-60	16						
Females							
1881-90	34	39.9[4]	1885	33.3	33.3	33.3	35.9
1891-00	32	35.3	1895	31.1	31.3	31.5	34.3
1901-10	28	26.0	1905	27.4	27.6	28.1	31.2
1911-20	23	21.3	1915	21.9	22.5	23.7	26.0
1921-30	21	20.0	1925	20.5	21.2	22.6	23.4
1931-40	20		1935	19.3	20.1	21.8	21.1
1941-50	18						
1951-60	17						

Note
1 F = first year of smoking, D = drift
2 Source: Haenszel(1956). Survey in 1955 as supplement to Current Population Survey.
3 Selected single year-of-birth cohorts
4 Born before 1890

TABLE 11

Estimated percentage of smokers seen in a surviving population, starting with an original percentage of 50%

Age at start		Length of follow-up (years) 10	20	30	40
25	S	98.0	93.8	82.5	61.1
	NS	98.7	96.4	90.9	77.7
	%	49.8	49.3	47.6	44.0
	A	35	45	55	65
35	S	95.7	84.2	62.3	30.9
	NS	97.7	92.1	78.7	53.0
	%	49.5	47.8	44.2	36.8
	A	45	55	65	75
45	S	88.0	65.1	32.3	7.7
	NS	94.3	80.6	54.3	19.9
	%	48.3	44.7	37.3	27.9
	A	55	65	75	85
55	S	74.1	36.7	8.7	
	NS	85.5	57.5	21.1	
	%	46.4	39.0	29.2	
	A	65	75	85	
65	S	49.6	11.8		
	NS	67.3	24.7		
	%	42.4	32.3		
	A	75	85		
75	S	23.8			
	NS	36.7			
	%	39.3			
	A	85			

Note
S = % smokers surviving, NS = % nonsmokers surviving, % = observed percentage of smokers (= S / (S + NS)), A = age at follow-up.

TABLE 12

Comparison of cigarettes sales with estimated sales based on Harris prevalence data

Year	Sales[1]	Harris				Harris as percentage of sales
		Age[2]		Total smokers[3] (thousands)	Total cigs[4] (millions)	
		min	max			
1951	391925	15	66	42031	306823	78.3
1952	405809	15	67	42760	312148	76.9
1953	397426	15	68	43569	318051	80.0
1954	378925	15	69	44486	324747	85.7
1955	391861	15	70	45673	333414	85.1
1956	401954	15	71	46542	339754	84.5
1957	418136	15	72	47415	346127	82.8
1958	445754	15	73	48298	352578	79.1
1959	462681	15	74	49085	358324	77.4
1960	479236	15	75	49921	364426	76.0
1961	497219	15	76	50712	370195	74.5
1962	503263	15	77	51546	376285	74.8
1963	518388	15	78	52372	382315	73.8
1964	507747	15	79	53074	387442	76.3
1965	520264	15	80	53728	392218	75.4
1966	531133	15	81	54252	396037	74.6
1967	536100	15	82	54991	401439	74.9
1968	532208	15	83	55707	406663	76.4
1969	520931	15	84	56388	411631	79.0
1970	545969	15	85	57056	416507	76.3
1971	540858	16	86	58114	424231	78.4
1972	559717	17	87	59257	432573	77.3
1973	600100	18	88	59206	432203	72.0
1974	607500	19	89	60068	438496	72.2
1975	613800	20	89	60635	442635	72.1
1976	620300	21	89	60925	444753	71.7
1977	620900	22	89	61557	449367	72.4
1978	620500	23	89	61859	451570	72.8
1979	626100	24	89	62057	453020	72.4
1980	635900	25	89	62422	455679	71.7

Notes.

1 Sales of manufactured cigarettes, plus estimated total numbers of hand-rolled cigarettes. From International Smoking Statistics, Tables 22.1.1/2

2 Age range available from Harris, see text for method of extension to full age range.

3 Using WHO population data

4 Assuming 20 cigarettes per smoker per day.

TABLE 13A

Estimates of prevalence of cigarette smoking from International Smoking Statistics

	Year of birth										
Age	1916	1921	1926	1931	1936	1941	1946	1951	1956	1961	1966
Male											
15-19	30.0	30.8	31.5	32.0	31.5	35.5	29.0	27.4	25.6	24.2	18.7
20-24	56.9	51.6	61.5	51.5	55.2	60.8	51.6	42.3	39.5	34.8	
25-29	55.4	66.0	62.3	59.2	61.5	56.6	48.8	44.7	38.2		
30-34	61.1	61.5	58.5	59.3	54.6	47.0	43.1	37.9			
35-39	62.5	60.1	59.3	53.9	48.2	45.1	36.7				
40-44	55.5	58.9	53.5	47.8	44.7	36.8					
45-49	55.9	51.2	44.0	42.5	36.6						
50-54	48.6	41.7	40.3	34.5							
55-59	39.2	39.5	33.9								
60-64	34.8	29.3									
65-69	25.4										
Female											
15-19	10.0	17.4	24.8	19.3	18.3	26.7	19.1	18.6	25.9	25.9	19.7
20-24	36.7	47.4	41.7	31.0	42.2	40.9	38.0	37.0	34.2	33.1	
25-29	49.6	43.6	37.7	44.1	42.6	42.1	38.2	36.6	33.2		
30-34	43.4	36.1	42.2	42.2	41.2	36.1	34.5	31.3			
35-39	34.9	41.2	41.8	40.0	37.5	37.6	30.8				
40-44	43.5	40.9	39.4	37.2	37.1	30.3					
45-49	36.9	37.3	37.2	37.9	30.6						
50-54	35.6	33.4	34.4	28.9							
55-59	30.9	32.6	29.2								
60-64	24.8	24.2									
65-69	19.5										

TABLE 13B

Difference between estimates of prevalence of cigarette smoking from Harris and from International Smoking Statistics

Age	Year of birth 1915	1925	1935	1945	1955
Male					
15-19	+2.5	+4	+2	+2	+3.5
20-24	+1	+2.5	+4	+3.5	+4
25-29	+12	+7.0	+2.5	+8	
30-34	+6.5	+10.5	+5.5	+7.5	
35-39	+5	+7.0	+5.5		
40-44	+9.5	+7.5	+3		
45-49	+4.5	+9			
50-54	+6.5	+6.5			
55-59	+7.5				
60-64	+4				
Female					
15-19	-1	-13	-2	0	-4
20-24	-13	-12	-6	-0.5	+4
25-29	-18	+1	+1.5	+3	
30-34	-7.5	0	+3	+4.5	
35-39	+2	+0.5	+4.5		
40-44	-5.5	+1.5	+2		
45-49	+0.5	+1.5			
50-54	+1	+1			
55-59	+2				
60-64	+4				

Note. Differences are Harris - IntSS. For Harris, year of birth is midpoint of 10 year cohort. IntSS data relate to cohort born one year later.

TABLE 14

Observed and predicted 10 year changes in risk from basic model, using alternative data from International Smoking Statistics

Sex	Male			Female		
Age	45-54		55-64	45-54		55-64
Period	1966 1975	1976 1985	1976 1985	1966 1975	1976 1985	1976 1985
Lung cancer rates						
Observed	22.5	-9.3	7.2	87.4	23.5	56.1
Obs -0.5*Background	23.6	-9.7	7.4	108.0	26.0	63.3
Obs - Background	24.7	-10.1	7.7	141.3	29.1	72.6
Absolute risk estimates						
Swartz 1 Brit Docs	-1.8	-10.4	-7.5	8.2	-4.3	9.5
Swartz 1 US Vets	-1.8	-9.5	-6.9	5.1	-3.6	6.0
Swartz 2 Brit Docs	0.3	-6.8	-2.9	2.3	-2.7	1.2
Swartz 2 US Vets	0.3	-6.1	-2.6	2.0	-2.3	1.1
Excess risk estimates						
Duration **k-1	-0.9	-11.8	-7.5	40.9	-7.2	34.5
Multistage 1:0	3.3	-2.0	3.0	21.1	-6.2	13.4
Multistage 5:1	-0.6	-9.1	-4.4	13.5	-5.5	12.5
Multistage 1:1	-1.9	-11.0	-7.2	3.9	-4.6	4.0
Multistage 1:2	-2.2	-11.5	-8.0	3.2	-4.5	3.3
Multistage 1:2E	-2.4	-12.0	-8.8	3.6	-4.6	4.1
Multistage 1:5	-2.4	-11.8	-8.5	2.7	-4.5	2.8
Multistage 0:1	-2.6	-12.0	-9.1	-0.5	-4.1	-1.4
Smoking indices						
Av % smkrs lifetime	-0.6	-7.6	-4.0	1.8	-2.5	0.8
Av % first 10 years	2.9	-0.2	3.1	19.6	-3.3	19.0
Av % last 10 years	-5.1	-15.8	-14.7	1.2	-7.3	-2.2
% 20 yrs ago	11.6	-1.6	-4.2	-15.1	12.6	12.7
% dur 30+ years	-0.2	-7.5	-3.5	131.3	-6.1	10.3

TABLE 15

Observed and predicted 10 year changes in risk in Italy, using basic model and data from La Vecchia

Male

Age	45-54			55-74		65-74
Period	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1976 1985
Lung cancer rate						
Observed	12.7	54.1	-2.9	20.9	30.1	25.2
Obs - 0.5*Background	13.6	57.7	-3.1	21.8	31.2	26.4
Obs - Background	14.7	61.7	-3.2	22.9	32.4	27.6
Absolute risk estimates						
Swartz 1 Brit Docs	8.2	2.3	-4.2	6.8	1.1	7.1
Swartz 1 US Vets	6.9	2.2	-4.1	5.6	1.1	5.9
Swartz 2 Brit Docs	6.8	3.4	-4.7	6.2	2.8	5.8
Swartz 2 US Vets	6.1	3.0	-4.2	5.6	2.6	5.4
Excess risk estimates						
Duration **k-1	15.1	1.2	-2.2	11.5	0.3	11.2
Multistage 1:0	13.4	2.8	-3.0	11.3	3.4	10.0
Multistage 5:1	10.5	2.6	-4.3	8.9	1.9	8.8
Multistage 1:1	8.1	3.1	-5.4	6.5	1.9	6.7
Multistage 1:2	7.9	3.1	-5.5	6.2	1.8	6.3
Multistage 1:2E	7.9	3.0	-5.5	6.1	1.6	6.3
Multistage 1:5	7.7	3.1	-5.6	5.9	1.7	6.1
Multistage 0:1	7.0	3.3	-6.0	4.9	1.9	4.9
Smoking indices						
Av % smkrs lifetime	7.6	2.8	-5.1	6.6	2.7	6.4
Av % first 10 yrs	11.4	2.1	-3.2	11.7	2.3	11.4
Av % last 10 yrs	7.1	1.9	-5.8	3.3	1.0	5.6
% 20 yrs ago	5.3	4.3	-5.2	6.6	4.6	7.1
% dur 30+ years	48.5	-3.5	2.5	7.2	2.1	6.6

TABLE 15 (cont)

Observed and predicted 10 year changes in risk in Italy, using basic model and data from La Vecchia

Female

Age	45-54			55-74		65-74
Period	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1976 1985
Lung cancer rate						
Observed	15.1	8.2	-7.2	12.4	29.4	30.2
Obs - 0.5*Background	27.5	13.2	-10.5	21.0	45.2	50.4
Obs - Background	---	---	-19.9	68.0	97.0	151.1
Absolute risk estimate						
Swartz 1 Brit Docs	21.6	20.7	22.4	24.5	24.8	29.9
Swartz 1 US Vets	14.1	14.7	16.6	16.0	17.4	19.3
Swartz 2 Brit Docs	21.1	20.2	21.8	26.8	25.0	32.2
Swartz 2 US Vets	12.8	13.3	15.3	17.1	17.5	21.5
Excess risk estimates						
Duration **k-1	82.3	44.9	46.2	78.1	56.0	97.7
Multistage 1:0	85.1	43.9	44.4	81.6	45.4	77.7
Multistage 5:1	73.0	46.0	41.6	72.4	48.9	79.9
Multistage 1:1	68.3	46.7	39.9	67.6	48.1	75.2
Multistage 1:2	67.8	46.8	39.8	66.9	48.3	75.0
Multistage 1:2E	67.6	46.8	39.8	66.6	48.8	75.9
Multistage 1:5	67.5	46.9	39.7	66.5	48.4	74.9
Multistage 0:1	66.4	47.0	39.2	65.2	47.7	72.5
Smoking indices						
Av % smkrs lifetime	68.8	45.5	41.8	67.1	46.2	70.4
Av % first 10 yrs	78.5	43.1	46.8	80.4	43.4	79.8
Av % last 10 yrs	63.0	45.4	38.1	62.4	46.8	81.2
% 20 yrs ago	73.3	46.6	47.1	68.7	46.8	58.9
% dur 30+ years	151.4	21.8	58.4	75.3	55.1	79.1

TABLE 16

Observed and predicted 10 year changes in risk in Norway, using basic model and data from Rønneberg

Male

Age	45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate								
Observed	76.9	37.6	-15.9	23.7	65.1	24.0	55.0	35.7
Obs - 0.5*Background	95.5	43.3	-17.4	27.6	74.1	25.8	62.4	38.8
Obs - Background	---	51.2	-19.2	33.1	86.0	27.9	72.1	42.5
Absolute risk estimate								
Swartz 1 Brit Docs	4.5	-4.4	-17.5	9.1	-1.4	-8.1	2.0	-5.7
Swartz 1 US Vets	4.1	-4.3	-15.9	8.5	-1.3	-7.8	2.1	-5.2
Swartz 2 Brit Docs	5.8	-0.1	-10.5	11.3	3.1	-2.9	8.4	0.9
Swartz 2 US Vets	5.3	-0.1	-9.6	10.4	2.8	-2.7	7.9	0.8
Excess risk estimates								
Duration **k-1	5.3	-2.3	-21.3	8.8	-1.6	-7.2	0.9	-6.4
Multistage 1:0	7.5	8.5	-1.5	12.0	7.5	6.9	11.9	7.1
Multistage 5:1	5.7	-1.0	-13.8	10.7	1.5	-3.2	5.8	-0.9
Multistage 1:1	5.1	-4.4	-16.7	10.9	-0.2	-7.3	4.7	-3.6
Multistage 1:2	5.0	-5.0	-17.7	10.8	-0.8	-8.5	4.0	-4.8
Multistage 1:2E	4.9	-5.6	-18.9	10.6	-1.5	-9.7	3.0	-6.2
Multistage 1:5	4.9	-5.5	-18.3	10.8	-1.2	-9.3	3.6	-5.6
Multistage 0:1	4.8	-6.2	-18.4	11.1	-1.5	-10.4	3.8	-6.1
Smoking indices								
Av % smkrs lifetime	6.2	-0.9	-13.2	11.2	3.0	-4.0	8.2	0.4
Av % first 10 yrs	7.6	10.5	-3.4	10.6	7.6	10.5	10.4	7.6
Av % last 10 yrs	2.9	-10.3	-21.8	7.7	-6.1	-14.1	-3.3	-12.1
% 20 yrs ago	9.1	6.1	-11.6	14.1	6.9	-5.1	12.9	-1.4
% dur 30+ years	3.3	-1.5	-24.5	10.0	-0.1	-6.6	11.5	-1.8

TABLE 16 (cont)

Observed and predicted 10 year changes in risk in Norway, using basic model and data from RVnneberg

Female

Age	45-54			55-64			65-74	
Period	1956	1966	1976	1956	1966	1976	1966	1976
	1965	1975	1985	1965	1975	1985	1975	1985
Lung cancer rate								
Observed	18.5	132.9	71.5	11.2	79.1	122.0	38.8	63.9
Obs - 0.5*Background	---	523.2	109.8	---	---	192.5	182.9	119.3
Obs - Background	---	---	---	---	---	456.2	---	---
Absolute risk estimate								
Swartz 1 Brit Docs	41.0	28.3	10.5	48.5	30.4	28.4	40.5	29.4
Swartz 1 US Vets	31.4	21.4	8.0	35.2	22.8	21.1	28.8	21.5
Swartz 2 Brit Docs	45.5	31.0	13.1	57.3	37.9	28.2	53.8	35.6
Swartz 2 US Vets	33.3	25.0	11.1	39.8	29.9	23.7	39.7	29.2
Excess risk estimates								
Duration **k-1	102.6	91.1	28.1	114.3	83.9	75.7	109.3	75.0
Multistage 1:0	114.0	94.4	43.1	126.4	102.4	78.2	129.2	91.0
Multistage 5:1	72.9	49.6	21.1	100.4	57.6	49.4	87.5	57.6
Multistage 1:1	62.0	31.9	11.7	92.3	40.2	30.4	70.1	38.5
Multistage 1:2	60.9	30.1	10.3	91.4	38.2	28.0	67.9	35.7
Multistage 1:2E	60.6	29.8	9.6	91.0	37.3	27.4	66.8	34.3
Multistage 1:5	60.2	29.0	9.5	90.8	36.9	26.4	66.5	33.8
Multistage 0:1	58.1	25.0	7.2	89.2	32.9	20.7	62.4	28.6
Smoking indices								
Av % smkrs lifetime	69.7	39.7	13.8	103.1	51.4	33.0	84.6	44.9
Av % first 10 yrs	105.9	111.7	37.2	121.4	105.4	111.5	123.0	104.6
Av % last 10 yrs	45.4	16.5	3.3	66.6	19.6	16.4	41.8	23.1
% 20 yrs ago	102.2	88.5	3.7	154.6	78.6	13.5	111.6	17.2
% dur 30+ years	194.1	102.5	114.5	137.4	75.6	95.6	130.0	67.2

P.N. LEE STATISTICS AND COMPUTING LTD.

Hamilton House
17 Cedar Road
Sutton
Surrey SM2 5DA
Telephone: 081-642 8265 (4 lines)
Fax: 081-642 2135
VAT Reg. No. 318 4017 78

PNL/pw

16 December 1993

Dr J B Swartz
Department of Health Services
Environmental Epidemiology and Toxicology Branch
5900 Hollis Street
Suite E
Emeryville
CA 94608
USA

Dear Dr Swartz,

I have been looking recently at your 1992 paper in the Journal of Epidemiology and Community Health on "Use of a multistage model to predict time trends in smoking induced lung cancer". I would like to try one or two other mathematical models using the Harris data you cite and also to try your model with other data. Unfortunately Harris's paper only gives graphical results and your paper only gives selected data. I would be extremely grateful if you could supply me with a listing or floppy disk containing all the smoking and lung cancer data you used to test your model. This would help me considerably in ensuring I could reproduce your findings and be able to identify clearly any differences as being model and not data dependent. If you could let me have copies also of any software you used to fit the model I would be grateful too. I would of course be happy to pay any reasonable charge for any expenses involved.

Thank you in advance.

Yours sincerely,

pp D Wassell (sec'y)

Peter N Lee

125 Moss Avenue, #120
Oakland, California, 94611
United States of America
January 3, 1994

Dr. P. N. Lee
P.N. Lee Statistics and Computing
Hamilton House, 17 Cedar Road, Sutton
Surrey SM2 5DA, United Kingdom

Dear Dr. Lee:

Thank you for your recent letter concerning the article " Use of a Multistage Model...." Here are the answers to your questions as best as I can answer, not necessarily in order.

1. For fitting the empirical multistage model(equation 1) I used the parameters from Whittemore's article(Stat. Med, 1988; 7, 223-238). So I did no fitting of my own for this model.

For the alternate model(equations 3 and 4) I did a fit to the data presented in table 1. I don't know if I still have this software available, although I expect to find the parameters when I reaccess my software. In any case I should have the values of the parameters. I hope to be able to find these within a month. Please see below.

2. I obtained my prevalence estimates from the Harris paper using a ruler and a straight edge to extract the prevalences from the graphs. I used the 10 year endpoints, and interpolated for the prevalences in between.

The numbers which I actually used are included in my programs. Because I have moved several times the programs are not easily accessible at this time. However, I am in the process of of getting them out for future use. I expect to have a printout and a tape or disk of the program within a month. So at that time I will be able to send you the numbers which I extracted from the Harris tables and/or the actual software. The main function of the software is to produce smoking spectra, i.e. the number of people who started and stopped smoking by age , in given years. I suggest that in the meantime you use whatever numbers you can extract from the Harris graphs with a straight edge. Even in the era of high tech this method does not work too badly. I suspect that prevalences each of us obtains from the graphs will not be very different, but I completely agree with you that it would be better if we used the precise same values for the prevalences.

I am very excited that you are interested in trying out different models, and in using other data sets. I am also planning to use the model on additonal data sets in the near future. I would appreciate your keeping me informed of your progress, and I will do the same. Also if you can think of any additional uses for this type of modelling I would be very interested. Please use the above address for the time being. I will send you my new address shortly.

Yours truly,

Joel B. Swartz

Joel B. Swartz, Ph. D.

P.N. LEE STATISTICS AND COMPUTING LTD.

Hamilton House
17 Cedar Road
Sutton
Surrey SM2 5DA
Telephone: 081-642 8265 (4 lines)
Fax: 081-642 2135
VAT Reg. No. 318 4017 78

PNL/pw

7 April 1994

Dr Joel B Swartz
125 Moss Avenue, #120
Oakland
California 94611
USA

Dear Dr Swartz,

You will remember that we corresponded a few months ago about your 1992 paper in the Journal of Epidemiology and Community Health. In your letter of 3 January you said that you expected to have a printout and a tape or disk of the programs you used within a month (including the actual numbers you extracted from the Harris tables), but I have heard nothing since. I am writing to remind you of this and also to raise some further points that have come up when trying to reproduce your results.

1. In your formula 1, you state that c is packs per day, but in fact Whittemore's source paper uses c as cigarettes per day. I believe you actually used c as cigarettes per day since we get much closer agreement to your results with c as cigarettes than with c as packs, but can you confirm this please.

2. Why did you use $p = 0.207$, Whittemore's fitted value for British Doctors' data, rather than $p = 0.128$, the value which fitted both sets of US data better? After all you were concerned with US data.

3. I assume you only applied formula 1 for $t_1<t-5$ as Whittemore's paper indicates. Or equivalently evaluated the formula with $t^{4.5}$ rather than $(t-5)^{4.5}$ to give a value which was then taken to be the risk of someone five years later (i.e. ignoring smoking history up to five years before death).

4. The "actual" lung cancer data you used were for white males. Why so? Harris's data are regardless of race and British Doctors are not all white either (though the ethnic distribution is very different from US blacks). Can you let me have the actual US lung cancer data you used? Your reference 30 (see Table III) does not appear in the reference list.

5. You state that mortality rates are for the age range 42 to 70. Why? How did you get US data for these ages? Normally rates are given only for five-year periods starting with e.g. 40, 45, 50 ...

6. On p313 you state that the rate for 1970 was computed using the rate for 69 year olds born in 1901, 68 years olds born in 1902 and so on. Why were the 70 year olds born in 1900 not included?

7. Whittemore's formula may be wrong! See copy of a letter of mine to her.

I await your answer with great interest. Could you give me your phone number when you reply so that I can pursue any other points easily.

Best wishes.

Yours sincerely,

Peter N Lee

enc

Charles Drew University of Medicine
Epidemiology and Statistics Unit
Mail Stop 30
Los Angeles, California, 90066
United States of America
April 22,1994

Dr. P. N. Lee
Hamilton House
17 Cedar Road
Sutton
Surrey SM2 5DA
England
United Kingdom

Dear Dr. Lee:

I just received your recent letter. I have been involved with moving, and have not had a chance to write to you. I will be able to write you to answer your questions within two weeks. Most of your question can be answered easily. If you do not hear from me within two weeks pleae contact me. My work phone is213-563--4842.

Yours truly,

Joel B. Swartz

Joel B. Swartz, Ph.D.

Reply to Correspondence Dated April 7,1994

Dr. Lee,

These are nearly complete answers to your questions. I will send final answers in a week. Please let me know of your results.I am going to perform some additional calculations of my own. Would you be interested in possibly arranging for joint publication of results and comments?

1. You are correct. "c" is the smoking rate in cigarettes per day.

2. I used parameters from the British physician's study for the following reason
A: This study provided the best fit of the three listed in the Whittemore paper.
B. I have done some work with the Dorn study. It does not have complete smoking histories, so it has the largest possibility of an error.
C. I doubt that there would be any fundamental difference in lung cancer function between the U.S. and Great Britain, although there are a number of factors which affect this function for which we are unable to control.

4. I used data for white males because I did not have time to apply it to other groups in the population. I do not think that there is any underlying difference between blacks and whites in lung cancer susceptibility. The ethnic mix in the U.S. is obviously somewhat different from that in Great Britain. I think the number of black physicians in England at that time was very small.The ability to perform these calculations, and also the validity of most epidemiologic studies depends on the relative similarity in exposure effects across populations.

I think it is simplistic to try to identify two populations as equivalent just because they each had some portion of blacks and whites.I also think it is simplistic to assume that the largest differences in populations are due to racial factors as opposed to other factors such as type of work, income ,etc.

The lung cancer data come from two sources:
A. National Cancer Institute (U.S.), Division of Cancer Prevention and Control, Statistical Review, 1987, U.S. Dept of Health and Human Services

B. an article by Pollack and Horm in JNCI 64:1091,1980.

As I recall the base period for age adjustment was different for the two data source, so I made some adjustment to insure that the trends were correct.

5. Please remember that the predicted mortality rates come from the model. There is no problem in computing these by 1 year age intervals. The smoking prevalences by age and cohort were computed by linear interpolation from the nearest age and cohort categories. Naturally the population lung cancer mortality rates are computed by age adjusting over 5 year periods.

7. I took Whittemore's equation to be a semi-empirical equation, based on the multistage model, but not identical to the appropriate model equation. Under the strict multistage model the exponents would have to be integral, but here the exponent is 4.5.

Yours truly,

Joel B. Swartz

Joel B. Swartz, Ph.D.

May 18

Dear Dr. Lee:

Just as an afterthought to my previous letter, I am in process of getting my old programs out of the archives. I expect to have them running within a few weeks. Perhaps it would be interesting to plan parallel computations.

Yours truly,

Joel B. Swartz, Ph.D.

Appendix B
Correspondence with Whittemore

P.N. LEE STATISTICS AND COMPUTING LTD.

Hamilton House
17 Cedar Road
Sutton
Surrey SM2 5DA
Telephone: 081-642 8265 (4 lines)
Fax: 081-642 2135
VAT Reg. No. 318 4017 78

PNL/pw

7 April 1994

Prof A Whittemore
Department of Family, Community
and Preventive Medicine
Stanford University School of Medicine
Stanford
California 94305
USA

Dear Alice,

I hope you still remember me from what must be almost 20 years ago! I have recently been asked to carry out a detailed critique of a paper by Joel Swartz in the Journal of Epidemiology and Community Health (1992, 46, 311-315) in which he estimated lung cancer rates in the US population based on extrapolated smoking prevalence data and a function linking mortality to smoking which you derived in your 1988 Statistics of Medicine paper (7, 223-238). Formula 12 in your paper (Formula 1 in Swartz's paper) gave the mortality rate as

$$2.01\times10^{-12}\{(t-5)^{4.5} + pc(1+2pc)(t_1-t_0)^{4.5} + 2pc(t_1^{4.5}-t_0^{4.5})\}$$

where t is age, t_0 is time of starting to smoke, t_1 is time of stopping smoking, p is a constant (Swartz uses your British Doctors fitted value of 0.207), and c is cigarettes per day (Swartz erroneously states c is packs per day). This is derived assuming a multistage model with the first and penultimate stages affected, the penultimate twice as much as the first. The death rate at age t corresponds to smoking experience up to five years before death.

/Trying unsuccessfully

Trying unsuccessfully to reproduce Swartz's findings and checking everything, I tried to derive the formula you gave and found that I could not. Ignoring the lag time of five years my calculations gave a function of the form

$$t^{k-1} + pc[(t-t_0)^{k-1}-(t-t_1)^{k-1}] + 2pc[t_1^{k-1}-t_0^{k-1}] + 2(pc)^2[t_1-t_0]^{k-1}$$

Compared with your formula mine differs by having a term in $pc[(t-t_0)^{k-1}-(t-t_1)^{k-1}]$ rather than your term in $pc(t_1-t_0)^{k-1}$. For continuous exposure ($t=t_1$) the two formulae are identical, but for discontinuous exposure they are not.

Having come up with this discrepancy, I then looked at the paper by Brown and Chu in J Chron Diseases (1987, 40, 171S-179S), which gives the formula as

$$t^{k-1} + r_1[(d+f)^{k-1}-f^{k-1}] + r_{k-1}[(t-f)^{k-1}-(t-d-f)^{k-1}] + r_1 r_{k-1} d^{k-1}$$

where r_1 and r_{k-1} are the stage effects, d is duration, and f is time elapsed since exposure. Substituting $r_1=pc$, $r_{k-1}=2pc$, $d=t_1-t_0$, $f=t-t_1$, one gets exactly my formula.

My questions to you are:

(1) Do you agree my formula actually is correct?

(2) Did you actually use the formula you cited when carrying out your fits to the New Mexico data or is it just that the formula was wrongly printed in the paper?

/(3) If you

(3) If you did use the formula you cited and it was the wrong one, would using the right one have fitted the New Mexico data better?

I look forward to your reply.

Best wishes.

Yours sincerely,

Peter N Lee

P.S. I also noted that in Table 1 of your paper, the pack years stated to be in hundreds are actually in thousands (referring back to the original source). I think you used the correct data in your analysis but just gave the footnote wrong.

STANFORD UNIVERSITY SCHOOL OF MEDICINE
DEPARTMENT OF HEALTH RESEARCH AND POLICY
DIVISION OF EPIDEMIOLOGY

April 18, 1994

HEALTH RESEARCH AND POLICY BUILDING
STANFORD, CALIFORNIA 94305-5092

(415) 723-5460
FAX (415) 725-6951

Peter N. Lee
P. N. Lee Statistics and Computing Ltd.
Hamilton House
17 Cedar Road
Sutton
Surrey SM2 5DA

Dear Peter:

It's a treat to hear from you, after (eecks!) almost 20 years. I hope that the intervening decades have been good ones for you and your family.

Alas, it appears that formula (12) in my paper is incorrect, as you note. I agree with your formula. I have dusted off my old records, and it appears that I used the correct formula in fitting the New Mexico data. Although I have records of my fortran programs using (12) for the British smokers and the US Veterans (for which (12) is okay because they were assumed to smoke continuously), a colleague, Jerry Halpern, did the GLIM programming for the NM data. My records contain a note to him on November 3, 1986 giving him the integral formula (11). (Incidentally, formula (11) is missing an exponent of 2.5 on the term s-u in the third integral.) Jerry used (11) to program the g_2 function for each case and control, based on his smoking history (some may have started and stopped more than once).

I feel badly that this error has misled Joel Swartz (and possibly others). Do you recommend that I publish an erratum at this late date? Should I contact Swartz? If so, do you wish your identity kept secret?

Do you ever get to the west coast of the US? If so, it would be fun to get together to swap stories. We never did finish that work on overdispersed tumor counts for the shaved backs of mice!

Thanks for the good calculations.

Sincerely,

Alice S. Whittemore, Ph.D., M.A.
Professor of Epidemiology and
Biostatistics
Director for Epidemiology,
Northern California Cancer Center

ASW:eem

Appendix C

Detailed examples of smoking models

In this Appendix, detailed tables show how prevalence data by single ages are used under the three smoking models to simulate the progress of an individual cohort through their smoking lives. Males born in 1900 are used for these examples. Ages 15-40 and 70 are shown, except for the Swartz model with drift, where ages 15-20, 30, 40 and 70 are shown.

The prevalences were obtained by linear interpolation as described in section 2.4 from the Harris data for 1895 (1891-1900) and 1905 (1910-1910) and are shown in Table C1.

The Swartz smoking model is shown in Table C2. For each age, the percentages of smokers are shown in a triangular matrix, where each row contains persons who started smoking at an given age. The first column contains current smokers, and subsequent columns contain ex-smokers divided according to their duration of smoking.

The youngest age considered is 15, so the 14.5% of the population who smoke are all assigned to age started 15. At age 16, the prevalence had increased to 18.5%, so the previous smokers carry forward with no ex-smokers and a further 4.0% are assigned to age started 16.

Prevalence increases until age 31, when it decreases by 0.25%. So there are no "starters" (bottom of current smokers columns). The ex-smokers are subtracted proportionally from all the available current smokers and assigned to the final column in each row, which represents giving up at the current age.

Table C3 presents the Swartz smoking model with drift. At each

year, "starters" or "stoppers" are added in order to match the required prevalence, as for the basic Swartz model. Then the drift is applied where 0.5% of current smokers are re-assigned as "stoppers" (added to the final column of the same row), and a corresponding number of never smokers are re-assigned as "starters" (at the bottom of the current smokers column).

Table C4 shows the Townsend model. Here, the population is divided into a number of groups, the first of which is never smokers. For each group, the first column shows the percentage of the population in the group. The next two columns show the duration of smoking, firstly up to the current age and secondly (of relevance to lagged mortality models) the duration up to 5 years ago; these are shown as negative for ex-smokers. The final columns show the number of changes in smoking status made by the group and the ages of the changes, which are alternately starting and stopping.

At age 15, 85.5% in group 1 were never smokers, while the remaining 14.5% in group 2 were current smokers, started at age 15. At age 16, as the prevalence increased, a further 4.0% were transferred from group 1 to a new group 3. As the prevalence increased steadily, the model continues with a new group being added each year, up to age 31. Then a prevalence drop requires 0.25% stoppers and these are selected from the group with lowest desire to smoke, the shortest duration of smoking, namely group 17. They are set up as a new group 18, and their duration marked negative to indicate that they are ex-smokers. The following year there is another prevalence increase and the ex-smokers in group 18 are

selected as re-starters. Since their number is exactly as required, no new group is created.

At higher ages the prevalence decreases fairly steadily and at each stage the group with the longest duration of smoking is either converted completely to ex-smokers, or split to create a new group of ex-smokers. Comparing the output at age 40 and age 70, it can be seen, for instance, that of group 16 who started smoking at age 29, 0.25% gave up at age 43 (group 20), and 0.25% gave up at age 44 (group 21). The rest gave up at age 45 (still group 16), along with all who started at age 28 (group 15) and part of those who started at age 27 (groups 14/22).

By age 70, the only remaining smokers are those who started at ages 15-19.

Table C1

Prevalence of smoking by cohort, data from Harris for 1895 and 1905 cohorts, and by interpolation for 1900 cohort

Age	Cohort 1895	1900	1905
15	10.0	14.50	19.0
16	13.0	18.50	24.0
17	16.5	22.75	29.0
18	20.0	27.00	34.0
19	23.5	30.00	36.5
20	27.5	33.75	40.0
21	32.5	38.75	45.0
22	37.0	43.00	49.0
23	41.0	47.00	53.0
24	44.0	49.50	55.0
25	45.0	51.00	57.0
26	46.0	52.00	58.0
27	47.0	53.00	59.0
28	47.5	53.50	59.5
29	48.5	54.25	60.0
30	49.0	54.75	60.5
31	48.5	54.50	60.5
32	48.5	54.75	61.0
33	49.0	55.25	61.5
34	49.5	55.25	61.0
35	49.0	55.25	61.5
36	48.5	54.75	61.0
37	48.5	54.75	61.0
38	48.5	54.75	61.0
39	48.5	54.50	60.5
40	48.5	54.25	60.0
41	48.5	54.25	60.0
42	49.5	54.50	59.5
43	49.0	54.00	59.0
44	49.0	53.75	58.5
45	47.5	52.50	57.5
46	47.0	52.00	57.0
47	47.0	52.00	57.0
48	46.5	51.50	56.5
49	46.0	51.50	56.0
50	45.0	49.75	54.5
51	45.0	49.25	53.5
52	45.0	49.25	53.5
53	45.0	49.00	53.0
54	44.0	48.00	52.0
55	43.0	46.00	49.0
56	45.0	46.50	48.0
57	41.0	44.25	47.5
58	41.0	44.00	47.0
59	40.0	43.00	46.0
60	37.5	40.75	44.0
61	37.5	40.25	43.0
62	37.5	39.75	42.0
63	37.0	39.00	41.0
64	36.0	37.50	39.0
65	34.0	35.50	37.0
66	32.0	33.50	35.0
67	31.5	32.25	33.0
68	31.0	31.50	32.0
69	29.5	29.75	30.0
70	28.0	28.50	29.0

Table C2

Example working of Swartz smoking model, Male 1900 cohort
(See explanation p C1)

Age 15 Prevalence 14.50

Age Started	Current Smokers
15	14.500

Age 16 Prevalence 18.50

Age Started	Current Smokers	Ex Smokers, Duration 1
15	14.500	0.000
16	4.000	

Age 17 Prevalence 22.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2
15	14.500	0.000	0.000
16	4.000	0.000	
17	4.250		

Age 18 Prevalence 27.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3
15	14.500	0.000	0.000	0.000
16	4.000	0.000	0.000	
17	4.250	0.000		
18	4.250			

Age 19 Prevalence 30.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4
15	14.500	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	
17	4.250	0.000	0.000		
18	4.250	0.000			
19	3.000				

Age 20 Prevalence 33.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5
15	14.500	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000		
18	4.250	0.000	0.000			
19	3.000	0.000				
20	3.750					

Age 21 Prevalence 38.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000			
19	3.000	0.000	0.000				
20	3.750	0.000					
21	5.000						

Age 22 Prevalence 43.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000				
20	3.750	0.000	0.000					
21	5.000	0.000						
22	4.250							

Table C2 (cont)

Age 23 Prevalence 47.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000					
21	5.000	0.000	0.000						
22	4.250	0.000							
23	4.000								

Age 24 Prevalence 49.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000						
22	4.250	0.000	0.000							
23	4.000	0.000								
24	2.500									

Age 25 Prevalence 51.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000							
23	4.000	0.000	0.000								
24	2.500	0.000									
25	1.500										

Age 26 Prevalence 52.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000	0.000							
23	4.000	0.000	0.000	0.000								
24	2.500	0.000	0.000									
25	1.500	0.000										
26	1.000											

Age 27 Prevalence 53.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000	0.000	0.000							
23	4.000	0.000	0.000	0.000	0.000								
24	2.500	0.000	0.000	0.000									
25	1.500	0.000	0.000										
26	1.000	0.000											
27	1.000												

Table C2 (cont)

Age 28 Prevalence 53.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000	0.000	0.000	0.000							
23	4.000	0.000	0.000	0.000	0.000	0.000								
24	2.500	0.000	0.000	0.000	0.000									
25	1.500	0.000	0.000	0.000										
26	1.000	0.000	0.000											
27	1.000	0.000												
28	0.500													

Age 29 Prevalence 54.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
23	4.000	0.000	0.000	0.000	0.000	0.000	0.000								
24	2.500	0.000	0.000	0.000	0.000	0.000									
25	1.500	0.000	0.000	0.000	0.000										
26	1.000	0.000	0.000	0.000											
27	1.000	0.000	0.000												
28	0.500	0.000													
29	0.750														

Table C2 (cont)

Age 30 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	14.500	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
17	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
18	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
19	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
20	3.750	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
21	5.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
22	4.250	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
23	4.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000								
24	2.500	0.000	0.000	0.000	0.000	0.000	0.000									
25	1.500	0.000	0.000	0.000	0.000	0.000										
26	1.000	0.000	0.000	0.000	0.000											
27	1.000	0.000	0.000	0.000												
28	0.500	0.000	0.000													
29	0.750	0.000														
30	0.500															

Age 31 Prevalence 54.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
15	14.434	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066
16	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	
17	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019		
18	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019			
19	2.986	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014				
20	3.733	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017					
21	4.977	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023						
22	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019							
23	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018								
24	2.489	0.000	0.000	0.000	0.000	0.000	0.000	0.011									
25	1.493	0.000	0.000	0.000	0.000	0.000	0.007										
26	0.995	0.000	0.000	0.000	0.000	0.005											
27	0.995	0.000	0.000	0.000	0.005												
28	0.498	0.000	0.000	0.002													
29	0.747	0.000	0.003														
30	0.498	0.002															
31	0.000																

Table C2 (cont)

Age 32 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
15	14.434	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000
16	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	
17	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000		
18	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000			
19	2.986	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000				
20	3.733	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000					
21	4.977	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000						
22	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000							
23	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000								
24	2.489	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000									
25	1.493	0.000	0.000	0.000	0.000	0.000	0.007	0.000										
26	0.995	0.000	0.000	0.000	0.000	0.005	0.000											
27	0.995	0.000	0.000	0.000	0.005	0.000												
28	0.498	0.000	0.000	0.002	0.000													
29	0.747	0.000	0.003	0.000														
30	0.498	0.002	0.000															
31	0.000	0.000																
32	0.250																	

Age 33 Prevalence 55.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
15	14.434	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000
16	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	
17	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000		
18	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000			
19	2.986	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000				
20	3.733	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000					
21	4.977	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000						
22	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000							
23	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000								
24	2.489	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000									
25	1.493	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000										
26	0.995	0.000	0.000	0.000	0.000	0.005	0.000	0.000											
27	0.995	0.000	0.000	0.000	0.005	0.000	0.000												
28	0.498	0.000	0.000	0.002	0.000	0.000													
29	0.747	0.000	0.003	0.000	0.000														
30	0.498	0.002	0.000	0.000															
31	0.000	0.000	0.000																
32	0.250	0.000																	
33	0.500																		

Table C2 (cont)

Age 34 Prevalence 55.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
15	14.434	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000
16	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	
17	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000		
18	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000			
19	2.986	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000				
20	3.733	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000					
21	4.977	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000						
22	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000							
23	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000								
24	2.489	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000									
25	1.493	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000										
26	0.995	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000											
27	0.995	0.000	0.000	0.000	0.005	0.000	0.000	0.000												
28	0.498	0.000	0.000	0.002	0.000	0.000	0.000													
29	0.747	0.000	0.003	0.000	0.000	0.000														
30	0.498	0.002	0.000	0.000	0.000															
31	0.000	0.000	0.000	0.000																
32	0.250	0.000	0.000																	
33	0.500	0.000																		
34	0.000																			

Age 35 Prevalence 55.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.434	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	
17	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000		
18	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000			
19	2.986	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000				
20	3.733	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000					
21	4.977	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000						
22	4.231	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000							
23	3.982	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000								
24	2.489	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000									
25	1.493	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000										
26	0.995	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000											
27	0.995	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000												
28	0.498	0.000	0.000	0.002	0.000	0.000	0.000	0.000													
29	0.747	0.000	0.003	0.000	0.000	0.000	0.000														
30	0.498	0.002	0.000	0.000	0.000	0.000															
31	0.000	0.000	0.000	0.000	0.000																
32	0.250	0.000	0.000	0.000																	
33	0.500	0.000	0.000																		
34	0.000	0.000																			
35	0.000																				

Table C2 (cont)

Age 36 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.303	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	
18	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038		
19	2.959	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027			
20	3.699	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034				
21	4.932	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045					
22	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038						
23	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036							
24	2.466	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023								
25	1.480	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014									
26	0.986	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009										
27	0.986	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009											
28	0.493	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005												
29	0.740	0.000	0.003	0.000	0.000	0.000	0.000	0.007													
30	0.493	0.002	0.000	0.000	0.000	0.000	0.005														
31	0.000	0.000	0.000	0.000	0.000	0.000															
32	0.248	0.000	0.000	0.000	0.002																
33	0.495	0.000	0.000	0.005																	
34	0.000	0.000	0.000																		
35	0.000	0.000																			
36	0.000																				

	21
15	0.131

Table C2 (cont)

Age 37 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.303	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000
18	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	
19	2.959	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027	0.000		
20	3.699	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034	0.000			
21	4.932	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045	0.000				
22	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000					
23	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036	0.000						
24	2.466	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023	0.000							
25	1.480	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.000								
26	0.986	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000									
27	0.986	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000										
28	0.493	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000											
29	0.740	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.000												
30	0.493	0.002	0.000	0.000	0.000	0.000	0.005	0.000													
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000														
32	0.248	0.000	0.000	0.000	0.002	0.000															
33	0.495	0.000	0.000	0.005	0.000																
34	0.000	0.000	0.000	0.000																	
35	0.000	0.000	0.000																		
36	0.000	0.000																			
37	0.000																				

	21	22
15	0.131	0.000
16	0.000	

Table C2 (cont)

Age 38 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.303	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000
18	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000
19	2.959	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027	0.000	0.000	
20	3.699	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034	0.000	0.000		
21	4.932	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045	0.000	0.000			
22	4.192	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000				
23	3.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036	0.000	0.000					
24	2.466	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023	0.000	0.000						
25	1.480	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.000	0.000							
26	0.986	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000								
27	0.986	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000									
28	0.493	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000										
29	0.740	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.000	0.000											
30	0.493	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000												
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000													
32	0.248	0.000	0.000	0.000	0.002	0.000	0.000														
33	0.495	0.000	0.000	0.005	0.000	0.000															
34	0.000	0.000	0.000	0.000	0.000																
35	0.000	0.000	0.000	0.000																	
36	0.000	0.000	0.000																		
37	0.000	0.000																			
38	0.000																				

Age Started	21	22	23
15	0.131	0.000	0.000
16	0.000	0.000	
17	0.000		

Table C2 (cont)

Age 39 Prevalence 54.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.238	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.928	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	4.173	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000
18	4.173	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000
19	2.946	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.014
20	3.682	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034	0.000	0.000	0.017	
21	4.910	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045	0.000	0.000	0.023		
22	4.173	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.019			
23	3.928	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036	0.000	0.000	0.018				
24	2.455	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.011					
25	1.473	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.007						
26	0.982	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005							
27	0.982	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005								
28	0.491	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002									
29	0.736	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.003										
30	0.491	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002											
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000												
32	0.247	0.000	0.000	0.000	0.002	0.000	0.000	0.001													
33	0.493	0.000	0.000	0.005	0.000	0.000	0.002														
34	0.000	0.000	0.000	0.000	0.000	0.000															
35	0.000	0.000	0.000	0.000	0.000																
36	0.000	0.000	0.000	0.000																	
37	0.000	0.000	0.000																		
38	0.000	0.000																			
39	0.000																				

	21	22	23	24
15	0.131	0.000	0.000	0.065
16	0.000	0.000	0.018	
17	0.000	0.019		
18	0.019			

Table C2 (cont)

Age 40 Prevalence 54.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	14.173	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	3.910	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	4.154	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000
18	4.154	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000
19	2.932	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.014
20	3.665	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034	0.000	0.000	0.017	0.017
21	4.887	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045	0.000	0.000	0.023	0.023	
22	4.154	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.019	0.019		
23	3.910	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036	0.000	0.000	0.018	0.018			
24	2.444	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.011	0.011				
25	1.466	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.007	0.007					
26	0.977	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005	0.005						
27	0.977	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005	0.005							
28	0.489	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.002								
29	0.733	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.003	0.003									
30	0.489	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.002										
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000											
32	0.245	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.001												
33	0.491	0.000	0.000	0.005	0.000	0.000	0.002	0.002													
34	0.000	0.000	0.000	0.000	0.000	0.000	0.000														
35	0.000	0.000	0.000	0.000	0.000	0.000															
36	0.000	0.000	0.000	0.000	0.000																
37	0.000	0.000	0.000	0.000																	
38	0.000	0.000	0.000																		
39	0.000	0.000																			
40	0.000																				

	21	22	23	24	25
15	0.131	0.000	0.000	0.065	0.065
16	0.000	0.000	0.018	0.018	
17	0.000	0.019	0.019		
18	0.019	0.019			
19	0.014				

Table C2 (cont)

Age 70 Prevalence 28.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	7.332	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.066	0.000	0.000	0.000	0.000
16	2.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036
17	2.149	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000
18	2.149	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000
19	1.517	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.027	0.000	0.000	0.014
20	1.896	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.000	0.000	0.000	0.000	0.034	0.000	0.000	0.017	0.017
21	2.528	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.000	0.000	0.045	0.000	0.000	0.023	0.023	0.000
22	2.149	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.038	0.000	0.000	0.019	0.019	0.000	0.000
23	2.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.018	0.000	0.000	0.000	0.000	0.036	0.000	0.000	0.018	0.018	0.000	0.000	0.036
24	1.264	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.023	0.000	0.000	0.011	0.011	0.000	0.000	0.022	0.011
25	0.758	0.000	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.007	0.007	0.000	0.000	0.013	0.007	0.034
26	0.506	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005	0.005	0.000	0.000	0.009	0.004	0.022	0.009
27	0.506	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.009	0.000	0.000	0.005	0.005	0.000	0.000	0.009	0.004	0.022	0.009	0.000
28	0.253	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.002	0.000	0.000	0.004	0.002	0.011	0.004	0.000	0.004
29	0.379	0.000	0.003	0.000	0.000	0.000	0.000	0.007	0.000	0.000	0.003	0.003	0.000	0.000	0.007	0.003	0.017	0.007	0.000	0.007	0.007
30	0.253	0.002	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.002	0.002	0.000	0.000	0.004	0.002	0.011	0.004	0.000	0.004	0.004	0.011
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32	0.127	0.000	0.000	0.000	0.002	0.000	0.000	0.001	0.001	0.000	0.000	0.002	0.001	0.006	0.002	0.000	0.002	0.002	0.006	0.002	0.000
33	0.254	0.000	0.000	0.005	0.000	0.000	0.002	0.002	0.000	0.000	0.005	0.002	0.011	0.005	0.000	0.005	0.005	0.011	0.005	0.000	0.002
34	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
35	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
36	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
42	0.129	0.002	0.001	0.006	0.002	0.000	0.002	0.002	0.006	0.002	0.000	0.001	0.005	0.009	0.000	0.010	0.001	0.005	0.010	0.002	0.002
43	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
44	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
45	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
46	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
47	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
48	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
49	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
51	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
52	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		
53	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000			
54	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
55	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
56	0.306	0.024	0.003	0.011	0.024	0.005	0.005	0.008	0.016	0.022	0.022	0.013	0.008	0.019	0.013						
57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
58	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000								
59	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000									
60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000										
61	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000											
62	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000												
63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000													
64	0.000	0.000	0.000	0.000	0.000	0.000	0.000														
65	0.000	0.000	0.000	0.000	0.000	0.000															
66	0.000	0.000	0.000	0.000	0.000																
67	0.000	0.000	0.000	0.000																	
68	0.000	0.000	0.000																		
69	0.000	0.000																			
70	0.000																				

Table C2 (cont)

Age 70 (continued)

Age Started	Ex Smokers, Duration 21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
15	0.131	0.000	0.000	0.065	0.065	0.000	0.000	0.130	0.065	0.325	0.130	0.000	0.130	0.130	0.325	0.130	0.000	0.065	0.260	0.520
16	0.000	0.000	0.018	0.018	0.000	0.000	0.036	0.018	0.090	0.036	0.000	0.036	0.036	0.090	0.036	0.000	0.018	0.072	0.143	0.000
17	0.000	0.019	0.019	0.000	0.000	0.038	0.019	0.095	0.038	0.000	0.038	0.038	0.095	0.038	0.000	0.019	0.076	0.152	0.000	0.170
18	0.019	0.019	0.000	0.000	0.038	0.019	0.095	0.038	0.000	0.038	0.038	0.095	0.038	0.000	0.019	0.076	0.152	0.000	0.170	0.019
19	0.014	0.000	0.000	0.027	0.013	0.067	0.027	0.000	0.027	0.027	0.067	0.027	0.000	0.013	0.054	0.108	0.000	0.120	0.013	0.053
20	0.000	0.000	0.034	0.017	0.084	0.034	0.000	0.034	0.034	0.084	0.034	0.000	0.017	0.067	0.135	0.000	0.150	0.017	0.067	0.150
21	0.000	0.045	0.022	0.112	0.045	0.000	0.045	0.045	0.112	0.045	0.000	0.022	0.090	0.179	0.000	0.200	0.022	0.089	0.200	0.044
22	0.038	0.019	0.095	0.038	0.000	0.038	0.038	0.095	0.038	0.000	0.019	0.076	0.152	0.000	0.170	0.019	0.075	0.170	0.038	0.038
23	0.018	0.090	0.036	0.000	0.036	0.036	0.090	0.036	0.000	0.018	0.072	0.143	0.000	0.160	0.018	0.071	0.160	0.035	0.035	0.053
24	0.056	0.022	0.000	0.022	0.022	0.056	0.022	0.000	0.011	0.045	0.090	0.000	0.100	0.011	0.044	0.100	0.022	0.022	0.033	0.067
25	0.013	0.000	0.013	0.013	0.034	0.013	0.000	0.007	0.027	0.054	0.000	0.060	0.007	0.027	0.060	0.013	0.013	0.020	0.040	0.053
26	0.000	0.009	0.009	0.022	0.009	0.000	0.004	0.018	0.036	0.000	0.040	0.004	0.018	0.040	0.009	0.009	0.013	0.027	0.035	0.035
27	0.009	0.009	0.022	0.009	0.000	0.004	0.018	0.036	0.000	0.040	0.004	0.018	0.040	0.009	0.009	0.013	0.027	0.035	0.035	0.022
28	0.004	0.011	0.004	0.000	0.002	0.009	0.018	0.000	0.020	0.002	0.009	0.020	0.004	0.004	0.007	0.013	0.018	0.018	0.011	0.007
29	0.017	0.007	0.000	0.003	0.013	0.027	0.000	0.030	0.003	0.013	0.030	0.007	0.007	0.010	0.020	0.027	0.027	0.017	0.010	0.023
30	0.004	0.000	0.002	0.009	0.018	0.000	0.020	0.002	0.009	0.020	0.004	0.004	0.007	0.013	0.018	0.018	0.011	0.007	0.016	0.011
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
32	0.001	0.005	0.009	0.000	0.010	0.001	0.004	0.010	0.002	0.002	0.003	0.007	0.009	0.009	0.006	0.003	0.008	0.006		
33	0.009	0.018	0.000	0.020	0.002	0.009	0.020	0.004	0.004	0.007	0.013	0.018	0.018	0.011	0.007	0.016	0.011			
34	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000				
35	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000					
36	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000						
37	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000							
38	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000								
39	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000									
40	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000										
41	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000											
42	0.003	0.007	0.009	0.009	0.006	0.003	0.008	0.006												
43	0.000	0.000	0.000	0.000	0.000	0.000	0.000													
44	0.000	0.000	0.000	0.000	0.000	0.000														
45	0.000	0.000	0.000	0.000	0.000															
46	0.000	0.000	0.000	0.000																
47	0.000	0.000	0.000																	
48	0.000	0.000																		
49	0.000																			

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
15	0.000	0.579	0.064	0.257	0.579	0.129	0.129	0.193	0.386	0.515	0.515	0.322	0.193	0.450	0.322
16	0.160	0.018	0.071	0.160	0.035	0.035	0.053	0.106	0.142	0.142	0.089	0.053	0.124	0.089	
17	0.019	0.075	0.170	0.038	0.038	0.057	0.113	0.151	0.151	0.094	0.057	0.132	0.094		
18	0.075	0.170	0.038	0.038	0.057	0.113	0.151	0.151	0.094	0.057	0.132	0.094			
19	0.120	0.027	0.027	0.040	0.080	0.106	0.106	0.067	0.040	0.093	0.067				
20	0.033	0.033	0.050	0.100	0.133	0.133	0.083	0.050	0.116	0.083					
21	0.044	0.067	0.133	0.177	0.177	0.111	0.067	0.155	0.111						
22	0.057	0.113	0.151	0.151	0.094	0.057	0.132	0.094							
23	0.106	0.142	0.142	0.089	0.053	0.124	0.089								
24	0.089	0.089	0.055	0.033	0.078	0.055									
25	0.053	0.033	0.020	0.047	0.033										
26	0.022	0.013	0.031	0.022											
27	0.013	0.031	0.022												
28	0.016	0.011													
29	0.017														

Table C3

Example working of Swartz smoking model with drift, Male 1900 cohort
(See explanation p C2)

Age 15 Prevalence 14.50

Age Started	Current Smokers
15	14.500

Age 16 Prevalence 18.50

Age Started	Current Smokers	Ex Smokers, Duration 1
15	14.500	0.000
16	4.000	

Age 16 Drift

15	14.427	0.073
16	4.073	

Age 17 Prevalence 22.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2
15	14.427	0.073	0.000
16	4.073	0.000	
17	4.250		

Age 17 Drift

15	14.355	0.073	0.072
16	4.052	0.020	
17	4.342		

Age 18 Prevalence 27.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3
15	14.355	0.073	0.072	0.000
16	4.052	0.020	0.000	
17	4.342	0.000		
18	4.250			

Age 18 Drift

15	14.284	0.073	0.072	0.072
16	4.032	0.020	0.020	
17	4.321	0.022		
18	4.364			

Age 19 Prevalence 30.00

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4
15	14.284	0.073	0.072	0.072	0.000
16	4.032	0.020	0.020	0.000	
17	4.321	0.022	0.000		
18	4.364	0.000			
19	3.000				

Age 19 Drift

15	14.212	0.073	0.072	0.072	0.071
16	4.012	0.020	0.020	0.020	
17	4.299	0.022	0.022		
18	4.342	0.022			
19	3.135				

Age 20 Prevalence 33.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5
15	14.212	0.073	0.072	0.072	0.071	0.000
16	4.012	0.020	0.020	0.020	0.000	
17	4.299	0.022	0.022	0.000		
18	4.342	0.022	0.000			
19	3.135	0.000				
20	3.750					

Age 20 Drift

15	14.141	0.073	0.072	0.072	0.071	0.071
16	3.992	0.020	0.020	0.020	0.020	
17	4.278	0.022	0.022	0.021		
18	4.320	0.022	0.022			
19	3.119	0.016				
20	3.900					

Table C3 (cont)

Age 30 Prevalence 54.75

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
15	13.517	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.000
16	3.816	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.000	
17	4.089	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.000		
18	4.130	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.000			
19	2.982	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.000				
20	3.728	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.000					
21	4.966	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.000						
22	4.291	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.000							
23	4.090	0.021	0.021	0.021	0.021	0.021	0.021	0.000								
24	2.667	0.014	0.014	0.014	0.013	0.013	0.000									
25	1.713	0.009	0.009	0.009	0.009	0.000										
26	1.236	0.006	0.006	0.006	0.000											
27	1.247	0.006	0.006	0.000												
28	0.761	0.004	0.000													
29	1.018	0.000														
30	0.500															

Age 30 Drift

15	13.450	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.068
16	3.797	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019	
17	4.069	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.020		
18	4.109	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021			
19	2.967	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015				
20	3.709	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019					
21	4.941	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025						
22	4.269	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.021							
23	4.070	0.021	0.021	0.021	0.021	0.021	0.021	0.020								
24	2.654	0.014	0.014	0.014	0.013	0.013	0.013									
25	1.704	0.009	0.009	0.009	0.009	0.009										
26	1.230	0.006	0.006	0.006	0.006											
27	1.241	0.006	0.006	0.006												
28	0.757	0.004	0.004													
29	1.012	0.005														
30	0.771															

Table C3 (cont)

ıge 40 Prevalence 54.25

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	12.566	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.068	0.128	0.067	0.066	0.066	0.066
16	3.547	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019	0.036	0.019	0.019	0.019	0.019	0.052
17	3.801	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019
18	3.839	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.039	0.020	0.020	0.020	0.020	0.056	0.020	0.020
19	2.772	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.028	0.015	0.015	0.015	0.014	0.040	0.014	0.014	0.027
20	3.466	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.035	0.018	0.018	0.018	0.018	0.050	0.018	0.018	0.034	0.016
21	4.616	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025	0.047	0.024	0.024	0.024	0.024	0.067	0.024	0.024	0.045	0.021	
22	3.989	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.021	0.041	0.021	0.021	0.021	0.021	0.058	0.020	0.020	0.039	0.018		
23	3.802	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019	0.019	0.037	0.018			
24	2.480	0.014	0.014	0.014	0.013	0.013	0.013	0.025	0.013	0.013	0.013	0.013	0.036	0.013	0.013	0.024	0.011				
25	1.592	0.009	0.009	0.009	0.009	0.009	0.016	0.008	0.008	0.008	0.008	0.023	0.008	0.008	0.015	0.007					
26	1.149	0.006	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.005						
27	1.160	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.005							
28	0.708	0.004	0.004	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.003								
29	0.946	0.005	0.010	0.005	0.005	0.005	0.005	0.014	0.005	0.005	0.009	0.004									
30	0.721	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.003										
31	0.257	0.001	0.001	0.001	0.001	0.004	0.001	0.001	0.002	0.001											
32	0.495	0.003	0.003	0.003	0.007	0.003	0.003	0.005	0.002												
33	0.737	0.004	0.004	0.011	0.004	0.004	0.007	0.003													
34	0.265	0.001	0.004	0.001	0.001	0.003	0.001														
35	0.266	0.004	0.001	0.001	0.003	0.001															
36	0.267	0.001	0.001	0.003	0.001																
37	0.269	0.001	0.003	0.001																	
38	0.270	0.003	0.001																		
39	0.271	0.001																			
40	0.000																				

	21	22	23	24	25
15	0.183	0.064	0.064	0.122	0.058
16	0.018	0.018	0.034	0.016	
17	0.019	0.037	0.018		
18	0.037	0.018			
19	0.013				

Table C3 (cont)

Age 40 Age Started	Drift Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	12.503	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.068	0.128	0.067	0.066	0.066	0.066
16	3.529	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019	0.036	0.019	0.019	0.019	0.019	0.052
17	3.782	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019
18	3.820	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.039	0.020	0.020	0.020	0.020	0.056	0.020	0.020
19	2.758	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.028	0.015	0.015	0.015	0.014	0.040	0.014	0.014	0.027
20	3.448	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.035	0.018	0.018	0.018	0.018	0.050	0.018	0.018	0.034	0.033
21	4.593	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025	0.047	0.024	0.024	0.024	0.024	0.067	0.024	0.024	0.045	0.044	
22	3.969	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.021	0.041	0.021	0.021	0.021	0.021	0.058	0.020	0.020	0.039	0.038		
23	3.783	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019	0.019	0.037	0.037			
24	2.467	0.014	0.014	0.014	0.013	0.013	0.013	0.025	0.013	0.013	0.013	0.013	0.036	0.013	0.013	0.024	0.024				
25	1.584	0.009	0.009	0.009	0.009	0.009	0.016	0.008	0.008	0.008	0.008	0.023	0.008	0.008	0.015	0.015					
26	1.144	0.006	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011						
27	1.154	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011							
28	0.704	0.004	0.004	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007								
29	0.941	0.005	0.010	0.005	0.005	0.005	0.005	0.014	0.005	0.005	0.009	0.009									
30	0.717	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007										
31	0.256	0.001	0.001	0.001	0.001	0.004	0.001	0.001	0.002	0.002											
32	0.493	0.003	0.003	0.003	0.007	0.003	0.003	0.005	0.005												
33	0.734	0.004	0.004	0.011	0.004	0.004	0.007	0.007													
34	0.263	0.001	0.004	0.001	0.001	0.003	0.003														
35	0.265	0.004	0.001	0.001	0.003	0.003															
36	0.266	0.001	0.001	0.003	0.003																
37	0.267	0.001	0.003	0.003																	
38	0.269	0.003	0.003																		
39	0.270	0.003																			
40	0.271																				
15		0.183	0.064	0.064	0.122	0.121															
16		0.018	0.018	0.034	0.034																
17		0.019	0.037	0.037																	
18		0.037	0.037																		
19		0.027																			

Table C3 (cont)

Age 70 Prevalence 28.50

Age Started	Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	5.593	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.068	0.128	0.067	0.066	0.066	0.066
16	1.579	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019	0.036	0.019	0.019	0.019	0.019	0.052
17	1.692	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019
18	1.709	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.039	0.020	0.020	0.020	0.020	0.056	0.020	0.020
19	1.234	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.028	0.015	0.015	0.015	0.014	0.040	0.014	0.014	0.027
20	1.543	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.035	0.018	0.018	0.018	0.018	0.050	0.018	0.018	0.034	0.033
21	2.055	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025	0.047	0.024	0.024	0.024	0.024	0.067	0.024	0.024	0.045	0.044	0.023
22	1.775	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.021	0.041	0.021	0.021	0.021	0.021	0.058	0.020	0.020	0.039	0.038	0.020	0.020
23	1.692	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019	0.019	0.037	0.037	0.019	0.019	0.053
24	1.104	0.014	0.014	0.014	0.013	0.013	0.013	0.025	0.013	0.013	0.013	0.013	0.036	0.013	0.013	0.024	0.024	0.012	0.012	0.035	0.023
25	0.709	0.009	0.009	0.009	0.009	0.009	0.016	0.008	0.008	0.008	0.008	0.023	0.008	0.008	0.015	0.015	0.008	0.008	0.022	0.015	0.043
26	0.512	0.006	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011	0.006	0.006	0.016	0.011	0.031	0.016
27	0.516	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011	0.006	0.006	0.016	0.011	0.031	0.016	0.005
28	0.315	0.004	0.004	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.019	0.010	0.003	0.009
29	0.421	0.005	0.010	0.005	0.005	0.005	0.005	0.014	0.005	0.005	0.009	0.009	0.005	0.005	0.013	0.009	0.026	0.013	0.004	0.013	0.013
30	0.321	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.020	0.010	0.003	0.010	0.010	0.019
31	0.114	0.001	0.001	0.001	0.001	0.004	0.001	0.001	0.002	0.002	0.001	0.001	0.004	0.002	0.007	0.003	0.001	0.003	0.003	0.007	0.003
32	0.220	0.003	0.003	0.003	0.007	0.003	0.003	0.005	0.005	0.002	0.002	0.007	0.005	0.013	0.007	0.002	0.007	0.007	0.013	0.006	0.002
33	0.328	0.004	0.004	0.011	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.020	0.010	0.003	0.010	0.010	0.019	0.010	0.003	0.006
34	0.118	0.001	0.004	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006
35	0.118	0.004	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010
36	0.119	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001
37	0.120	0.001	0.003	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001	0.011
38	0.120	0.003	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001	0.011	0.002
39	0.121	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005
40	0.121	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005	0.011
41	0.122	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005	0.011	0.003
42	0.236	0.007	0.005	0.014	0.007	0.002	0.007	0.007	0.014	0.007	0.002	0.005	0.011	0.020	0.002	0.022	0.004	0.011	0.021	0.006	0.006
43	0.124	0.003	0.008	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.011	0.002	0.006	0.011	0.003	0.003	0.004
44	0.124	0.008	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008
45	0.125	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010
46	0.126	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010
47	0.126	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010	0.006
48	0.127	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004
49	0.128	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009
50	0.128	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009	0.006
51	0.129	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009	0.006	
52	0.129	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006		
53	0.130	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006			
54	0.131	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006				
55	0.131	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.005	0.008	0.010	0.010	0.007	0.004	0.009	0.006					
56	0.419	0.039	0.007	0.019	0.038	0.011	0.011	0.014	0.026	0.033	0.033	0.021	0.013	0.028	0.018						
57	0.134	0.002	0.006	0.012	0.003	0.003	0.005	0.008	0.011	0.010	0.007	0.004	0.009	0.006							
58	0.135	0.006	0.012	0.003	0.003	0.005	0.008	0.011	0.010	0.007	0.004	0.009	0.006								
59	0.136	0.012	0.003	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.006									
60	0.136	0.004	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.006										
61	0.137	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.006											
62	0.138	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.006												
63	0.138	0.008	0.011	0.011	0.007	0.004	0.009	0.006													
64	0.139	0.011	0.011	0.007	0.004	0.009	0.006														
65	0.140	0.011	0.007	0.004	0.009	0.006															
66	0.140	0.007	0.005	0.009	0.006																
67	0.141	0.005	0.009	0.006																	
68	0.142	0.009	0.006																		
69	0.143	0.006																			
70	0.000																				

Table C3 (cont)

Age 70 (cont)

Age Started	Ex Smokers, Duration 21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
15	0.183	0.064	0.064	0.122	0.121	0.063	0.062	0.175	0.117	0.340	0.170	0.058	0.168	0.166	0.329	0.163	0.053	0.107	0.267	0.477
16	0.018	0.018	0.034	0.034	0.018	0.018	0.049	0.033	0.096	0.048	0.016	0.047	0.047	0.093	0.046	0.015	0.030	0.075	0.135	0.014
17	0.019	0.037	0.037	0.019	0.019	0.053	0.035	0.103	0.051	0.018	0.051	0.050	0.099	0.049	0.016	0.032	0.081	0.144	0.015	0.157
18	0.037	0.037	0.019	0.019	0.053	0.036	0.104	0.052	0.018	0.051	0.051	0.100	0.050	0.016	0.033	0.081	0.146	0.015	0.158	0.030
19	0.027	0.014	0.014	0.039	0.026	0.075	0.038	0.013	0.037	0.037	0.073	0.036	0.012	0.024	0.059	0.105	0.011	0.114	0.022	0.056
20	0.017	0.017	0.048	0.032	0.094	0.047	0.016	0.046	0.046	0.091	0.045	0.015	0.029	0.074	0.132	0.013	0.143	0.027	0.069	0.140
21	0.023	0.064	0.043	0.125	0.062	0.021	0.062	0.061	0.121	0.060	0.020	0.039	0.098	0.175	0.018	0.190	0.036	0.093	0.186	0.053
22	0.056	0.037	0.108	0.054	0.018	0.053	0.053	0.104	0.052	0.017	0.034	0.085	0.151	0.016	0.164	0.031	0.080	0.161	0.046	0.045
23	0.035	0.103	0.051	0.018	0.051	0.050	0.099	0.049	0.016	0.032	0.081	0.144	0.015	0.157	0.030	0.076	0.153	0.044	0.043	0.058
24	0.067	0.034	0.011	0.033	0.033	0.065	0.032	0.011	0.021	0.053	0.094	0.010	0.102	0.019	0.050	0.100	0.028	0.028	0.038	0.067
25	0.022	0.007	0.021	0.021	0.042	0.021	0.007	0.014	0.034	0.060	0.006	0.066	0.012	0.032	0.064	0.018	0.018	0.024	0.043	0.056
26	0.005	0.015	0.015	0.030	0.015	0.005	0.010	0.024	0.044	0.004	0.047	0.009	0.023	0.046	0.013	0.013	0.018	0.031	0.040	0.040
27	0.015	0.015	0.030	0.015	0.005	0.010	0.025	0.044	0.005	0.048	0.009	0.023	0.047	0.013	0.013	0.018	0.031	0.040	0.040	0.026
28	0.009	0.019	0.009	0.003	0.006	0.015	0.027	0.003	0.029	0.006	0.014	0.029	0.008	0.008	0.011	0.019	0.025	0.024	0.016	0.010
29	0.025	0.012	0.004	0.008	0.020	0.036	0.004	0.039	0.007	0.019	0.038	0.011	0.011	0.014	0.026	0.033	0.033	0.021	0.014	0.028
30	0.009	0.003	0.006	0.015	0.027	0.003	0.030	0.006	0.014	0.029	0.008	0.008	0.011	0.020	0.025	0.025	0.016	0.010	0.021	0.014
31	0.001	0.002	0.005	0.010	0.001	0.011	0.002	0.005	0.010	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005	
32	0.004	0.011	0.019	0.002	0.020	0.004	0.010	0.020	0.006	0.006	0.008	0.013	0.017	0.017	0.011	0.007	0.015	0.010		
32	0.016	0.028	0.003	0.030	0.006	0.015	0.030	0.008	0.008	0.011	0.020	0.026	0.025	0.016	0.011	0.022	0.014			
33	0.010	0.001	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005				
34	0.001	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005					
35	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005						
36	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005							
37	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005								
38	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.005									
39	0.003	0.003	0.004	0.007	0.010	0.009	0.006	0.004	0.008	0.005										
40	0.003	0.004	0.007	0.010	0.009	0.006	0.004	0.008	0.005											
41	0.008	0.014	0.018	0.018	0.012	0.008	0.016	0.010												
42	0.008	0.010	0.010	0.006	0.004	0.008	0.005													
43	0.010	0.010	0.006	0.004	0.008	0.005														
44	0.010	0.006	0.004	0.008	0.005															
45	0.006	0.004	0.008	0.006																
46	0.004	0.008	0.006																	
47	0.008	0.006																		
48	0.006																			

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
15	0.049	0.518	0.098	0.252	0.506	0.144	0.143	0.192	0.341	0.438	0.434	0.281	0.180	0.374	0.245
16	0.146	0.028	0.071	0.143	0.041	0.040	0.054	0.096	0.124	0.123	0.079	0.051	0.106	0.069	
17	0.030	0.076	0.153	0.044	0.043	0.058	0.103	0.133	0.131	0.085	0.054	0.113	0.074		
18	0.077	0.155	0.044	0.044	0.059	0.104	0.134	0.133	0.086	0.055	0.114	0.075			
19	0.112	0.032	0.031	0.042	0.075	0.097	0.096	0.062	0.040	0.083	0.054				
20	0.040	0.039	0.053	0.094	0.121	0.120	0.078	0.050	0.103	0.068					
21	0.052	0.071	0.125	0.161	0.159	0.103	0.066	0.138	0.090						
22	0.061	0.108	0.139	0.138	0.089	0.057	0.119	0.078							
23	0.103	0.133	0.131	0.085	0.054	0.113	0.074								
24	0.086	0.086	0.055	0.035	0.074	0.048									
25	0.055	0.036	0.023	0.047	0.031										
26	0.026	0.016	0.034	0.022											
27	0.017	0.035	0.023												
28	0.021	0.014													
29	0.018														

Table C3 (cont)

Age 70 Age Started	Drift Current Smokers	Ex Smokers, Duration 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
15	5.565	0.073	0.072	0.072	0.071	0.071	0.071	0.070	0.070	0.070	0.069	0.069	0.069	0.068	0.068	0.068	0.128	0.067	0.066	0.066	0.066
16	1.571	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.020	0.019	0.019	0.019	0.019	0.019	0.036	0.019	0.019	0.019	0.019	0.052
17	1.683	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019
18	1.700	0.022	0.022	0.022	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.021	0.039	0.020	0.020	0.020	0.020	0.056	0.020	0.020
19	1.228	0.016	0.016	0.016	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.015	0.028	0.015	0.015	0.015	0.014	0.040	0.014	0.014	0.027
20	1.535	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.019	0.035	0.018	0.018	0.018	0.018	0.050	0.018	0.018	0.034	0.033
21	2.044	0.026	0.026	0.026	0.025	0.025	0.025	0.025	0.025	0.025	0.047	0.024	0.024	0.024	0.024	0.067	0.024	0.024	0.045	0.044	0.023
22	1.766	0.022	0.022	0.022	0.022	0.022	0.022	0.022	0.021	0.041	0.021	0.021	0.021	0.021	0.058	0.020	0.020	0.039	0.038	0.020	0.020
23	1.684	0.021	0.021	0.021	0.021	0.021	0.021	0.020	0.039	0.020	0.020	0.020	0.020	0.055	0.019	0.019	0.037	0.037	0.019	0.019	0.053
24	1.098	0.014	0.014	0.014	0.013	0.013	0.013	0.025	0.013	0.013	0.013	0.013	0.036	0.013	0.013	0.024	0.024	0.012	0.012	0.035	0.023
25	0.705	0.009	0.009	0.009	0.009	0.009	0.016	0.008	0.008	0.008	0.008	0.023	0.008	0.008	0.015	0.015	0.008	0.008	0.022	0.015	0.043
26	0.509	0.006	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011	0.006	0.006	0.016	0.011	0.031	0.016
27	0.514	0.006	0.006	0.006	0.012	0.006	0.006	0.006	0.006	0.017	0.006	0.006	0.011	0.011	0.006	0.006	0.016	0.011	0.031	0.016	0.005
28	0.313	0.004	0.004	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.019	0.010	0.003	0.009
29	0.419	0.005	0.010	0.005	0.005	0.005	0.005	0.014	0.005	0.005	0.009	0.009	0.005	0.005	0.013	0.009	0.026	0.013	0.004	0.013	0.013
30	0.319	0.007	0.004	0.004	0.004	0.004	0.010	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.020	0.010	0.003	0.010	0.010	0.019
31	0.114	0.001	0.001	0.001	0.001	0.004	0.001	0.001	0.002	0.002	0.001	0.001	0.004	0.002	0.007	0.003	0.001	0.003	0.003	0.007	0.003
32	0.219	0.003	0.003	0.003	0.007	0.003	0.003	0.005	0.005	0.002	0.002	0.007	0.005	0.013	0.007	0.002	0.007	0.007	0.013	0.006	0.002
33	0.326	0.004	0.004	0.011	0.004	0.004	0.007	0.007	0.004	0.004	0.010	0.007	0.020	0.010	0.003	0.010	0.010	0.019	0.010	0.003	0.006
34	0.117	0.001	0.004	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006
35	0.118	0.004	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010
36	0.118	0.001	0.001	0.003	0.003	0.001	0.001	0.004	0.002	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001
37	0.119	0.001	0.003	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001	0.011
38	0.120	0.003	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.003	0.001	0.002	0.006	0.010	0.001	0.011	0.002
39	0.120	0.003	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005
40	0.121	0.001	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005	0.011
41	0.121	0.001	0.004	0.003	0.007	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.010	0.001	0.011	0.002	0.005	0.011	0.003
42	0.234	0.007	0.005	0.014	0.007	0.002	0.007	0.007	0.014	0.007	0.002	0.005	0.011	0.020	0.002	0.022	0.004	0.011	0.021	0.006	0.006
43	0.123	0.003	0.008	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.011	0.002	0.006	0.011	0.003	0.003	0.004
44	0.124	0.008	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008
45	0.124	0.004	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010
46	0.125	0.001	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010
47	0.126	0.004	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010	0.006
48	0.126	0.004	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.011	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004
49	0.127	0.007	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009
50	0.128	0.004	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009	0.006
51	0.128	0.001	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.006	0.004	0.009	0.006	
52	0.129	0.002	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006		
53	0.129	0.006	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006			
54	0.130	0.011	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.004	0.008	0.010	0.010	0.007	0.004	0.009	0.006				
55	0.131	0.001	0.012	0.002	0.006	0.012	0.003	0.003	0.005	0.008	0.010	0.010	0.007	0.004	0.009	0.006					
56	0.417	0.039	0.007	0.019	0.038	0.011	0.011	0.014	0.026	0.033	0.033	0.021	0.013	0.028	0.020						
57	0.134	0.002	0.006	0.012	0.003	0.003	0.005	0.008	0.011	0.010	0.007	0.004	0.009	0.007							
58	0.134	0.006	0.012	0.003	0.003	0.005	0.008	0.011	0.010	0.007	0.004	0.009	0.007								
59	0.135	0.012	0.003	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.007									
60	0.136	0.004	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.007										
61	0.136	0.003	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.007											
62	0.137	0.005	0.008	0.011	0.011	0.007	0.004	0.009	0.007												
63	0.138	0.008	0.011	0.011	0.007	0.004	0.009	0.007													
64	0.138	0.011	0.011	0.007	0.004	0.009	0.007														
65	0.139	0.011	0.007	0.004	0.009	0.007															
66	0.140	0.007	0.005	0.009	0.007																
67	0.140	0.005	0.009	0.007																	
68	0.141	0.009	0.007																		
69	0.142	0.007																			
70	0.143																				

Table C3 (cont)

Age 70 Drift (cont)

Age Started	Ex Smokers, Duration 21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
15	0.183	0.064	0.064	0.122	0.121	0.063	0.062	0.175	0.117	0.340	0.170	0.058	0.168	0.166	0.329	0.163	0.053	0.107	0.267	0.477
16	0.018	0.018	0.034	0.034	0.018	0.018	0.049	0.033	0.096	0.048	0.016	0.047	0.047	0.093	0.046	0.015	0.030	0.075	0.135	0.014
17	0.019	0.037	0.037	0.019	0.019	0.053	0.035	0.103	0.051	0.018	0.051	0.050	0.099	0.049	0.016	0.032	0.081	0.144	0.015	0.157
18	0.037	0.037	0.019	0.019	0.053	0.036	0.104	0.052	0.018	0.051	0.051	0.100	0.050	0.016	0.033	0.081	0.146	0.015	0.158	0.030
19	0.027	0.014	0.014	0.039	0.026	0.075	0.038	0.013	0.037	0.037	0.073	0.036	0.012	0.024	0.059	0.105	0.011	0.114	0.022	0.056
20	0.017	0.017	0.048	0.032	0.094	0.047	0.016	0.046	0.046	0.091	0.045	0.015	0.029	0.074	0.132	0.013	0.143	0.027	0.069	0.140
21	0.023	0.064	0.043	0.125	0.062	0.021	0.062	0.061	0.121	0.060	0.020	0.039	0.098	0.175	0.018	0.190	0.036	0.093	0.186	0.053
22	0.056	0.037	0.108	0.054	0.018	0.053	0.053	0.104	0.052	0.017	0.034	0.085	0.151	0.016	0.164	0.031	0.080	0.161	0.046	0.045
23	0.035	0.103	0.051	0.018	0.051	0.050	0.099	0.049	0.016	0.032	0.081	0.144	0.015	0.157	0.030	0.076	0.153	0.044	0.043	0.058
24	0.067	0.034	0.011	0.033	0.033	0.065	0.032	0.011	0.021	0.053	0.094	0.010	0.102	0.019	0.050	0.100	0.028	0.028	0.038	0.067
25	0.022	0.007	0.021	0.021	0.042	0.021	0.007	0.014	0.034	0.060	0.006	0.066	0.012	0.032	0.064	0.018	0.018	0.024	0.043	0.056
26	0.005	0.015	0.015	0.030	0.015	0.005	0.010	0.024	0.044	0.004	0.047	0.009	0.023	0.046	0.013	0.013	0.018	0.031	0.040	0.040
27	0.015	0.015	0.030	0.015	0.005	0.010	0.025	0.044	0.005	0.048	0.009	0.023	0.047	0.013	0.013	0.018	0.031	0.040	0.040	0.026
28	0.009	0.019	0.009	0.003	0.006	0.015	0.027	0.003	0.029	0.006	0.014	0.029	0.008	0.008	0.011	0.019	0.025	0.024	0.016	0.010
29	0.025	0.012	0.004	0.008	0.020	0.036	0.004	0.039	0.007	0.019	0.038	0.011	0.011	0.014	0.026	0.033	0.033	0.021	0.014	0.028
30	0.009	0.003	0.006	0.015	0.027	0.003	0.030	0.006	0.014	0.029	0.008	0.008	0.011	0.020	0.025	0.025	0.016	0.010	0.021	0.016
31	0.001	0.002	0.005	0.010	0.001	0.011	0.002	0.005	0.010	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006	
32	0.004	0.011	0.019	0.002	0.020	0.004	0.010	0.020	0.006	0.006	0.008	0.013	0.017	0.017	0.011	0.007	0.015	0.011		
33	0.016	0.028	0.003	0.030	0.006	0.015	0.030	0.008	0.008	0.011	0.020	0.026	0.025	0.016	0.011	0.022	0.016			
34	0.010	0.001	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006				
35	0.001	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006					
36	0.011	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006						
37	0.002	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006							
38	0.005	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006								
39	0.011	0.003	0.003	0.004	0.007	0.009	0.009	0.006	0.004	0.008	0.006									
40	0.003	0.003	0.004	0.007	0.010	0.009	0.006	0.004	0.008	0.006										
41	0.003	0.004	0.007	0.010	0.009	0.006	0.004	0.008	0.006											
42	0.008	0.014	0.018	0.018	0.012	0.008	0.016	0.012												
43	0.008	0.010	0.010	0.006	0.004	0.008	0.006													
44	0.010	0.010	0.006	0.004	0.008	0.006														
45	0.010	0.006	0.004	0.008	0.006															
46	0.006	0.004	0.008	0.006																
47	0.004	0.008	0.006																	
48	0.008	0.006																		
49	0.006																			

	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55
15	0.049	0.518	0.098	0.252	0.506	0.144	0.143	0.192	0.341	0.438	0.434	0.281	0.180	0.374	0.273
16	0.146	0.028	0.071	0.143	0.041	0.040	0.054	0.096	0.124	0.123	0.079	0.051	0.106	0.077	
17	0.030	0.076	0.153	0.044	0.043	0.058	0.103	0.133	0.131	0.085	0.054	0.113	0.083		
18	0.077	0.155	0.044	0.044	0.059	0.104	0.134	0.133	0.086	0.055	0.114	0.083			
19	0.112	0.032	0.031	0.042	0.075	0.097	0.096	0.062	0.040	0.083	0.060				
20	0.040	0.039	0.053	0.094	0.121	0.120	0.078	0.050	0.103	0.075					
21	0.052	0.071	0.125	0.161	0.159	0.103	0.066	0.138	0.100						
22	0.061	0.108	0.139	0.138	0.089	0.057	0.119	0.087							
23	0.103	0.133	0.131	0.085	0.054	0.113	0.083								
24	0.086	0.086	0.055	0.035	0.074	0.054									
25	0.055	0.036	0.023	0.047	0.035										
26	0.026	0.016	0.034	0.025											
27	0.017	0.035	0.025												
28	0.021	0.015													
29	0.021														

Table C4

Example working of Townsend smoking model, Male 1900 cohort
(See explanation p C2)

Age 15 Prevalence 14.50

Group	%	Duration current	Duration lagged	Changes	Age started
1	85.500	0	0	0	
2	14.500	0	0	1	15

Age 16 Prevalence 18.50

Group	%	Duration current	Duration lagged	Changes	Age started
1	81.500	0	0	0	
2	14.500	1	0	1	15
3	4.000	0	0	1	16

Age 17 Prevalence 22.75

Group	%	Duration current	Duration lagged	Changes	Age started
1	77.250	0	0	0	
2	14.500	2	0	1	15
3	4.000	1	0	1	16
4	4.250	0	0	1	17

Age 18 Prevalence 27.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	73.000	0	0	0	
2	14.500	3	0	1	15
3	4.000	2	0	1	16
4	4.250	1	0	1	17
5	4.250	0	0	1	18

Age 19 Prevalence 30.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	70.000	0	0	0	
2	14.500	4	0	1	15
3	4.000	3	0	1	16
4	4.250	2	0	1	17
5	4.250	1	0	1	18
6	3.000	0	0	1	19

Age 20 Prevalence 33.75

Group	%	Duration current	Duration lagged	Changes	Age started
1	66.250	0	0	0	
2	14.500	5	0	1	15
3	4.000	4	0	1	16
4	4.250	3	0	1	17
5	4.250	2	0	1	18
6	3.000	1	0	1	19
7	3.750	0	0	1	20

Age 21 Prevalence 38.75

Group	%	Duration current	Duration lagged	Changes	Age started
1	61.250	0	0	0	
2	14.500	6	1	1	15
3	4.000	5	0	1	16
4	4.250	4	0	1	17
5	4.250	3	0	1	18
6	3.000	2	0	1	19
7	3.750	1	0	1	20
8	5.000	0	0	1	21

Age 22 Prevalence 43.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	57.000	0	0	0	
2	14.500	7	2	1	15
3	4.000	6	1	1	16
4	4.250	5	0	1	17
5	4.250	4	0	1	18
6	3.000	3	0	1	19
7	3.750	2	0	1	20
8	5.000	1	0	1	21
9	4.250	0	0	1	22

Table C4 (cont)

Age 23 Prevalence 47.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	53.000	0	0	0	
2	14.500	8	3	1	15
3	4.000	7	2	1	16
4	4.250	6	1	1	17
5	4.250	5	0	1	18
6	3.000	4	0	1	19
7	3.750	3	0	1	20
8	5.000	2	0	1	21
9	4.250	1	0	1	22
10	4.000	0	0	1	23

Age 24 Prevalence 49.50

Group	%	Duration current	Duration lagged	Changes	Age started
1	50.500	0	0	0	
2	14.500	9	4	1	15
3	4.000	8	3	1	16
4	4.250	7	2	1	17
5	4.250	6	1	1	18
6	3.000	5	0	1	19
7	3.750	4	0	1	20
8	5.000	3	0	1	21
9	4.250	2	0	1	22
10	4.000	1	0	1	23
11	2.500	0	0	1	24

Age 25 Prevalence 51.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	49.000	0	0	0	
2	14.500	10	5	1	15
3	4.000	9	4	1	16
4	4.250	8	3	1	17
5	4.250	7	2	1	18
6	3.000	6	1	1	19
7	3.750	5	0	1	20
8	5.000	4	0	1	21
9	4.250	3	0	1	22
10	4.000	2	0	1	23
11	2.500	1	0	1	24
12	1.500	0	0	1	25

Age 26 Prevalence 52.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	48.000	0	0	0	
2	14.500	11	6	1	15
3	4.000	10	5	1	16
4	4.250	9	4	1	17
5	4.250	8	3	1	18
6	3.000	7	2	1	19
7	3.750	6	1	1	20
8	5.000	5	0	1	21
9	4.250	4	0	1	22
10	4.000	3	0	1	23
11	2.500	2	0	1	24
12	1.500	1	0	1	25
13	1.000	0	0	1	26

Age 27 Prevalence 53.00

Group	%	Duration current	Duration lagged	Changes	Age started
1	47.000	0	0	0	
2	14.500	12	7	1	15
3	4.000	11	6	1	16
4	4.250	10	5	1	17
5	4.250	9	4	1	18
6	3.000	8	3	1	19
7	3.750	7	2	1	20
8	5.000	6	1	1	21
9	4.250	5	0	1	22
10	4.000	4	0	1	23
11	2.500	3	0	1	24
12	1.500	2	0	1	25
13	1.000	1	0	1	26
14	1.000	0	0	1	27

Table C4 (cont)

Age 28 Prevalence 53.50

Group	%	Duration current	Duration lagged	Changes	Age started
1	46.500	0	0	0	
2	14.500	13	8	1	15
3	4.000	12	7	1	16
4	4.250	11	6	1	17
5	4.250	10	5	1	18
6	3.000	9	4	1	19
7	3.750	8	3	1	20
8	5.000	7	2	1	21
9	4.250	6	1	1	22
10	4.000	5	0	1	23
11	2.500	4	0	1	24
12	1.500	3	0	1	25
13	1.000	2	0	1	26
14	1.000	1	0	1	27
15	0.500	0	0	1	28

Age 29 Prevalence 54.25

Group	%	Duration current	Duration lagged	Changes	Age started
1	45.750	0	0	0	
2	14.500	14	9	1	15
3	4.000	13	8	1	16
4	4.250	12	7	1	17
5	4.250	11	6	1	18
6	3.000	10	5	1	19
7	3.750	9	4	1	20
8	5.000	8	3	1	21
9	4.250	7	2	1	22
10	4.000	6	1	1	23
11	2.500	5	0	1	24
12	1.500	4	0	1	25
13	1.000	3	0	1	26
14	1.000	2	0	1	27
15	0.500	1	0	1	28
16	0.750	0	0	1	29

Age 30 Prevalence 54.75

Group	%	Duration current	Duration lagged	Changes	Age started
1	45.250	0	0	0	
2	14.500	15	10	1	15
3	4.000	14	9	1	16
4	4.250	13	8	1	17
5	4.250	12	7	1	18
6	3.000	11	6	1	19
7	3.750	10	5	1	20
8	5.000	9	4	1	21
9	4.250	8	3	1	22
10	4.000	7	2	1	23
11	2.500	6	1	1	24
12	1.500	5	0	1	25
13	1.000	4	0	1	26
14	1.000	3	0	1	27
15	0.500	2	0	1	28
16	0.750	1	0	1	29
17	0.500	0	0	1	30

Age 31 Prevalence 54.50

Group	%	Duration current	Duration lagged	Changes	Age started	stopped
1	45.250	0	0	0		
2	14.500	16	11	1	15	
3	4.000	15	10	1	16	
4	4.250	14	9	1	17	
5	4.250	13	8	1	18	
6	3.000	12	7	1	19	
7	3.750	11	6	1	20	
8	5.000	10	5	1	21	
9	4.250	9	4	1	22	
10	4.000	8	3	1	23	
11	2.500	7	2	1	24	
12	1.500	6	1	1	25	
13	1.000	5	0	1	26	
14	1.000	4	0	1	27	
15	0.500	3	0	1	28	
16	0.750	2	0	1	29	
17	0.250	1	0	1	30	
18	0.250	-1	0	2	30	31

Table C4 (cont)

Age 32 Prevalence 54.75

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	45.250	0	0	0			
2	14.500	17	12	1	15		
3	4.000	16	11	1	16		
4	4.250	15	10	1	17		
5	4.250	14	9	1	18		
6	3.000	13	8	1	19		
7	3.750	12	7	1	20		
8	5.000	11	6	1	21		
9	4.250	10	5	1	22		
10	4.000	9	4	1	23		
11	2.500	8	3	1	24		
12	1.500	7	2	1	25		
13	1.000	6	1	1	26		
14	1.000	5	0	1	27		
15	0.500	4	0	1	28		
16	0.750	3	0	1	29		
17	0.250	2	0	1	30		
18	0.250	1	0	3	30	31	32

Age 33 Prevalence 55.25

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	18	13	1	15		
3	4.000	17	12	1	16		
4	4.250	16	11	1	17		
5	4.250	15	10	1	18		
6	3.000	14	9	1	19		
7	3.750	13	8	1	20		
8	5.000	12	7	1	21		
9	4.250	11	6	1	22		
10	4.000	10	5	1	23		
11	2.500	9	4	1	24		
12	1.500	8	3	1	25		
13	1.000	7	2	1	26		
14	1.000	6	1	1	27		
15	0.500	5	0	1	28		
16	0.750	4	0	1	29		
17	0.250	3	0	1	30		
18	0.250	2	0	3	30	31	32
19	0.500	0	0	1	33		

Age 34 Prevalence 55.25

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	19	14	1	15		
3	4.000	18	13	1	16		
4	4.250	17	12	1	17		
5	4.250	16	11	1	18		
6	3.000	15	10	1	19		
7	3.750	14	9	1	20		
8	5.000	13	8	1	21		
9	4.250	12	7	1	22		
10	4.000	11	6	1	23		
11	2.500	10	5	1	24		
12	1.500	9	4	1	25		
13	1.000	8	3	1	26		
14	1.000	7	2	1	27		
15	0.500	6	1	1	28		
16	0.750	5	0	1	29		
17	0.250	4	0	1	30		
18	0.250	3	0	3	30	31	32
19	0.500	1	0	1	33		

Table C4 (cont)

Age 35 Prevalence 55.25

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	20	15	1	15		
3	4.000	19	14	1	16		
4	4.250	18	13	1	17		
5	4.250	17	12	1	18		
6	3.000	16	11	1	19		
7	3.750	15	10	1	20		
8	5.000	14	9	1	21		
9	4.250	13	8	1	22		
10	4.000	12	7	1	23		
11	2.500	11	6	1	24		
12	1.500	10	5	1	25		
13	1.000	9	4	1	26		
14	1.000	8	3	1	27		
15	0.500	7	2	1	28		
16	0.750	6	1	1	29		
17	0.250	5	0	1	30		
18	0.250	4	0	3	30	31	32
19	0.500	2	0	1	33		

Age 36 Prevalence 54.75

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	21	16	1	15		
3	4.000	20	15	1	16		
4	4.250	19	14	1	17		
5	4.250	18	13	1	18		
6	3.000	17	12	1	19		
7	3.750	16	11	1	20		
8	5.000	15	10	1	21		
9	4.250	14	9	1	22		
10	4.000	13	8	1	23		
11	2.500	12	7	1	24		
12	1.500	11	6	1	25		
13	1.000	10	5	1	26		
14	1.000	9	4	1	27		
15	0.500	8	3	1	28		
16	0.750	7	2	1	29		
17	0.250	6	1	1	30		
18	0.250	5	-1	3	30	31	32
19	0.500	-3	0	2	33	36	

Age 37 Prevalence 54.75

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	22	17	1	15		
3	4.000	21	16	1	16		
4	4.250	20	15	1	17		
5	4.250	19	14	1	18		
6	3.000	18	13	1	19		
7	3.750	17	12	1	20		
8	5.000	16	11	1	21		
9	4.250	15	10	1	22		
10	4.000	14	9	1	23		
11	2.500	13	8	1	24		
12	1.500	12	7	1	25		
13	1.000	11	6	1	26		
14	1.000	10	5	1	27		
15	0.500	9	4	1	28		
16	0.750	8	3	1	29		
17	0.250	7	2	1	30		
18	0.250	6	1	3	30	31	32
19	0.500	-3	0	2	33	36	

Table C4 (cont)

Age 38 Prevalence 54.75

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started
1	44.750	0	0	0			
2	14.500	23	18	1	15		
3	4.000	22	17	1	16		
4	4.250	21	16	1	17		
5	4.250	20	15	1	18		
6	3.000	19	14	1	19		
7	3.750	18	13	1	20		
8	5.000	17	12	1	21		
9	4.250	16	11	1	22		
10	4.000	15	10	1	23		
11	2.500	14	9	1	24		
12	1.500	13	8	1	25		
13	1.000	12	7	1	26		
14	1.000	11	6	1	27		
15	0.500	10	5	1	28		
16	0.750	9	4	1	29		
17	0.250	8	3	1	30		
18	0.250	7	2	3	30	31	32
19	0.500	-3	0	2	33	36	

Age 39 Prevalence 54.50

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started	stopped
1	44.750	0	0	0				
2	14.500	24	19	1	15			
3	4.000	23	18	1	16			
4	4.250	22	17	1	17			
5	4.250	21	16	1	18			
6	3.000	20	15	1	19			
7	3.750	19	14	1	20			
8	5.000	18	13	1	21			
9	4.250	17	12	1	22			
10	4.000	16	11	1	23			
11	2.500	15	10	1	24			
12	1.500	14	9	1	25			
13	1.000	13	8	1	26			
14	1.000	12	7	1	27			
15	0.500	11	6	1	28			
16	0.750	10	5	1	29			
17	0.250	9	4	1	30			
18	0.250	-8	3	4	30	31	32	39
19	0.500	-3	1	2	33	36		

Age 40 Prevalence 54.25

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started	stopped
1	44.750	0	0	0				
2	14.500	25	20	1	15			
3	4.000	24	19	1	16			
4	4.250	23	18	1	17			
5	4.250	22	17	1	18			
6	3.000	21	16	1	19			
7	3.750	20	15	1	20			
8	5.000	19	14	1	21			
9	4.250	18	13	1	22			
10	4.000	17	12	1	23			
11	2.500	16	11	1	24			
12	1.500	15	10	1	25			
13	1.000	14	9	1	26			
14	1.000	13	8	1	27			
15	0.500	12	7	1	28			
16	0.750	11	6	1	29			
17	0.250	-10	5	2	30	40		
18	0.250	-8	4	4	30	31	32	39
19	0.500	-3	2	2	33	36		

Table C4 (cont)

Age 70 Prevalence 28.50

Group	%	Duration current	Duration lagged	Changes	Age started	stopped	started	stopped
1	44.750	0	0	0				
2	14.500	55	50	1	15			
3	4.000	54	49	1	16			
4	4.250	53	48	1	17			
5	4.250	52	47	1	18			
6	1.500	51	46	1	19			
7	1.500	-49	45	2	20	69		
8	1.750	-45	44	2	21	66		
9	0.250	-42	-42	2	22	64		
10	1.000	-36	-36	2	23	59		
11	1.000	-31	-31	2	24	55		
12	0.250	-26	-26	2	25	51		
13	0.500	-23	-23	2	26	49		
14	0.500	-19	-19	2	27	46		
15	0.500	-17	-17	2	28	45		
16	0.250	-16	-16	2	29	45		
17	0.250	-11	-11	4	30	40	42	43
18	0.250	-8	-8	4	30	31	32	39
19	0.500	-3	-3	2	33	36		
20	0.250	-14	-14	2	29	43		
21	0.250	-15	-15	2	29	44		
22	0.500	-18	-18	2	27	45		
23	0.500	-22	-22	2	26	48		
24	1.250	-25	-25	2	25	50		
25	0.250	-27	-27	2	24	51		
26	0.250	-29	-29	2	24	53		
27	1.000	-30	-30	2	24	54		
28	0.500	-32	-32	2	23	55		
29	0.500	-33	-33	4	23	55	56	57
30	1.750	-34	-34	2	23	57		
31	0.250	-35	-35	2	23	58		
32	2.250	-38	-38	2	22	60		
33	0.500	-39	-39	2	22	61		
34	0.500	-40	-40	2	22	62		
35	0.750	-41	-41	2	22	63		
36	1.250	-43	-43	2	21	64		
37	2.000	-44	-44	2	21	65		
38	0.250	-46	45	2	20	66		
39	1.250	-47	45	2	20	67		
40	0.750	-48	45	2	20	68		
41	0.250	-50	46	2	19	69		
42	1.250	-51	46	2	19	70		

Appendix D

Mathematical models for the relationship of smoking to lung cancer

I. The multistage model

Authors: P.N. Lee and Mrs B.A. Forey

Date: June 1994

Index

Glossary of abbreviations

a_i	transition probabilities during first period considered for stage i
B	constant relating incidence to a power of time
b_i	transition probabilities during second period considered for stage i
C	proportion of susceptible
c	power of dose relationship
c_i	transition probabilities during third period considered for stage i
D	duration of exposure
d	dose of carcinogen
F	length of period after stopping exposure
G_T	cumulative density function at time T
g_1	Whittemore's packs function
g_2	Whittemore's multistage function
I_T	incidence rate at time T
k	number of stages of the multistage process
N	number of cells at risk
P_i	transition probability for stage i
R	ratio of incidences of smoker and nonsmoker
S	age of starting to smoke
S_i	time at which ith period of exposure ends
T	time
$\underline{T}$	median time of tumour induction
u	transition probability for affected stage during first period considered
v	transition probability for affected stage during second period considered
W	waiting time between last transition and appearance of cancer
α	background transition probabilities for stage i
β	increase in transition probability for stage i per unit dose of carcinogen

1. INTRODUCTION

1.1 Value of models

A number of mathematical models have been used to attempt to quantify the relationship between lung cancer and various aspects of the smoking habit, such as age of starting to smoke, amount smoked, duration of smoking, and, in ex-smokers, time since stopping. Use of an appropriate model may allow prediction of future lung cancer rates and judgement as to the extent to which trends over time or differences between countries in incidence of lung cancer are explicable in terms of smoking habits or depend on other lung cancer risk factors. Ideally, a good model should not only describe well how incidence depends on smoking, but should have some biological meaning, giving insight into the mechanisms by which cancer develops. Even a good model will, however, only be an approximation to the truth and cannot be expected to take into account precisely the interplay of susceptibility, exposure and disease.

1.2 Power law relationship of mortality rates with age and the multistage model

Early interest in mathematical models for cancer started shortly after the second World War with the observation (e.g. Fisher and Holloman, 1951; Nordling, 1953) that, for many types of cancer, mortality rates rose with age according to an approximate power law, with the exponent often about 6. There are a number of difficulties in interpreting published mortality rates, described in section 1.3 below. Despite these difficulties, and despite it being apparent

that the simple power law relationship did not fit for all types of cancer (as later confirmed in a detailed analysis of 338 data sets by Cook, Doll and Fellingham (1969)), a number of models have been postulated in an attempt to try to explain this relationship. The most important of these has been the multistage model of Armitage and Doll (1954), which predicts a power law when exposure is constant and continuous, and a more complex relationship when it is not. The multistage model is discussed in detail in this document, which not only gives its derivation, but also describes how well it explains a variety of aspects of the smoking/lung cancer relationship. Other models will be considered in a separate document.

1.3 Difficulties in interpreting published mortality rates

The major difficulties in interpreting published mortality rates can be summarized as follows:

(a) For some cancers, though not for lung cancer, which usually is rapidly fatal, mortality rates may not bear a close correspondence to incidence rates;

(b) Recorded mortality rates, based on death certificates, usually carried out in the absence of a post-mortem, will be inaccurate due to errors in diagnosis. For lung cancer, the techniques for diagnosing lung cancer have enormously improved between 1900 and 1950 due to the introduction of X-rays, bronchoscopy, intrathoracic surgery, sputum cytology, sulfa drugs and antibiotics (Doll and Peto, 1981), though even now the rate of

false-positive and false-negative diagnosis remains quite high (e.g. Szende et al, 1994), particularly at ages 80 or over (Doll, 1971).

(c) Mortality rates, and indeed incidence rates from cancer registries, do not distinguish between the different histological types of lung cancer, such as squamous cell cancer and adenocarcinoma, which may show different relationships with age, smoking habits and other factors.

(d) Experimental studies are often conducted on genetically similar animals and exposure to the agent of interest is carefully controlled. Human populations, however, vary widely both in susceptibility and exposure. The observed patterns of incidence may be very different for different subsets of the population.

(e) Studying variation in rates by age for one particular year inevitably means one is comparing different birth cohorts at each age, with differing patterns of smoking habits and exposure to other risk factors. The study of variation in rates by age for one particular birth cohort, on the other hand, means comparison over a long time period during which inter alia diagnostic standards may have changed.

(f) Because of competing risk of death from other diseases, people surviving to older ages may be unrepresentative, in respect of susceptibility and exposure, of the whole population from which they are derived. (Indeed, even in the absence of deaths from other causes, the surviving population may be unrepresentative, especially for genetic diseases, such as

familial polyposis coli and Huntingdon's chorea, where risk rises with age and then falls off, to zero, as the susceptible pool is eliminated.)

(g) There may be inadequate available comparable data on variation by age, sex and year in smoking habits. Data on cigarette consumption per head drawn from sales statistics are usually not age or sex specific; averages may be more appropriate to age groups 20 or 30 years younger than the ages at which lung cancer normally occurs.

(h) Published mortality rates typically do not take account of the effect of variations in exposure to other risk factors for lung cancer, such as occupational exposure, air pollution and diet.

2. DERIVATION AND ASSUMPTIONS

2.1 Assumptions

The multistage model involves the following assumptions:

(i) A person has a large and constant number of cells at risk, N;

(ii) All the cells start in an identical state at age zero;

(iii) A single cell can generate a malignant tumour only after it has undergone a certain number, k, of heritable changes.

Suppose that, when a cell (or its lineal descendants) has experienced exactly k-1 changes, the "transition" probability of occurrence of the kth change, in that line of descent, is p_k per unit time. Then the probability that the kth change occurs in the short time interval (t, t + dt) is approximately,

$$\frac{p_1 p_2 \cdot \cdot \cdot p_k t^{k-1}\, dt}{(k-1)!} \qquad (1)$$

as $t \to 0$. This result will be valid for large values of t (of the order of a human lifetime) provided that $p_1 t$, $p_2 t$, ... $p_k t$ are all sufficiently small. The incidence rate per person is obtained by multiplying (1) by N. For a rigorous proof, see Armitage (1953); for a less rigorous proof, see Armitage and Doll (1954).

2.2 Exposure constant throughout life

Providing that the transition probabilities remain constant throughout life, the incidence rate, I_T, of cancer at time T will be given by the simple formula

$$I_T = BT^{k-1} \qquad (2)$$

where B is a constant equal to $Np_1 p_2 \ldots p_k / (k-1)!$

This is the simple power law relationship observed by Fisher and Holloman (1951) and by Nordling (1953). The incidence rate is that for a Weibull distribution, where the cumulative density function, G_T, is given by

$$G_T = 1 - \exp(-BT^k) \qquad (3/1)$$

As noted by Pike (1966), this distribution may actually arise under quite broad assumptions concerning the distribution of time to onset of cancer in individual cells (i.e. the model implies the formula; but the formula does not imply the model). The Weibull distribution

is in fact also known as the "third asymptotic distribution of smallest values" discovered by Fréchet (1927) and by Fisher and Tippett (1928) (see Gumbel (1958) for a discussion of the derivation of the three distributions and of their properties). This distribution is often expressed with an extra parameter W as

$$G_T = 1 - \exp(-B(T-W)^k) \qquad (3/2)$$

In the context of the multistage model, W is often interpreted as the "waiting time" between the last transition occurring and clinical appearance of, or death from, lung cancer. To simplify the presentation that follows we ignore W, though note that some researchers, when fitting the multistage model, ignore exposure up to a short period (eg. 2 years) before recorded diagnosis or death to try to take account of this waiting time.

2.3 Exposure varying during life

In the simplest use of the multistage model, the transition probabilities are assumed to remain constant throughout life. A strength of the model is that incidence can readily be calculated for varying probabilities, e.g. resulting from varying exposure. Again assuming transition probabilities are small, and, for convenience, taking k=5, the incidence rate at time T is given by the formula

$$I_T = p_5 \int_0^T p_4 \int_0^{t_4} p_3 \int_0^{t_3} p_2 \int_0^{t_2} p_1 dt_1 dt_2 dt_3 dt_4 dt_5 \qquad (4)$$

where the p_i are the time-dependent transition probabilities for each stage.

Although it is in theory possible to take into account any form of functional dependence of the transition probabilities on age, the most common uses of the multistage model have been where transition probabilities are either unaffected by exposure, and take "background" values α_i which are invariant of age, or are affected by exposure, taking the constant value $\alpha_1 + \beta_i d = \gamma_i$ when exposure occurs, d being dose of carcinogen applied. In the simpler applications, dose is constant during exposure. In some contexts, $\beta_i d$ may be large with respect of α_i, so that the transition probability is approximately directly proportional to dose.

2.4 Two relevant periods - continuous smokers

One particularly useful form of the incidence rate formula applies where there are two periods of time, during the first of which [0,S] the transition probabilities are a_i and during the second of which [S,T] the transition probabilities are b_i. In the context of smoking, S can be viewed as the age of starting to smoke, smoking continuing subsequently. a_i are background probabilities in the absence of smoking, b_i the probabilities during smoking. Up to time S, the incidence rate is as for formula (2). Subsequently, the formula is given by

2 stage process

$$I_T = N\,[a_1 b_2 S + b_1 b_2\,(T-S)] \qquad (5/2)$$

3 stage process

$$I_T = N\left[\frac{a_1a_2b_3S^2}{2} + a_1b_2b_3S(T-S) + \frac{b_1b_2b_3(T-S)^2}{2}\right] \qquad (5/3)$$

4 stage process

$$I_T = N\left[\frac{a_1a_2a_3b_4S^3}{6} + \frac{a_1a_2b_3b_4S^2(T-S)}{2} + \ldots\right.$$

$$\left.\ldots + \frac{a_1b_2b_3b_4S(T-S)^2}{2} + \frac{b_1b_2b_3b_4(T-S)^3}{6}\right] \qquad (5/4)$$

5 stage process

$$I_T = N\left[\frac{a_1a_2a_3a_4b_5S^4}{24} + \frac{a_1a_2a_3b_4b_5S^3(T-S)}{6} + \ldots\right.$$

$$\left.\ldots + \frac{a_1a_2b_3b_4b_5S^2(T-S)^2}{4} + \frac{a_1b_2b_3b_4b_5S(T-S)^3}{6} + \frac{b_1b_2b_3b_4b_5(T-S)^4}{24}\right] \qquad (5/5)$$

More generally, for a k stage process, the formula can be derived noting that the terms within the square bracket arise from a binomial expansion of $[S + (T-S)]^{k-1} / (k-1)!$ with each term being multiplied by appropriate values of a_i or b_i, the first term relating to cancers where the first k-1 transitions occur before S, the second term to cancers where the first k-2 transitions occur before S, and so on (the last transition must occur after S, at time T, by definition).

Note that these formulae can be considerably simplified when only one, or a limited number of stages, are affected by exposure. As an example consider the four stage process where only the first stage is affected. If a_i are the background transition probabilities

for unaffected stages, u is the transition probability for the affected stage during the period [0,S] and v the transition probability for the affected stage during the period [S,T], we have

$$I_T = \frac{Na_2a_3a_4}{6} [uS^3 + 3uS^2(T-S) + 3uS(T-S)^2 + v(T-S)^3]$$

$$\simeq uT^3 + (v-u)(T-S)^3$$

More generally, for a k stage process with the <u>first stage</u> affected

$$I_T \simeq uT^{k-1} + (v-u)(T-S)^{k-1} \qquad (6/1)$$

With the <u>penultimate stage</u> affected, we have

$$I_T \simeq (u-v)S^{k-1} + vT^{k-1} \qquad (6/2)$$

With the <u>first and penultimate stages</u> affected, we have

$$I_T \simeq u_1u_2S^{k-1} + v_1v_2\ (T-S)^{k-1}$$

$$+ u_1v_2\ (T^{k-1}-S^{k-1}-(T-S)^{k-1}) \qquad (6/3)$$

(Here u_1 and v_1 refer to the first stage transition probabilities, and u_2 and v_2 refer to the penultimate stage transition probabilities.)

As discussed elsewhere, e.g. by Day and Brown (1980), Brown and Chu (1983b) and Brown and Chu (1987), these formulae allow some fairly simple conclusions. Let us consider firstly <u>excess</u> incidence at age T in relation to exposure starting at time S. Where only the <u>first stage</u> is affected, since the incidence at age T in the

absence of carcinogenic exposure would be uT^{k-1}, since the duration of exposure, D, equals (T-S) and since v-u is linearly proportional to dose d, we have (from formula 6/1)

$$I_T \simeq dD^{k-1} \qquad (7/1)$$

i.e. the excess risk at a given age is proportional to dose, depends (by a power-law relationship) on duration of exposure, but is independent of age of starting to smoke. Where the <u>penultimate stage</u> is affected we have (from formula 6/2)

$$I_T \simeq d[(D+S)^{k-1} - S^{k-1}] \qquad (7/2)$$

i.e. the excess risk is proportional to the dose d and is an increasing function of both duration given age of start, and of age of start given duration. Where the <u>first and penultimate stages</u> are affected, the excess risk can be expressed by the formula

$$I_T \simeq d_1D^{k-1} + d_2[(D+S)^{k-1} - S^{k-1}] + d_1d_2D^{k-1} \qquad (7/3)$$

Here d_1 and d_2 are the effective excess doses, relative to background, for the first and penultimate stages (i.e. if the dose increases the background risk by a factor q, the effective dose is q-1). Note that setting $d_2 = 0$ gives formula (7/1) and setting $d_1 = 0$ gives formula (7/2).

2.5 Three relevant periods - giving up smoking

The same authors note that inferences can similarly be made by examining the excess risk patterns for those individuals who have stopped their exposure. When the exposure starts at age S, continues

for a duration D, then stops, and follow-up continues for a period of length F, the excess risk at age S+D+F = T is given by

$$I_T \simeq d[(D+F)^{k-1} - F^{k-1}] \qquad (8/1)$$

when only the <u>first stage</u> is affected by the carcinogen, by

$$I_T \simeq d[(D+S)^{k-1} - S^{k-1}] \qquad (8/2)$$

where only the <u>penultimate stage</u> is affected, and by

$$I_T \simeq d_1[(D+F)^{k-1} - F^{k-1}] + d_2[(D+S)^{k-1} - S^{k-1}] + d_1 d_2 D^{k-1} \qquad (8/3)$$

where both the <u>first and penultimate stages</u> are affected. Note that Whittemore (1988) gives a version of this formula (her formula 12 using different notation) which is incorrect, including a term $d_1 D^{k-1}$ rather than the correct term $d_1[(D+F)^{k-1} - F^{k-1}]$. These terms are the same where exposure is not discontinued (F = 0) but not otherwise.

These inferences for stopping smoking can be derived from formulae (analogous to formulae 5) in which there are three periods of time, during the first of which $[0,S_1]$ the transition probabilities are a_i, during the second of which $[S_1,S_2]$ the transition probabilities are b_i, and during the third of which $[S_2,T]$ the transition probabilities are c_i. Below we give the formulae for a <u>4 stage process</u>.

$$I_T = N \left[\frac{a_1a_2a_3c_4S_1^3}{6} + \frac{a_1a_2b_3c_4S_1^2(S_2-S_1)}{2} + \ldots \right.$$

$$+ \frac{a_1a_2c_3c_4S_1^2(T-S_2)}{2} + \frac{a_1b_2b_3c_4S_1(S_2-S_1)^2}{2} + \ldots$$

$$+ \frac{a_1b_2c_3c_4S_1(S_2-S_1)(T-S_2)}{1} + \frac{a_1c_2c_3c_4S_1(T-S_2)^2}{2} + \ldots$$

$$+ \frac{b_1b_2b_3c_4(S_2-S_1)^3}{6} + \frac{b_1b_2c_3c_4(S_2-S_1)^2(T-S_2)}{2} + \ldots$$

$$\left. + \frac{b_1c_2c_3c_4(S_2-S_1)(T-S_2)^2}{2} + \frac{c_1c_2c_3c_4(T-S_2)^3}{6} \right] \qquad (9)$$

More generally, for a k stage process, the formula can be derived noting that the terms within the square brackets arise from a multinomial expansion of $[S_1+(S_2-S_1)+(T-S_2)]^{k-1}/(k-1)!$ with each term being multiplied by appropriate values of a_i, b_i or c_i, to describe the various sequences by which cancer can arise. For example the 5th term above describes the cases where the first transition occurs in $[0,S_1]$, with contribution a_1S_1 to the formula (probability x length of period), the second transition occurs in $[S_1,S_2]$, with contribution $b_2(S_2-S_1)$, and the third occurs in $[S_2, T]$, with contribution $c_3(T-S_2)$, the fourth occurring at T, with contribution c_4. Where multiple (z) transitions occur in one period, e.g. in the first term the first three changes occur in $[0,S_1]$, the denominator includes a term z! to take account of the fact that only one of the possible sequences of transition is allowed (the transitions must be in order).

Formulae 8 can readily be shown to be special cases of formula 9.

2.6 More than three relevant periods

It may also be useful to write down the formula for the situation where there are two periods of identical exposure, a person having periods of length U, V, W, X, Y respectively unexposed, exposed, unexposed, exposed and unexposed, i.e. the person starts smoking and gives up twice. Where both the first and penultimate stages are affected, the excess risk is given by

$$\begin{aligned} I_T \simeq\ & d_1[(V+W+X+Y)^{k-1} - (W+X+Y)^{k-1} + (X+Y)^{k-1} - Y^{k-1}] \\ & + d_2[(U+V+W+X)^{k-1} - (U+V+W)^{k-1} + (U+V)^{k-1} - V^{k-1}] \\ & + d_1 d_2[(V+W+X)^{k-1} + V^{k-1} - W^{k-1} + X^{k-1} - (V+W)^{k-1} - (W+X)^{k-1}] \end{aligned} \tag{10}$$

The simpler formulae when only the first or only the penultimate stages are affected are given by setting $d_2 = 0$ or $d_1 = 0$, respectively, in the above formula.

This formula can be extended to larger numbers of exposure periods by realizing that:

(a) the term in d_1 (the first stage effect) is the sum of (k-1)th powers of the length of all periods starting at the beginning of an exposure period and ending at t, minus the sum of (k-1)th powers of the length of all periods starting at the end of an exposure period and ending at t;

(b) the term in d_2 (the penultimate stage effect) is the sum of (k-1)th powers of the length of all periods starting at time 0

and ending at the end of an exposure period, minus the sum of (k-1)th powers of the length of all periods starting at time 0 and ending at the beginning of an exposure period;

(c) the term in $d_1 d_2$ (the joint effect) is the sum of (k-1)th powers of the length of all periods starting at the beginning of an exposure period and ending at the end of an exposure period, minus the sum of (k-1)th powers of the length of all periods which either start at the beginning of one exposure period and end at the beginning of another or start at the end of one exposure period and end at the end of another.

3. PREDICTIONS OF THE MULTISTAGE MODEL AND CONFORMITY WITH OBSERVATIONS

The multistage model makes a number of predictions as to how the cancer incidence rate will depend on various aspects of the data. These are considered in some detail, comparing the predictions as appropriate with epidemiological and animal data. Before looking at these various aspects in turn, we first summarize some of the key data sources we will use as reference for comparison.

3.1 Data sources

British Doctors Study. In 1951 Doll and Hill sent a questionnaire on smoking habits to all men and women on the British Medical Register. The 34,000 men and 6,000 women who replied have been followed up for mortality ever since. Results of 20 year follow-up for men are given in Doll and Peto (1976) and of 22 year follow-up for women are given in Doll *et al* (1980). Doll and Peto (1978) give a detailed

tabulation of lung cancers and man-years at risk by age and amount smoked for men who had never smoked and for men who started smoking at ages 16-25 and continued to smoke.

US Veterans' Study. In 1954 Dorn mailed questionnaires to US veterans, mainly of World War I, who held Government life insurance policies. Almost all policy holders were white males. Almost 250,000 responses were received. Kahn (1966) gives extensive tables or results relating to follow up after 8½ years. Rogot (1974) gives less detailed results for 16 years follow-up.

American Cancer Society (ACS) Cancer Prevention Studies I and II (CPS I and II). The ACS have sponsored two huge prospective studies of smoking and mortality in the United States. In the first study about 1 million persons were followed from 1959 until 1972, in the second study about 1.2 million persons were followed from 1982 until 1988. There have been a very large number of papers published about CPS I. In particular Hammond (1966) gave very detailed results for four years follow-up, and various reports of the US Surgeon-General (particularly 1979, 1982 and 1989) have presented summary results. The 1989 report has also presented some results for CPS II, though extensive tables have yet to be published. It should be noted that the sampling in both studies was by ACS volunteers and those interviewed are not representative of the US population. In particular they are far more likely than average to be white, have higher education and income and lower exposure to occupational carcinogens and lower mortality than average.

Studies of skin painting of mice. During the 1960's and early 1970's a large number of studies were carried out in which the backs of mice were painted regularly with tobacco smoke condensate or with known carcinogens as a model for human carcinogenesis. Studies were carried out by the Tobacco Research Council at Harrogate, by the Medical Research Council at Pollard's Wood and by other laboratories. Relevant papers include Lee (1974), Lee and O'Neill (1971), Lee, Rothwell and Whitehead (1977) and Peto *et al* (1975).

3.2 Relationships with age, duration and age of starting to smoke

As shown by formula 2, the multistage model predicts that if the transition probabilities remain constant throughout life the incidence rate of cancer will bear a simple power law relationship to *age*. Where the first stage is very strongly affected then, regardless of which other stages are affected, the incidence rate will have a simple power law relationship to *duration of exposure*. For example, take formula 6/3 and let u_1 tend to zero. However, where the first stage is not affected, one may get a more complex relationship (see formula 7/3).

As noted above, the multistage model was actually derived to explain the fact that, for many cancers, incidence (or mortality) rates tend to rise approximately according to a power of age (Fisher and Holloman, 1951; Nordling, 1953), although the relationship shows upward or downward curvature from this general pattern in many cases

(Cook, Doll and Fellingham, 1969), even if one excludes from analysis incidence rates observed at high age, where diagnosis is unreliable.

A particularly important study was that on mouse skin reported by Peto et al (1975). In this study a total of 950 mice with a normal lifespan of two to three years were exposed to regular application of benzpyrene (a proven carcinogen) starting at 10, 25, 40 or 55 weeks of age. In each group the incidence rate of malignant epithelial skin tumours among the survivors increased similarly according to a power of duration of exposure. Given duration of exposure, incidence was shown to be completely independent of age. These results suggested that observed approximate power-law increases in most human adult cancer incidence rates with age could exist merely because age equals duration of exposure to background and carcinogenic stimuli. The results could be explained without postulating any intrinsic effects of ageing (such as failing immunological surveillance or age related hormonal changes), and are consistent with our multistage hypotheses in which benzpyrene strongly affected the first stage (and perhaps also other stages) of a multistage process, with background transition probabilities invariant of age.

Another interesting observation consistent with the notion that age per se need not be relevant to risk of cancer occurrence is that reported by Lijinsky (1993). Collecting evidence from studies in 20 species of mammals, reptiles, birds, amphibians and fish exposed to

approximately 1000 mg/kg body weight lifetime dose of nitrosodiethylamine, he noted that, despite the great variation in lifespan (from 3 years in mice to over 50 years in snakes), tumours developed within a similar period, of about a year. He felt that "the evidence suggests that the time dependence of tumour development is more likely related to the cumulative dose of carcinogen than to lifespan and the rate of aging".

The results of a study by Stenbäck et al (1981), in which mouse skin tumours were induced by a single initiating dose of DMBA followed three weeks later by application of the tumour promoter TPA, do not fit in so well with the simple multistage theory. They reported a highly significantly lower yield of tumours when initiation took place at 68 weeks of age than when it took place at 8 or at 48 weeks of age. The authors suggested that this difference was chiefly due not to changes in the number of cells initiated by DMBA but rather to a decrease in the promotional efficacy of TPA in ageing mice.

Peto et al (1985) consider these and additional animal experiments, concluding that the observations "argue strongly that there is no systematic tendency for old animals to be more susceptible to the processes of carcinogenesis than younger animals are", a conclusion reflected in the provocative title of their paper, "There is no such thing as ageing, and cancer is not related to it".

Turning now to humans, Seidman (1985) and Peto et al (1982), have analysed data relating incidence of mesothelioma in asbestos workers to age, age at start of exposure and duration of exposure. Just as in the Peto et al (1975) benzpyrene mouse study, they found that, given duration of exposure, age at start of exposure was irrelevant. Peto et al (1982) concluded that their results support the multistage model of carcinogenesis "under which the increase in most cancer incidence rates with age is due to a constant incidence of genetic or epigenetic accidents, rather than to progressive generalized changes in regulatory or immune function".

Given duration of exposure, age at start of exposure is associated with risk of some cancers. One case in point is lung cancer due to arsenic exposure. Brown and Chu (1983a,b) compared risk of lung cancer in groups of copper smelter workers exposed to arsenic and found that risk increased steadily as age at start of exposure increased from <20, through 20-29 and 30-35, up to 40-49 years. However this does not of itself mean that their results are inconsistent with the multistage hypothesis, rather that one needs to assume that arsenic affects a late stage of the process in order to explain the results. In fact, Brown and Chu fitted the actual functional form of the excess cancer risk predicted by the multistage theory to their detailed data on risk of lung cancer by level of exposure, age at initial employment and duration of employment and found an excellent fit to formula 7/2, in which the penultimate stage of a four stage process is affected. This formula

fitted the data considerably better than formula 7/1, in which the first stage is affected and the authors concluded that "the results indicate that arsenic exerts a definite late stage effect though an additional effect at the initial stage cannot be ruled out".

Doll (1971), using data from his British Doctors Study, plotted, on a double logarithmic scale, lung cancer incidence rates in man

(a) for nonsmokers, against age,

(b) for smokers, against age, and

(c) for smokers, against duration of smoking.

Since the amount smoked varied with age, the incidence rates in smokers were standardized for smoking habits. Equations (a) and (b) both showed a good linear relationship (consistent with formula 2) but the slopes of the lines varied markedly, with k estimated as 5 for nonsmokers and about 8.5 for cigarette smokers. However, when plot (c) was considered, the position was changed. In this case the relationship remained linear, but the value of k for smokers became much lower and very similar to that for nonsmokers. The graphical results presented by Doll were consistent with lung cancer resulting from a 5 stage process, with risk related to duration of exposure. In nonsmokers exposure is from birth to a weak carcinogen; in smokers exposure is from start of smoking to a stronger carcinogen. Note that, in theory (see formula 7/1), <u>excess</u>, not <u>absolute</u>, risk in smokers should be proportional to a power of duration of

exposure. However, since risk in smokers is so much higher than in nonsmokers (relative risk of about 14 in the British Doctors Study), excess and absolute risk are very similar.

While many studies other than the British Doctors Study allow one to study how risk rises with age in smokers and nonsmokers, relatively few studies provide useful data on how risk varies by age of starting to smoke given duration of exposure. A problem of course is that most smokers tend to start smoking within a relatively short period of time and it is difficult to accumulate sufficient data on people starting very early or very late to allow reliable comparison. Perhaps the best data, reproduced in Table 1, comes from the Veterans' Study (Kahn, 1966). If one looks at the data for all cigarette smokers a striking fact emerges, namely that increasing age by 10 years has a virtually identical effect to decreasing age of starting to smoke by 10 years. Thus comparing two groups of smokers, both with a duration of about 43 years, one aged 55-64 and starting to smoke at age 15-19, the other aged 65-74 and starting to smoke at age 25+, we see their lung cancer rates (168 and 162 per 10^5 per year) are virtually identical. Similarly comparing two groups of smokers, both with a duration of about 48 years, one aged 55-64 and starting to smoke at age <15, the other aged 65-74 and starting to smoke at age 20-24, we again see lung cancer rates (251 and 241 per 10^5 per year) that are very similar. At first sight these results are consistent with the Peto et al (1975) mouse skin results showing irrelevance of age given duration of smoking. However, if one looks at the results in Table 1 broken

down further by amount smoked, the pattern is not so clear cut. Where adequate numbers of deaths are available (in the 10-20 and 21-39 cigs/day group) there is a consistent tendency for risk to be somewhat higher in the older smokers in the above comparisons. The simple comparison for all cigarette smokers appears to be somewhat biassed because it fails to take into account the fact that people who start to smoke younger smoke rather more cigarettes a day than those who start to smoke older. However the inference that age is important given duration is not totally secure, bearing in mind the uncertainty present in what the mean durations in the various groups are, given the relatively wide and in some cases open-ended intervals. Thus, for example, if the average age of starting in the <15 group is say 13.5 and that in the 20-24 group is say 21.5, one may not be comparing groups with identical durations (when one compares 55-64 year olds and 65-74 year olds) but groups which differ in duration by two years.

Another study that has provided relevant data is that by Lubin et al (1984). As described in more detail below (section 5.4), Brown and Chu (1987) found that a multistage model in which the first and penultimate stages were affected by smoking predicted reasonably well the variation observed in risk of lung cancer by age of starting to smoke, given age.

Hegmann et al (1993) have also presented data consistent with a major effect of age of starting to smoke. Based on a case-control study in Utah involving 282 lung cancer cases and 3282 population

controls they found that, after adjusting for age and amount smoked, men who started to smoke before age 20 had a substantially higher risk of lung cancer (RR compared to nonsmokers = 12.7, 95% CI 6.39-25.2) than men who started later (6.03, 2.82-12.9). For women the heavy increase in risk continued until age 25 (9.97, 4.68-21.2) compared with women who began smoking at age 26 or older (2.58, 0.53-12.4). No analyses were presented comparing risk in smokers of the same duration but of differing ages.

Perhaps the safest conclusions to draw are those given in the IARC (1986) monograph on tobacco smoking. They note that "the effects of the duration of smoking are so strong, and so closely correlated with age, that it is virtually impossible to determine exactly whether ageing _per se_ has any independent effect on excess lung cancer rates among people of different ages who have all smoked similarly for a similar number of years. If age has any independent effect, however, this would be small compared with the accumulative effect of duration of smoking (Peto _et al_, 1975, 1985; see also Likhachev _et al_, 1985)".

The data in Table 1 can be used not only to demonstrate that risk depends much more strongly on duration of smoking than on age given duration, but also to demonstrate an approximate power law relationship between duration and risk. Table 2 shows the result of fitting a fourth power relationship of duration to lung cancer risk. It can be seen that the fit is very adequate.

3.3 Relationships with dose

Given continuous exposure to a dose of a carcinogen, then under the multistage assumptions it has already been shown that the risk of lung cancer at a given age is proportional to the product of the individual transition probabilities. For a stage affected by the carcinogen one might assume that the transition probability, p_i, is linearly related to dose d by the formula

$$p_i = \alpha_i + \beta_i d \tag{11}$$

Here α_i is the background value of the transition probability, and β_i is the coefficient of the regression of the transition probability on dose. Where the carcinogen strongly affects risk, so that $\beta_i d \gg \alpha_i$ one would then get the approximate relationship

$$p_i = \beta_i d \tag{12}$$

i.e. a direct linear relationship of transition probability with dose. Where the particular stage is unaffected by the carcinogen, one would have $\beta_i = 0$ so that

$$p_i = \alpha_i \text{ (constant)} \tag{13}$$

Based on this formulation one would expect the following relationship between incidence rate and the number of stages affected:

(i) <u>One stage strongly affected</u>. Risk proportional to dose, linear through the origin.

(ii) <u>One stage weakly affected</u>. Risk proportional to dose, linear

not through the origin.

(iii) Two stages strongly affected. Risk directly proportional to dose squared.

(iv) Two stages affected, one or both weakly. Quadratic relationship of risk to dose.

(v) C stages strongly affected. Risk directly proportional to dose to the power c.

(vi) C stages affected, some weakly. Cth power polynomial relationship of risk to dose.

A striking example of data fitting the multistage hypothesis both in respect of dose and time comes from the mouse skin painting studies of Lee and O'Neill (1971). In two separate experiments benzopyrene was painted regularly on the backs of mice at different dose levels (6, 12, 24 and 48 μg per week in the Harrogate study; 1, 3, 9 and 27 μg per week in the Zurich study). In both studies the incidence, both of tumours and of infiltrating carcinomas, was very well fitted by the expression

$$I_T = d^2(T-W)^k \qquad (14)$$

where T is time from first application, d is the applied dose, and W and k are constants independent of dose. The direct quadratic relationship of incidence with dose was consistent with benzopyrene strongly affecting two stages of mouse skin carcinogenesis.

There are a number of reasons (some applicable to humans only, some to animals also) why one might not always expect to see such a simple relationship of incidence to dose. These include:

(i) Numbers of cigarettes smoked per day may not be a direct index of exposure to target tissues of relevant smoke constituents, e.g. smokers of differing numbers of cigarettes a day inhale differently;

(ii) Numbers of cigarettes smoked per day may be inaccurately reported; low numbers may be understatements, high numbers exaggerations. There are no data relating lung cancer risk to objective markers of smoke uptake. (Even if there were, current markers, such as cotinine, only quantify recent exposure to one constituent of smoke.)

(iii) Numbers of cigarettes smoked per day may depend on susceptibility to disease. Sufferers of symptoms may cut down; those with strong constitutions may stay smoking high numbers.

(iv) Smokers of different numbers of cigarettes may differ in respect of various other characteristics - age, age of starting to smoke, diet, occupation, etc, etc.

(v) At high doses cells may be killed off before they get the chance to be transformed into cancerous cells. It is generally believed (Major and Mole, 1978) that cell killing by radiation is an explanation for the fact that the risk of induced leukaemia flattens off and then falls above a given dose, and Davies et al (1974) suggest it may explain why in mouse skin painting studies with various cigarette smoke condensates the log incidence/log dose relationship becomes less steep at high

doses.

(vi) It may not be correct that the transition probability for a given stage is actually directly proportional to dose.

Despite these reasons, dose-response relationships consistent with the multistage formulation are found to fit many data sets quite well. Druckrey (1967) has summarized the results of extensive animal studies over more than 25 years involving a total of about 10,000 rats treated with a variety of carcinogenic substances. He noted that for all the carcinogens he studied, the relationship between dose d and median time of tumour induction $\underline{T}$ could be summarized by the general formula:

$$d\underline{T}^n = \text{constant} \qquad (15)$$

(N.B. His studies generally involved such high doses of carcinogenic substances that deaths from other causes did not obscure this simple relationship.) As shown in formula 3/1 the distribution of time to tumour in the absence of death from other causes is given by

$$G = 1 - \exp(-BT^k)$$

Substituting $B = d^c$ (where a carcinogen strongly affects c stages) we have

$$G = 1 - \exp(-d^c T^k) \qquad (16)$$

At the median G = 0.5, so we have

$$\exp(-d^c \underline{T}^k) = 0.5 \qquad (17/1)$$

or

$$d^c \underline{T}^k = \log_e 2 \qquad (17/2)$$

$$\text{or} \qquad d\underline{T}^{k/c} = (\log_e 2)^{1/c} = \text{constant} \qquad (17/3)$$

which is exactly of the form that Druckrey (who did not invoke multistage assumptions at all) found to hold in practice.

Though Druckrey's simple formula may only hold for studies such as his with strong carcinogens where essentially all the animals get tumours, and deaths from other causes rarely occur (so that the observed median time is close to the true median time in the absence of deaths from other causes), his results are completely consistent with what is predicted by the multistage model. It is interesting to note that Druckrey always found his n to be greater than 1, i.e. the carcinogen never affected all the stages of the multistage process. Peto (1977) has also pointed out the dose power is invariably less than the time power. As Armitage and Doll (1954) note, this observation is inconsistent with the Fisher and Holloman (1951) model (vide infra) which predicts that the two powers should be the same.

A number of the major prospective studies on smoking and health have presented data relating incidence rate of lung cancer with amount smoked (see e.g. USSG 1982). All the studies show that risk increases with amount smoked. Generally the dose-response seems to be approximately linear. In view of evidence described elsewhere in section 2 that risk of lung cancer in ex-smokers rapidly becomes less than that in continuing smokers (which suggests a late stage is affected), and evidence that risk of lung cancer in continuing smokers of a given age depends strongly on age of starting to smoke

(which suggests an early stage is affected) this linear dose-response seems somewhat surprising. If two stages are affected then surely the dose-response relationship should have a quadratic component?

Doll and Peto (1978) attempted to answer this point, put forward by Armitage (1971) when discussing a paper by Doll (1971). Based on 20-year follow-up data from the British Doctors study, they studied the relationship of annual lung cancer incidence rate to age and number of cigarettes smoked among cigarette smokers of age 40-79 who started to smoke at age 16-25 and who smoked 40 or less per day. They reported an adequate fit to the formula

$$\text{Lung cancer incidence} = 0.273 \times 10^{-12}(\text{cigs/day} + 6)^{2}(\text{age} - 22.5)^{4.5}$$

They noted that the form of the dependence on dose is "subject not only to random error but also to serious systematic biases", biases which they discussed in the paper. They emphasized that "there was certainly some statistically significant ($p<0.01$) upward curvature of the dose-response relationship in the range 0-40 cigarettes/day, which is what might be expected if more than one of the stages (in the multistage genesis of bronchial carcinoma) was strongly affected by smoking". To some extent their conclusions are dependent on the extent to which they were justified in omitting results for smokers of more than 40 cigarettes a day from their analysis, since risk in this group was clearly substantially less than predicted from their

formula. Some of their reasons for omitting this group from analysis (in whom only five lung cancers occurred) have already been discussed.

For a carcinogen continuously applied throughout life, the incidence rate at a given time, t, should, in theory, be proportional to the following function of dose and time

$$I \alpha t^k \prod_{i=1}^{k} (\alpha_i + \beta_{id}) \qquad (18)$$

It should be noted that, as described by e.g. Crump and Howe (1984) it is possible to fit a generalization of this function as follows

$$I \alpha t^k (q_0 + q_1 d + q_2 d^2 + \ldots q_k d^k) \qquad (19)$$

where all the coefficients q_i are ≥ 0. This model, along with related statistical methods, is routinely used by the EPA and other regulatory agencies to assess low dose cancer risks. It is often referred to as the "multistage model". However formula 19 is actually more general than formula 18, since it contains polynomials not contained in it.

In formulae 18 and 19, the relationships of incidence rate to dose and of incidence rate to time are separable functions which multiply together. Strictly this only applies to continuous exposure throughout life. Where exposure starts at a given point in time, the separability no longer applies, as illustrated by formulae 5 and 6.

Lee (1979) considered a version of the multistage model in which it was assumed that lung cancer was a seven stage multistage process, with smoking only affecting the first and sixth stages. Lee presented a table, reproduced as Table 3, in which relative risk at age 70-74 was related to number of cigarettes smoked under two hypotheses: A - equal effects on stages 1 and 6, and B - greater effect on stage 6 than stage 1. Under the column "linear fit" is shown how a straight line going through the dose points 0 and 6 would fit the data. Figure 1 (reproduced from Lee (1979)) shows that hypothesis B produced a dose-response relationship that is quite close to a linear relationship. In this figure one dose unit from Table 2 has arbitrarily (though not unreasonably in view of the knowledge of the magnitude of relative risk for 20 a day smokers) been taken to be five cigarettes a day. Although inspection of Table 2 shows that hypothesis B fits a linear relationship better than does hypothesis A, it is far from clear that hypothesis A is necessarily ruled out. As Doll and Peto (1978) point out (vide supra) there does appear to be some upward curvature of the dose relationship, and as we have already noted, there are a number of reasons why the observed dose-response may be shallower than the true dose-response. Lee (1979) concluded that it would be difficult to infer reliably from existing data whether late stage effects are stronger than early stage effects. In any event, it is clear that apparent approximate linearity of the dose-response relationship

does not exclude the possibility of two stages being affected by the carcinogen, especially when the effects on the transition probabilities, relative to background, may not be very large.

3.4 Relationships with stopping exposure

Formulae 8/1, 8/2 and 8/3 relate incidence rate to age T for individuals starting to smoke at age S and then smoking for a duration of D. Using these formulae a number of authors have shown that the rise in incidence with time following stopping depends dramatically on which stages are assumed to be affected. If the first stage only is affected, then for a considerable time after stopping the risk rises nearly as fast as if exposure had been continued. This is illustrated in the table below, using formula 8/1 with k = 5, S = 20, d = 10 and D = 20.

	Excess lung cancer risk (10^4)	
Age	Continued smoking	Stopped at age 40
40	160	160
50	810	800
60	2560	2400
70	6250	5440
80	12960	10400

The relative lack of effect of giving up smoking here results from the fact that most cancers arising come from cells which have undergone their first transition early in life. Giving up after this first transition has occurred has no effect at all on risk of cancer arising from a cell.

If the penultimate stage only is affected, then the effect of stopping is much more dramatic, excess risk not rising at all after stopping, though absolute risk does rise. This is illustrated in the table below, using formula 8/2 - again with k = 5, S = 20, d = 10 and D = 20.

	Lung cancer risk (10^4)		
Age	Nonsmoker	Stopped at age 40	Excess
40	256	2656	2400
50	625	3025	2400
60	1296	3696	2400
70	2401	4801	2400
80	4096	6496	2400

Compared with the situation where the first stage is affected, where absolute risk after stopping rises from 416 at age 40 to 14496 at age 80 (i.e. by a factor of 34.8), absolute risk only rises by a factor of 2.4 in the situation where only the last stage is affected.

Lee (1979) has investigated how lung cancer risk varies by time since stopping for a multistage model with seven stages where only the first and sixth stages were affected. Taking S = 20 and D = 20 and using various assumed values of the two stage effects all of which predicted the same multiplication in risk (25) at age 60-64 for continuous smoking, he showed that provided that the sixth stage was affected at least as much as the first stage there was

relatively little increase in risk with giving up smoking for at least 10 years after stopping smoking. Some of his results are reproduced below:

Hypothesis	Description	Stage effects 1	Stage effects 6	Risk relative to risk at age 50-54: 50-54	60-64	70-74
1	Only stage 1 affected	275	1	100	544	2039
2	Stage 1 strongly affected, stage 6 more weakly	25	8.05	100	142	272
3	Both stages affected similarly	12.47	12.47	100	123	191
4	Stage 1 affected less than stage 6, but still quite strongly	5	18.52	100	113	147
5	Stage 1 affected weakly, stage 6 strongly	2	23.01	100	109	132
6	Stage 6 only affected	1	25.03	100	108	126

There are certain problems in interpreting epidemiological data on ex-smokers since those who give up may be unrepresentative in various ways of those who continue to smoke. Inter alia, those who give up may:

(a) be less committed smokers, smoking less, inhaling less, smoking lower tar brands and starting to smoke later;

(b) be more health conscious, a decision to give up smoking being linked to reduced levels of other risk factors; or

(c) be more unhealthy, illness precipitating the decision to give up.

Nevertheless study of trends in rates after giving up smoking gives useful insight into the validity of the multistage model and clues as to the stages likely to be affected.

Data from the British Doctors Study in relation to ex-smoking has been presented in various papers. Doll (1971) gives a detailed table giving man-years at risk and numbers of deaths by amount last smoked, age stopped and period since stopping, Doll and Peto (1976) give estimates of mortality relative to that in continuing smokers and in lifelong nonsmokers, while Doll (1978) gives graphs showing how absolute incidence in ex-smokers, by years stopped, compares with that in continuing smokers and in lifelong nonsmokers. Doll (1978) summarizes the data as follows:

> "The effect of stopping smoking is evident with 5 years. On stopping the rate ceases to increase as it would have if smoking had continued, but whether it actually falls is uncertain because the numbers are small ... The trend, however, suggests a fall followed by an increase, which keeps the rate ahead of that in lifelong nonsmokers".

Compared with continuing smokers, ex-smokers were found to have 35% of the lung cancer rate 5-9 years after stopping and 11% of the lung cancer rate 15+ years after stopping. For those periods after stopping risks relative to lifelong nonsmokers were respectively 5.9 and 2.0 times higher.

The multistage model cannot, of course, predict a declining risk after stopping unless the final stage of the process is

affected. However, as Doll notes, a true decline may not have occurred, the slight drop being explained by sampling variation or unrepresentativeness of ex-smokers. Doll's results seem not inconsistent with the multistage model, but clearly require that a late stage be affected to fit. The drop off, relative to continuing smoking, is far too large and rapid to be explained if only an early stage were affected. It will be interesting to see whether, when the 40 year results are published, the apparent approximate freezing of incidence rate on stopping continues for a longer period after stopping. As shown in the calculations above, the multistage model does not actually predict that the rate will stay constant on stopping, only that it will approximately do so for a period.

Kahn (1966) presented detailed tabulations, for smokers of cigarettes only, giving observed numbers of lung cancer deaths and annual death rates per 100,000 per year broken down by age (55-64, 65-74), age of starting to smoke (<15, 15-19, 20-24, 25+), maximum number of cigarettes smoked per day (1-9, 10-20, 21-39, 40+), and years since cigarette smoking stopped (continuing, 1-4, 5-9, 10-14 and 15+) based on 8½ years follow-up of the US Veterans Study. Those who had stopped smoking because of "doctor's orders" were excluded from analysis. Given age, it was generally evident that those who had given up smoking for more than 5 years had lower risks than those who continued to smoke, with risk declining with time given up. Smokers of age 65-74 who had given up for 10-14 years had higher risks (258) than those of age 55-64 who continued to

smoke (158), suggesting that the absolute risk did not freeze on stopping. A limitation of this study is the fact that smoking habits were only determined at one point in time.

Freedman and Navidi (1987, 1990) describe results of analyses based on a longer follow-up of the US Veterans Study, from 1954/57 to 1969. Again smokers giving up because of doctor's orders are omitted from analysis. 169 lung cancer deaths in ex-smokers of cigarettes only are considered compared to 113 reported by Kahn (1966). Freedman and Navidi compare risk by years of giving up smoking, standardized for amount smoked and age at giving up, i.e. they are testing whether absolute risk freezes on giving up smoking. For years of giving up of 0-4, 5-9, 10-14, 15-19, 20-24, 25-29, 30-34 and 35+ the standardized risks (numbers of lung cancers) were respectively 87 (26), 98 (45), 88 (52), 74 (25), 48 (11), 16 (6), 520 (4) and 0 (0). The risks for long-term giving up are based on small numbers of deaths and are difficult to interpret, but the pattern suggests some decline over a 20 year period. Compared to nonsmokers, the risk declined with increasing time of giving up, with no excess evident by 25 years. Without detailed study of the data, which are not presented so as to allow this, it is unclear why Freedman and Navidi's analysis appears to differ in conclusions from that of Kahn.

Hammond (1966) presents only limited data on ex-smoking from the first American Cancer Society Cancer Prevention Study. For men of age 50-69 who smoked (or had smoked) 20+ cigarettes a day

age-standardized death rates for lung cancer were 15 in never smokers, 205 in current smokers and, respectively, 437, 180, 108 and 16 in smokers who had given up for <1, 1-4, 5-9 and 10+ years. Following an initially higher rate for very short term ex-smokers, presumably related to why they gave up, the risk declined until no increase was evident for smokers who had given up for 10+ years. The pattern was similar for ex-smokers of 1-19 cigarettes a day, though less stable, being based on only 10 deaths in ex-smokers as against 93 for ex-smokers of 20+ cigarettes a day.

Freedman and Navidi (1987, 1990) also describe results of analyses based on the first ACS study. Based on five years follow-up and a total of 294 deaths in ex-smokers, they again compared risks by years of giving up smoking standardized for amount smoked and age at giving up. For years since quitting of <1, 1-4, 5-9 and 10+ years the standardized risks (numbers of lung cancers) were 158 (69), 114 (111), 83 (108) and 53 (6). Relative to nonsmokers the risks were estimated as 12.8, 7.8, 3.5 and 0.4. The decline in absolute risk, with risk going <u>below</u> that of nonsmokers after 10 years of quitting, are notable features of the data. Freedman and Navidi note that declining excess risk is not compatible with the versions of the multistage model normally considered. They consider various modifications of the model that might help to fit the data better (allowing for variability in waiting times from malignancy to clinical endpoint; allowing for rates of progression through the stages to vary from person to person; and allowing for individual variation in susceptibility),

but feel that "a more interesting idea is that the body can repair the lesions caused by smoking, and once the insult stops, the repair process is reasonably fast". They note that repair mechanisms are not compatible with the multistage model in standard form, but note that the idea is incorporated into the model used by Gaffney and Altshuler (1988). Freedman and Navidi do not, however, consider the possibility of bias due to non-representativeness of ex-smokers.

As described in more detail below (section 5.4), Brown and Chu (1987) found that a multistage model in which the first and penultimate stages were affected by smoking predicted reasonably well the variation seen in the Lubin et al (1984) study in risk of lung cancer in ex-smokers by years since smoking stopped, given age and duration of smoking.

Lubin et al (1984) themselves present some less detailed analysis of these data. One table gives risks of lung cancer by number of years since smoking is stopped (0, 1-4, 5-9, ≥10) and duration of smoking habit (1-19, 20-39, 40-49, ≥50). Another table gives risks by sex, number of years since smoking is stopped, and number of cigarettes a day (1-9, 10-19, 20-29, ≥30). There are some obvious limitations in these analyses. Firstly, duration of smoking habit, which is used directly in the first analysis, and as a standardizing variable in the second analysis, is not separated out into fine enough categories. Secondly, age at interview does not appear to have been adjusted for in any analysis.

In a study where cases and controls are matched on age, such adjustment is necessary to avoid marked bias in estimating risk by duration. Patterns reported of variation in risk by time of giving up smoking are, however, similar to those described by Brown and Chu (1987) (vide supra).

Halpern et al (1993) presented detailed data based on over 4000 lung cancer deaths occurring over a six year follow-up period in the American Cancer Society Cancer Prevention Study II (see Table 4). The observed patterns were similar in both sexes. For those quitting smoking between ages 30 and 49 lung cancer death rate rose gradually with age at a rate slightly greater than that for those who had never smoked. For those quitting between ages 50 and 64 risk levelled off near to that attained at the time of quitting until around age 75, when it rose sharply. At age 75, compared with the risk for current smokers, relative risks were approximately 0.45, 0.20, 0.10 and 0.05 for, respectively, those quitting in their early 60s, those quitting in their early 50s, those quitting in their 30s and those who had never smoked. The authors do not actually fit multistage models to their data, instead fitting a logistic model which contains terms in sex, education, age, cigarettes per day, years smoked and smoking status (and in some cases higher order terms and interactions). They note that the "plateau of risk in the age-at-quitting cohorts covering ages 50-64 is inconsistent with ... the Armitage-Doll multistage model, which predicts continuous increases" without pointing out that various forms of the multistage model predict

approximate constancy of risk for a period after stopping. They also note that their results are "inconsistent with the results of Freedman and Navidi (1990) who suggest that the absolute risk declines for about 20 years after cessation of smoking". Looking at Table 4, it is in fact notable that, in contrast to the data from the US Veterans Study and the first ACS study, there appears to be no real evidence at all of a decline in absolute risk following stopping. For example compare the risk in continuing smokers of age 54-58 (156.8) with that of ex-smokers who had given up at ages 55-59 (which is 244.0, 270.5 and 353.6 at, respectively, ages 64-68, 69-73 and 74-80). A similar conclusion can be reached for other ages of stopping.

Sobue et al (1993) describe analyses of data from a Japanese case-control study involving 776 lung cancer cases (553 current and 223 former smokers) and 772 controls (490 current and 282 ex smokers) all of whom started to smoke at age 18-22. Risk of lung cancer in ex-smokers according to the number of years given up was compared with that in continuing smokers, separate analyses being conducted for the overlapping age groups 55-64, 60-69, 65-74 and 70-79. The decline in relative risk was more rapid in the younger age groups (e.g. at age 55-64 RRs = 1.00, 0.85, 0.47 and 0.34 for current smokers and smokers giving up for 1-4, 5-9 and 10+ years) than in the older age groups (e.g. at age 70-79 RRs = 1.00, 0.85, 0.49 and 0.50), reflecting the fact that the smoking period as a fraction of total lifetime was greater at younger ages. Based on assumed values of risk by age for nonsmokers and continuing smokers

(these could not be assessed directly as cases and controls had been matched on age), the authors used their relative risk estimates to compute estimates of absolute risk by age at cessation, age at admission and years since cessation. The pattern was of a clearly increasing absolute risk after stopping smoking, though to less of an extent than occurs if smoking is continued. In interpreting the results from this study one should note that no adjustment has been made for number of cigarettes smoked. Nor has any attempt been made to exclude patients who gave up smoking for health reasons. Nevertheless the results clearly seem to conflict with those of the studies considered by Freedman and Navidi (1990) which suggested a decline in absolute risk on giving up smoking.

Lee (1974) analyzed the results from a mouse skin painting experiment in which groups of mice were treated with 180 mg/wk cigarette smoke condensate (CSC), with 600 mg/wk Fraction G of CSC, or with 36 or 60 μg/wk on benzo[a]pyrene for life or for various periods of time ranging from 10 to 50 weeks. Lee compared the tumour incidence observed with that expected under three hypotheses: no effect of stopping; tumour rate remaining constant at the time of stopping painting; and tumour rate remaining constant in weeks after stopping painting. For all types of treatment, it was clear that stopping painting reduced the tumour incidence compared with continuing painting. It was also clear that the tumour rate did not remain constant after stopping, this being evident from the simple observation that the groups painted for only 10 weeks had a zero tumour rate at 10 weeks (and indeed at 30 weeks for CSC and G) and

yet had an overall tumour yield far in excess of the untreated controls. In the benzo[a]pyrene treated groups incidence continued to rise after stopping painting but very much less steeply than it would have done had painting been continued. In the CSC and G groups painted for long enough for tumours to be seen before painting, incidence declined somewhat for 20 or 30 weeks after stopping and then rose, eventually markedly exceeding that seen at the time of stopping. A multistage model in which the carcinogens affected at least two stages of the cancer process, one early and one late, fitted the observed results quite well. For all the treatments the fitted effect relative to background was greater for the early stage than for the later stage, this being far more marked for benzo[a]pyrene than for CSC or G. It would be noted that the best fitted models for each treatment generally assumed that there was an effect on the <u>final</u> stage (as well as on other stages). Models in which only the first and penultimate stage were affected did not explain the drop-off in incidence observed after stopping in the CSC and G groups. It is interesting to note that for continuous painting best fitted Weibull distributions of the form $I = b(t - w)^k$ generally fit a positive value for w of about 10 weeks. This is consistent with the observation that, for benzo[a]pyrene, even at very high doses indeed, tumours are never seen before 11 or 12 weeks. The general interpretation of the w parameter is the time taken between the final mutation occurring and the tumour becoming clinically evident, and Lee carried out his model-fitting work under this assumption, i.e. he used the formulae in section 2 to estimate risk at time $t + w$ resulting from exposure occurring up to time t.

Lee actually points out that w may arise as the sum of constants $w_1 + w_2 + w_3 + \ldots$ representing fixed delays between a cell undergoing one mutation and being at risk of the next. He derived formulae for the risk in this more complex situation but never actually fitted them, due to the extensive and expensive nature of the computing involved. Such an extension of the model would seem required to try to reconcile the observation that there is a minimum time below which tumours cannot occur and the observation that risk may decline quickly after stopping.

3.5 Variation with age in relative risk associated with exposure

Many epidemiological studies appear to show that the ratio of the risk of lung cancer of a smoker of a fixed number of cigarettes a day to that of a nonsmoker (or to that of a smoker of a different fixed number of cigarettes a day) is approximately invariant of age, and indeed the formula proposed by Doll and Peto (1978) (vide supra) predicts exact invariance, with the terms in dose and age completely separable. However, inspection of formulae 6/1-6/3 shows that this simple relationship does not hold exactly. If, for example, one considers formula 6/3, taking $u_1 = u_2 = 1$, and $v_1 = v_2 = d$ for a smoker, and $v_1 = v_2 = 1$ for a nonsmoker, one can express the ratio of incidences at age T for a smoker (starting to smoke at age S) to that of a nonsmoker of the same age as

$$R = \frac{S^{k-1} + d(T^{k-1} - S^{k-1} - (T-S)^{k-1}) + d^2(T-S)^{k-1}}{T^{k-1}} \qquad (20)$$

For S = 20 years, k = 5 and d = 5, for example, one can readily calculate R for various values of T

T	R
50	7.50
60	8.90
70	10.18
80	11.31

The fact that R increases with T is not dependent on the precise values chosen of S, k or d, but is a general property, reflecting the fact that the greater the proportion of time one is exposed ((T-S)/T) the greater the relative risk. The rapidity of the rise in R with increasing age does however depend on which stages are most affected. Lee (1979) presents results of some illustrative calculations for a model in which the first and penultimate stages are affected and in which the relative risk at age 60-64 is assumed constant, the only variation being in the relative contribution of the first and penultimate stage effects (v_1 and v_2). Where v_1 is relatively small and v_2 relatively large, the increase in R with increasing age is quite modest, but as v_1 increases and v_2 decreases the increase in R with increasing age becomes relatively steep. This is illustrated by further calculations showing the rise in R with increasing T for S = 20, k = 5, d = 20 using formulae 6/1 (first stage only affected) and 6/2 (penultimate stage only affected)

T	First stage affected	Penultimate stage affected
50	3.46	19.51
60	4.75	19.77
70	5.95	19.87
80	7.01	19.93

There is rather little published data showing how the relative risk for smokers/nonsmokers varies with increasing age. Hammond

(1966) did observe some increase, with relative risks of 7.17 at age 35-54, 9.84 at age 55-69, and 10.67 at age 70-84, but Kahn (1966) did not, with relative risks of 11.30 at ages 55-64, and 7.03 at ages 65-74. However considerable sampling variation (due especially to relatively small numbers of lung cancer deaths among younger subjects) and failure to standardize for smoking duration (at that time the older men would certainly have tended to start smoking later than the younger men) makes these results difficult to interpret. The findings certainly do not seem inconsistent with the predictions of the multistage model, but they may be inconsistent with versions of the model in which the main effect of cigarette smoking arises from an early stage.

3.6 Effects of joint exposures

For continuous exposure to two agents, the joint dose response relationship will be very different depending on whether the agents affect the same or different stages of the cancer process. If the agents affected the same stage then the relationship should be additive, with the effect of a dose x of one agent being interchangeable with the effect of a dose y of the other, the ratio x/y reflecting the relative effectiveness of the different agents. If the agents affect different stages, however, the joint dose response should have a multiplicative component, the relationship becoming more multiplicative with higher doses as background effects become relatively weaker.

Evidence in favour of there being more than two stages comes from a number of studies which have shown multiplicative (or at least super additive) relationships between incidence and exposure to two agents. Selikoff and Hammond (1975) have reviewed some of the evidence on multiple risk factors in environmental cancer. Factors which show evidence of a multiplicative relationship with lung cancer include smoking and uranium mining, smoking and exposure to radiation from atomic bombs, and smoking and asbestos. The evidence for smoking and asbestos exposure is quite strong, with Hammond et al (1979) reporting lung cancer relative risks of 1, 5.2, 10.9 and 53.2 for exposure to, respectively, neither asbestos nor smoking, asbestos only, smoking only, or both asbestos and smoking (though small numbers of deaths in the group exposed to neither asbestos nor smoking may mean the apparent very multiplicative relationship was to some extent a chance finding). It would be interesting to see multistage models fitted to detailed joint exposure data but I am not aware that this has been attempted. One reason may be the lack of large studies providing detailed data on level, time of start and time of cessation of exposure.

Although, as noted below (see section 4.1), there is good animal evidence for some combinations of exposures that agent A followed by agent B elicits far more tumours than agent B followed by agent A, there appears to be little or no relevant

epidemiological evidence here. Peto (1984) in fact notes that the initiation/promotion phenomenon has never actually been observed directly in human carcinogenesis.

3.7 Effect of changing the type of cigarette smoked

Lee (1993a) recently reviewed the available epidemiological evidence relating risk of lung cancer to type of cigarette smoked. Although evidence relating to smoking cigarettes of tar 12 mg or less is still very sparse, there is quite substantial evidence that switching from plain to filter cigarettes or from higher to lower tar cigarettes is associated with some reduction in risk of lung cancer. Of 38 relative risk estimates associated with tar reduction or the plain/filter switch, 32 are less than 1.0, with the median 0.65. The fact that an apparent reduction in risk has been seen, despite the fact that in many studies smoking of the filter or lower tar cigarettes has only been for a relatively short period, is consistent with other evidence that smoking affects a late stage of the cancer process. As far as I am aware, however, no-one appears to have carried out formal multistage model fitting to such data.

3.8 Relationship of dose to age of onset of exposure

Passey (1962) noted that in a sample of hospital patients, age of onset of lung cancer appeared to be the same almost irrespective of their daily cigarette consumption, and argued that this provided evidence that cigarette smoke does not act as a carcinogen. That this line of reasoning was wrong was made clear by Pike and Doll (1965) in a paper which emphasized how misleading a statistic

average age at onset of a disease may be. While it is true that in animal experiments involving different doses of a strong carcinogen (which causes cancer in all or virtually all the exposed animals) increasing dose will lead to decreasing average age of tumour onset, this is not so for a weak carcinogen which leaves overall survival of the exposed population materially unaffected. If the function relating incidence rate to dose and time can be separated into terms dependent on dose and terms dependent on time, and the overall survivorship is similar in the various dose groups, it is apparent that the distribution of time of onset will be essentially independent of dose. Separability of dose and time is a characteristic of the Weibull expression $I = bd^c t^k$ and similarity of average age of onset in different dose groups is therefore consistent with this. In fact, two additional points which act in opposite directions need to be taken into account. The first is that, especially at higher ages, the proportion of heavy smokers surviving will be less than the proportion of lighter smokers, leading to some reduction in age of onset with increasing dose. The second is that, using a proper multistage formulation, and not the Weibull approximation, relative risk for heavy to light smokers increases with increasing age (see section 3.5), leading to some increase in age of onset with increasing dose. It should also be realized that variation in age distribution between heavy and light smokers and variation in age in the difference in mean age of starting to smoke between heavy and light smokers may upset any simple relationship.

Generally approximate similarity of mean age of onset of lung cancer in smokers of differing amounts is broadly consistent with the predictions of a multistage model, but the statistic is a difficult one to interpret and its use should be avoided if possible.

3.9 Other issues

Gaffney and Altshuler (1988) point out that, assuming a multistage model with the first and penultimate stages affected, the relative risk of heavy and lighter smokers will increase with increasing duration. Based on a best fit (six stage) to the Doll and Peto (1978) British Doctors data they point out that the relative risk comparing two packs a day and one pack a day smokers should increase from 2.5 at age 42.5 (smoking for 20 years) to 3.3 at age 72.5 (smoking for 50 years). In fact they noted that this prediction was not supported by the data. For smokers of, respectively, 17.5-27.5, 27.5-37.5, 37.5-47.5 and 47.5-57.5 years the relative risk of smokers of 25-40 cigarettes a day compared with smokers of 10-24 cigarettes a day was 2.5, 2.2, 2.5 and 1.6, i.e. there was no evidence of an increase in relative risk and indeed, in the highest duration category, some evidence of a decrease.

4. LIMITATIONS OF THE MULTISTAGE MODEL

4.1 Stages undefined

One obvious limitation of the multistage model is that it assumes that a number of stages must occur before the onset of cancer, but does not given any direct indication of what the stages might be. Although no clear evidence of what all the stages are has yet emerged (if indeed there are such stages and the model is not just a convenient mathematical approximation), there has been direct evidence for a long time that there are sequential aspects to carcinogenesis. It is over 50 years since it was demonstrated that the cocarcinogen croton oil was found capable of enhancing skin tumour induction when applied after a subeffective dose of carcinogenic hydrocarbon but not when applied beforehand. Such so-called "initiation/promotion" experiments led to the idea of "the two-stage hypothesis". See Berenblum (1982) for a comprehensive review of the evidence relating to sequential aspects of chemical carcinogenesis in the skin, where much of the work has been conducted. It is interesting to note that for many years it was unclear whether cocarcinogens of tumour promoter type were actually relevant to man. Recent observations by Hecker (1984) in the Caribbean island of Curaçao are of particular interest here. On this island the black and Creole population have an extremely high rate of oesophageal cancer and, as part of the local diet, the fresh green leaves of the aromatic bush known as "welensali" are commonly used to prepare a "bush tea". One cup of tea prepared from this

bush, Croton flavens L, contains very high levels indeed of known tumour promoters, and Hecker makes a strong case for this being responsible for the high oesophageal cancer rate.

It is possible that molecular genetic studies may help to identify the stages required for tumorigenesis. Renan (1993), in a paper attempting to answer the question as to how many mutations are required, notes that "molecular studies have strongly supported the idea that multiple genetic changes are required". He cites the example of colorectal malignancies, "which involve genetic alterations on chromosomes 5q, 12q, 18q and 17p and possibly other lesions as well".

4.2 Reversibility of effects may occur

As specified, the multistage model does not allow for reversibility of any of the stages. Over time the numbers of cells that have passed through the various stages can only increase. Conceivably, for some stages at least, damage may be repaired. Though, for continuous exposure, taking the possibility of reversibility into account should not affect the mathematical approximations (the transition probabilities can be viewed as differences between probability of damage minus probability of repair), this need not be the case for discontinuous exposure. Clear evidence that incidence declines in absolute terms after stopping would suggest reversibility and indicate the assumptions behind the multistage model are too simplistic.

4.3 Transition probabilities may vary from individual to individual for a given exposure

For a given exposure it is assumed by the multistage model that the transition probabilities for each stage do not vary from individual to individual. For a disease with a large genetic component this may be an inappropriate assumption. If the population actually consists of two groups of individuals, a susceptible group with non-zero transition probabilities for each stage, and a non-susceptible group with zero transition probabilities for one or more stages, then it is easy to see that one will not observe the simple relationship between incidence rate and age (formula 2) predicted for continuous exposure. Rather the incidence rate, instead of rising continuously with age, will fall off past a given point in time as the susceptibles are depleted, perhaps eventually reaching zero when only non-susceptibles remain. Sellers et al (1990), using segregation analysis, reported finding that lung cancer patients could be divided into three groups, one with a much higher risk of early onset disease (given smoking habits and occupation) than the other. This suggestion of a genetic component is supported by evidence (summarized by Lee, 1993b) that family history of lung cancer is an independent risk factor for lung cancer. The extent to which such genetic variation will modify predictions from the multistage model is not clear at this point in time.

In their analyses relating incidence of cancer (I) of 31 types in 11 populations to age (t), Cook et al (1969) found that in 54% of

cases there was evidence of downward curvature from the theoretical straight line relationship predicted by the Weibull formula $\log_e I = \log_e b + k \log_c t$. One possible explanation that they considered for this (apart from underdiagnosis in old age or differences in exposure between different age-cohorts) was that only a proportion of the population might be susceptible to cancer. If the initial proportion of susceptibles is C, it can be shown that instead of the simple relationship given above, the relationship will be of the form

$$\log_e I = b+k \log_e t - \log_e [C + (1-C)e^{F/C}] \qquad (21)$$

where $F = e^a t^{k+1}/(k+1)$.

They presented a graph showing that the extent of downward curvature is very small indeed for C even as low as 0.1 or 0.05. Only for C = 0.01 did substantial downward curvature occur with incidence falling off after age 60. They pointed out that if susceptibility were the explanation for the downward curvature one would expect to see an increased amount of curvature with increasing levels of incidence in genetically similar populations. However the data did not appear to support this. They concluded that there was "no evidence ..... to suggest that the shape of the observed relationship could be attributed to attenuation of a limited pool of susceptibles".

Peto et al (1985) cite data of Parish (1981) to support the idea that there is considerable variation among outbred mice in

their susceptibility to skin cancer induced by chronic benzo[a]pyrene treatment. A figure was presented comparing the new tumour incidence rate/time relationship of mice who had respectively 0, 1, 2 or 3 tumours already. There was a clear tendency for incidence at a given time to increase with the number of tumours already present, and for the log incidence/log(duration of exposure - 15) relationships to show downward curvature from a straight line. Peto et al note that their results are consistent with substantial heterogeneity of susceptibility with risk varying 100-fold between the upper and lower 95% extremes of the distribution. As they note, the more susceptible an animal is, the more tumours it is likely to have already, thus explaining the higher risk with increasing numbers of tumours present. They also note that failure to take into account variation in susceptibility will lead to underestimation of the true number of stages of the cancer process. Elsewhere, Doll (1978) makes it clear that substantial variation in susceptibility is not inconsistent with relatively small differences in risk associated with family history of cancer. Consider, for example, a recessive gene that increases the risk of a particular cancer 50-fold in homozygotes. The relative risk in the siblings of probands would then be just over 4-fold if the population frequency of the gene was approximately 10%.

One possibility apparently not considered in the literature is that, within an individual, all the cells capable of being transformed to cancer of a particular type may not be equally susceptible.

4.4 <u>The model may be inaccurate if the transition probabilities are not small</u>

Consider a two stage process in which both transition probabilities are equal, having the value a. The probability, $1-G_T^*$, of a cell surviving tumour free at time T is then given by the expression:

$$1-G_T^* = e^{-aT} + 2\int_0^T ae^{-au}e^{-a(T-a)}du \qquad (22/1)$$

$$= e^{-aT}(1+aT) \qquad (22/2)$$

The probability, $1-G_T$, of the organism, with N cells, surviving tumour free at time T is then given by:

$$1-G_T = (1-G_T^*)^N \qquad (23)$$

The incidence rate of cancer at time T, I_T, is then given by:

$$I_T = \frac{dG/dT}{1-G} = \frac{Na^2T}{1+aT} \qquad (24)$$

This compares with the standard approximate form of the incidence rate given by formula 1 in section 2, of:

$$I = Na^2T \qquad (25)$$

The exact form of the incidence rate would show some downward curvature when log I is plotted against log t, whereas the approximate form would not. This would also be true for the more general situation of a k stage process, with differing transition probabilities from stage to stage (see Hakama (1971) for the more general exact formulae).

The question arises as to how adequate the approximate form of the incidence rate formula actually is. In discussion on Hakama (1971), Moolgavkar (1977) noted the approximate Armitage-Doll formula can be viewed as the first term in an infinite (Taylor) series expansion of the solution, and that retention of additional terms in the power series would give a better approximation and might explain some of the deviations from the theoretical incidence curve noted by Cook et al (1969). Peto and Doll (1977) and Hakama (1977), in reply to Moolgavkar's letter, point out that in practice the Armitage-Doll approximation is extremely good, and that downward curvature in the lung cancer incidence rate curve is much more likely to result from underdiagnosis of lung cancer in the elderly, from cohort effects or from selective mortality, than it is to result from a poor approximation of the formula.

This can be illustrated by considering the two stage process above. Suppose we consider incidence at age 70. The annual incidence rate of lung cancer will not exceed 1 in 100. Given a fairly conservative number of cells at risk of 10,000, one can readily calculate that the annual transition probability per cell is about 1.2×10^{-4}. The difference between $1+aT = 1.008$ and 1 is really then quite small compared with other sources of variation. A similar conclusion can be reached using higher numbers of stages. The approximateness of the formula does not seem to be a problem in practice.

4.5 Other problems

As noted above, genetic heterogeneity may have the effect of altering the observed power of time, so that evidence of a kth power relationship between incidence and time (or duration of exposure) does not necessarily imply there are k+1 stages of cancer. Peto (1984) notes that other factors, including selective proliferation and diagnostic delay may also have this effect by altering the observed power of time.

Although the multistage model has been expressed in terms of mutations occurring since birth, it is possible that cancer may arise in individuals who are born with one (or more) of the mutations already present. See for example the retinoblastoma model proposed by Knudson (1971).

In his paper on multistage models, Peto (1977) points out that though they "hold out the most promise of being a useful framework for describing the process of neoplastic transformation, there are various observations which do not appear to fit naturally into the multistage formulation". These include:

(i) The fact that given age and dose of carcinogen, an animal is more likely to get a tumour if it already has a tumour of the same type than if it does not;

(ii) The existence of tumours of mixed cellularity; and

(iii) The fact that when mutagens are applied to cells in vitro it is much easier to cause neoplastic transformation than it is to cause gene mutation.

For all the problems, and a discussion, the interested reader should refer to Peto (1977).

5. APPLICATIONS OF THE MULTISTAGE MODEL

5.1 Using data on prevalence of smoking at different ages

Section 2 gives formulae, based on the multistage model, for one continuous period of smoking (formulae 7 and 8 and for two continuous period of smoking (formula 10). Formulae can also readily be derived for multiple periods. In cohort (or case-control) studies, where data are available on an individual basis concerning a person's lifetime smoking history, these formulae can be derived directly. However a number of coworkers have attempted to fit multistage (or other) models to national age-specific lung cancer incidence data where the only data available are cohort-specific percentages of smokers each year or each five years (sometimes accompanied by data on average consumption levels).

In order to convert these percentages into estimates of the frequency of people smoking for different periods of time (and hence use the multistage model formulae) it is necessary to make some assumptions. For example, if there were two time periods with 30% smokers in the first and 40% in the second there are various possibilities, including:

(i) 30% smoking throughout, 10% smoking only in the second period.

(ii) 30% smoking only in the first period, 40% smoking only in the second.

(iii) 20% smoking throughout, 10% smoking only in the first period, 20% smoking only in the second.

The first possibility maximizes the proportion of long duration smokers, the second minimizes it. The third is one of many intermediate possibilities.

When attempting to get round this problem, Townsend (1978) assumed that smokers can be ordered from "hard core" to "highly capricious", so that the frequency of longer duration smokers is maximized. If, for example, the percentages of smokers at six successive time periods are 20, 30, 45, 40, 50 and 35, one can divide the population into 20% smoking throughout, 10% (= 30-20) smoking at all times except in the first period, 5% (= 35-30) smoking at all times except in the first and second, 5% (= 40-35) smoking at all times except the first, second and sixth, and so on.

An alternative approach was used by Swartz (1992). Here it was assumed that smokers, once they give up, never start again. If, for example, the percentages of smokers at six successive time periods are 10, 20, 10, 20, 10, 20, Swartz would assume there are four groups of people, 10% who smoke throughout, 10% who smoke only in period 2, 10% who smoke only in period 4, and 10% who smoke only in period 6. This contrasts with Townsend's assumptions, which would

involve only two groups, 10% who smoke throughout and 10% who smoke only in periods 2, 4 and 6. Hakulinen and Pukkala (1981) appear to make similar assumptions to Swartz. It should be noted that the Swartz assumption may, with certain data, lead to more than 100% of the subjects being classified into smoking groups.

In theory it would be possible to investigate the validity of either approach using data from a study in which detailed lifetime smoking histories were collected, but no such investigation appears to have been carried out. On general grounds it seems that both approaches are likely to be incorrect, the first probably overestimating risk, the second probably underestimating it.

5.2 Applications to cohort data

Mazumdar et al (1991) describe techniques for fitting multistage models with two stages dose-related to cohort data. Their methodology and software allow for exposure to vary over intervals during the person's life as may be needed for occupational mortality studies with detailed exposure data. The method is illustrated using lung cancer mortality data for a cohort of non-white male coke oven workers exposed to coal tar pitch volatiles and shown to fit adequately. This group at the University of Pittsburgh are extending their software to fit alternative models proposed by Moolgavkar and his colleagues. Those intending to do detailed fitting of such complex data would do well to approach the authors, though note that the computing was done on a CRAY Super Computer!

5.3 Whittemore (1988)

Whittemore (1988) used data from three sources to test the fit of two functions relating lung cancer incidence to smoking habits. The first two sources, the British Doctors Study (Doll and Peto, 1978) and the US Veterans Study (Kahn, 1966), presented data on risk for current smokers and for lifelong nonsmokers. The third source, a case-control study of non-Hispanic white men in New Mexico, data for which were provided by Prof J Samet, had detailed lifetime smoking histories, and so provided a more rigorous test. The first function used, the packs function g_1, specified that the excess death rate at age t depended linearly on the cumulative amount smoked

$$g_1 = 2.01 \times 10^{-12}(t - 5)^{4.5}(1 + \alpha P) \qquad (26)$$

where P is the total number of packs of cigarettes smoked by age (t - 5) and α is a constant to be specified. The second function used, the multistage function g_2, specified that the death rate at age t is of the form

$$g_2 = 2.01 \times 10^{-12}[(t - 5)^{4.5} + pc(1 + 2pc)(t_1 - t_0)^{4.5} + 2pc(t_1^{4.5} - t_0^{4.5})] \qquad (27)$$

where c is the number of cigarettes per day and p is a constant to be specified.

Whittemore found that both functions fitted the British Doctors data with best-fitting parameters $\alpha = 1.13 \times 10^{-3}$ and p = 0.207, there being little to choose between the functions.

With the US Veterans' data, best-fitting parameters were lower, $\alpha = 0.59 \times 10^{-3}$ and $p = 0.128$, but neither function fitted the data very adequately, there being a notable tendency to overestimate risk at age 65-74 (624 deaths expected vs. 576 observed using g_2), and to underestimate it at age 55-64 (477 E vs. 547 O). For the New Mexico data, g_2 fitted markedly better than g_1. However there was some tendency for g_2 to overestimate risk in ex-smokers (68 E vs. 45 O) and to underestimate it in current smokers (166 E vs. 179 O). Both functions, however, explained substantially more variation in the New Mexico data than did any of several logistic regression models involving categorical variables for age and smoking.

Some points to note about this work are as follows:

(i) The function g_1 is stated to indicate excess risk. However as it is not zero for $P = 0$ it presumably actually was intended to indicate actual risk. The function is in any case not of a form predicted by the multistage model.

(ii) The function g_2, stated to be based on a multistage model in which the first and penultimate stages are affected, the penultimate stage being twice as strongly affected as the first, is actually incorrectly derived (or has been misreported). As noted elsewhere (see section 2), the term $pc(t_1 - t_0)^{4.5}$ should be replaced by $pc[(t - t_0)^{4.5} - (t - t_1)^{4.5}]$. This does not affect the fit for continuous

exposure, where $t_1 = t$, but gives different predictions for ex-smokers. The fit to the New Mexico data will therefore be in error.

(iii) The nonsmoker part of the function, $2.01 \times 10^{-12}(t - 5)^{4.5}$, was based on a fit to nonsmokers' data from the American Cancer Society CPS I study. Since these subjects are unrepresentative, and since there are a multitude of risk factors in nonsmokers, this function may not be fully appropriate for other data. It is surprising that Whittemore did not at least try the effect of fitting constants other than 2.01.

(iv) When fitting the New Mexico data, Whittemore tried using α and p values fitted to either the British Doctors data or the US Veterans data. The values for the US Veterans study fitted much better and were used in her main work. It was surprising that Whittemore did not try to determine the parameter values which best fitted the New Mexico data.

(v) Commenting on the lack of fit of the models to the US Veterans' data, Whittemore notes that this may be due to inadequate smoking data. Numbers smoked were determined only at the start of the study and may have changed both before and after.

5.4 Brown and Chu (1987)

Brown and Chu (1987) carried out detailed analyses relating cigarette smoking to lung cancer based on the large multicentre West European prospective study of Lubin *et al* (1984) involving 6920 male

patients and 13460 male controls. They compared the risk of lung cancer in smokers who had given up for 3, 4, 5-6, 7-8, 9-11, 12-15, 16-20, 21-26 or 27+ years of smoking with those who had continued to smoke (including those who had given up for 1 or 2 years in this group), after adjustment for reason for quitting, study area, age at interview, number of cigarettes smoked, duration of smoking, frequency of inhalation, and percent of time smoking nonfiltered cigarettes. The pattern of relative risks, 0.99, 0.78, 0.71, 0.69, 0.48, 0.47, 0.39, 0.44 and 0.40 for the nine ex-smoking groups, was shown by the authors to be quite well predicted by a multistage model in which the penultimate stage only was affected, and somewhat better predicted by a model in which both the first and penultimate stages were affected, the latter predicting a flattening out and eventual slight increase in the relative risk many years after giving up smoking. The authors emphasized the importance of adjustment for duration of smoking in their analyses. Had no adjustment been made, the fitted pattern of decline in relative risk with years given up smoking would have been much steeper, declining to 0.17 after 27+ years. Two features of the study design should be noted. One feature is the very large number of deaths, which means that the relative risk estimates have small sampling error (e.g. the estimate of 0.69 for having given up 7-8 years has 95% confidence limits of (0.56 - 0.84). The other feature is the fact that cases and controls were age matched. This means that comparisons cannot be made of risk of subjects in different age groups, so that one cannot compare risk in ex-smokers with that in smokers at the time they gave up.

Brown and Chu also carried out analyses relating risk in smokers who had started to smoke at ages ≤14, 15, 16, 17, 18, 19-20 and ≥21 with that in nonsmokers after adjustment for study area, age at interview, number of cigarettes smoked, frequency of inhalation and percent of time smoking nonfiltered cigarettes. The relative risks in general showed a declining pattern with increasing age of start (3.6, 4.1, 4.0, 4.0, 3.6, 3.4, 2.9 - 95% confidence limits are about ±0.8 on each estimate) with the exception of the group starting at age ≤14. The pattern of decline was found by the authors to be much better fitted by a multistage model in which the first and penultimate stages were affected than by models in which only the first, or only the penultimate stage was affected.

The authors also fitted the overall data to try to determine the relative effect of smoking on the first and penultimate stages, for smokers of 1-10, 11-20, 21-30 and 31+ cigarettes per day. The best fit values for all four smoking categories were found to indicate a higher penultimate stage than first stage effect (2.8 vs. 0.7 for 1-10 cigs/day, 5.0 vs. 2.5 for 11-20, 6.3 vs. 3.5 for 21-30, and 7.0 vs. 4.0 for 31+). On average smoking appeared to have about twice the effect per unit dose on the penultimate stage than on the first stage. This work was the basis of the assumption used by Whittemore (1988) that smoking had twice the effect on the penultimate stage that it had on the first stage. Especially as the various relationships seen were found to be consistent over subsets

of the data by age, duration of smoking and number of cigarettes smoked, the results appear to provide quite strong support for the multistage model.

5.5 Other authors

Brown and Chu (1983a,b) analyzed the incidence of lung cancer during the period 1938 to 1973 in a cohort of men occupationally exposed to arsenic and other contaminants. After adjustment for duration of exposure they found a clear tendency for risk to increase with increasing age of starting employment. They interpreted their findings as indicating that arsenic appeared to exert a definite effect on a late stage of the carcinogenic process, although their analyses could not conclusively rule out a possible additional effect on the initial stage. The data were found to be adequately fitted by a multistage model in which occupational exposure affected the penultimate stage. No data were available for cigarette smoking on this cohort, but evidence from other studies was cited by the authors in support of the view that this would not materially have biassed the results.

Day (1984) is a review paper demonstrating that a wide range of epidemiological phenomena can be described in terms of simple multistage models of carcinogenesis. He notes "the relationship of cancer risk with the different time variables considered corresponds closely with the behaviour predicted by theories of multistage process. Furthermore, the different behaviour associated with different agents enables one to attempt some classification as to

how an agent is acting". Day considers evidence inter alia on asbestos and mesothelioma and lung cancer, on ionizing radiation and cancer of various sites, on arsenic and lung cancer, on nickel and nasal sinus cancer, on chloromethylethers and lung cancer, on various risk factors for breast cancer, and on exogenous oestrogen exposure and endometrial cancer. The last is interesting in that it is the only well documented occasion in cancer epidemiology of a last stage agent, absolute excess risk disappearing after exposure stops.

An earlier review paper, reaching similar conclusions, is that by Day and Brown (1980). Included in this paper are some analyses of the Tobacco Research Council Stopping painting experiment, from which they concluded that Fraction G of smoke condensate T57 behaved like a carcinogen affecting predominantly a late stage carcinogen, in contrast to benzo[a]pyrene which behaved more like a carcinogen predominantly affecting an early stage carcinogen. These conclusions are not dissimilar from those by Lee (1974) described in section 3.4.

6. MODIFIED VERSIONS OF THE MULTISTAGE MODEL

Some authors have attempted to fit models based on the multistage model but using formulae not actually predicted by it.

6.1 Doll and Peto (1978)

Doll and Peto (1978) fitted the function

$$I = 0.273 \times 10^{-12}(\text{cigarettes/day} + 6)^2(\text{age} - 22.5)^{4.5} \qquad (28)$$

to 20 year follow-up data from the British Doctors, restricting attention to men aged 40-79, and to lifelong nonsmokers or to subjects who reported same amount of 40 or less per day at each interview. The fit was found to be adequate, but it should be realized that the functional form is not strictly multistage (it should contain terms in duration^k <u>and</u> in age^k), although it may be a fairly close approximation. The issues relating to exclusion of subjects smoking more than 40 cigarettes per day and of subjects aged 80+, justified by Doll and Peto at length in their paper, have already been discussed. One limitation of the British Doctors study is that it contains no data on age of starting to smoke.

6.2 Townsend (1978)

Another attempt to use a function related to the multistage model, but not actually predicted by it is that by Townsend (1978). Her model, described in detail in the original paper, was expressed in terms of the sum of three components:

(a) a product of a length of smoking effect and a level of smoking effect for cigarette smokers,

(b) a similar product for smokers of other products, and

(c) an effect for nonsmokers.

The length of smoking effect was of the form

$$\sum_i (e_i z_i^{\,k}) \;/\; \sum_i e_i \qquad (29)$$

the population being divided into i groups of smokers with frequency e_i who had smoked for duration z_i.

The level of smoking effect was of the form

$$\sum_t((t - w)^{\beta} e_t l_t f_t) \;/\; \sum_t (t - w)^{\beta} \qquad (30)$$

where t is age, w is age of starting to smoke, and e_t, l_t and f_t are respectively the values at time t of the proportion of smokers, the number smoked and a cigarette effect parameter (depending on weight of tobacco, tar content and plain/filter status). The function is a weighted mean of smoking levels at each age, the weight $(t - w)^{\beta}$ indicating the importance of recent relative to past smoking, recent smoking being more important for $\beta>0$.

Using national annual age and sex specific data on percentage of smokers, generated partly by extrapolation, and other data on type of cigarette, Townsend fitted the model to England and Wales lung cancer data from 1935 to 1970 by five-year age and time periods. The model tended to overestimate rates for 1935-1945 and to fit male data much better than female data. Even after putting in terms to account for likely greater underdiagnosis of lung cancer, the model did not fit the data well for females, predicting downturns in mortality at higher ages in the latter half of the period that were not seen.

The model, although intended to be based on multistage principles, is clearly not a true multistage model. Inter alia, the effects of length and of level of smoking are not separable, and the

effects of cigarette smoking, cigar/pipe smoking and nonsmoking are not independent. There are also problems with the extrapolated smoking data, detailed surveys only being carried out annually from 1948. This work does not really add to any conclusions regarding adequacy of the multistage model.

7. DISCUSSION AND CONCLUSIONS REGARDING THE MULTISTAGE MODEL

As a mathematical model for describing variation in lung cancer incidence rate by age, dose and duration of exposure, there is no doubt that the multistage model has proved useful and popular. Certainly its properties have been more widely discussed and are more widely understood than any of the other models which we will consider in a later document. The multistage model has a lot going for it: it is flexible, reasonably tractable, and in broad terms its predictions fit in with a number of observed facts. These include:

(i) the approximate power law relationship of incidence with duration of exposure when exposure is continuous;

(ii) evidence that age *per se* does not affect incidence of many cancers;

(iii) direct evidence from initiation/promotion studies that some cancers require multiple exposures in a specific order for cancer to arise;

(iv) the observation that tumour incidence may be increased as a result of exposure that has long since ceased;

(v) evidence of quadratic dose-response relationships for some carcinogens;

(vi) explaining why the joint effect of two carcinogens is often

multiplicative, or at least markedly super-additive; and

(vii) describing reasonably well patterns of incidence following cessation of exposure.

It would be asking too much of any model to describe adequately all aspects of the variations seen in lung cancer incidence rate. Even in a carefully controlled animal experiment in which precisely defined doses are given at predetermined points in time and animals are randomized to different groups there will inevitably be some sources of variation that will not be completely accounted for. Animals and cells within animals are unlikely to be totally homogeneous in susceptibility for example, so that the multistage assumption that each similarly exposed animal is effectively identical, containing an identical number of identical cells, can at best only be an approximation to reality. That, however, need not be an important limitation if models are seen in the light in which they are put forward, namely as a means of approximately explaining known facts and of making reasonable approximate predictions.

In judging the usefulness of a model, one has to consider whether its predictions materially break down in any circumstances. Much of the testing of the multistage model has been carried out on data from epidemiological studies, and it is important to be aware that such data are limited in a number of ways. These include:

(i) inaccuracy of diagnosis of disease;

(ii) inaccurate quantification of average extent of exposure;

(iii) inadequate details on changes in exposure;

(iv) inadequate information on other causes of the disease which may confound the smoking/lung cancer relationship. In this respect it is important to realize that nonsmokers, light smokers, heavy smokers and ex-smokers are not randomly selected and are likely to be systematically different in many respects. Comparison of ex-smokers with continuing smokers is a particular problem in this respect, since the decision to give up smoking may be related to several factors (including illness and increased health awareness) that are themselves linked to risk of disease.

Bearing in mind these difficulties in interpreting epidemiological data, are there any features of the smoking/lung cancer data that the multistage model notably fails to predict? Certainly, providing it is assumed that smoking affects two distinct stages of the process, probably the first and penultimate stage, the multistage model does not in general do too badly. There are, however, three aspects of the data where it appears that it may have some difficulty.

The first of these is the dose-response relationship, some studies indicating an apparent linear relationship of incidence with number of cigarettes smoked when the requirement for smoking to affect early and late stages of the process (needed to explain relationships of incidence to age at starting to smoke and to time since stopping smoking) would suggest a quadratic relationship. When one bears in mind that a multistage model with two stages

moderately affected only actually predicts a relationship that has only a modest quadratic component, and when one realizes that inaccuracies in measuring exposure are likely to reduce the slope of the dose-response relationship, it is not at all clear that this objection undermines the validity of the model. The evidence presented by Doll and Peto (1978) based on the British Doctors data and the arguments they put forward can be seen as a reasonable defence of the model.

A second apparent difficulty of the multistage model that has been referred to is the fact that in the British Doctors data there is no evident tendency for the ratio of risks of heavy to light smokers to increase with increasing age. Gaffney and Altshuler (1988) draw attention to this, pointing out that an increase with age in this ratio would be predicted by the multistage model. Bearing in mind the following facts:

(i) the predicted rise is not very large anyway;

(ii) the data on number smoked may not be completely reliable;

(iii) ability to smoke a large number in an old man may be an indicator of reasonable health (put another way, symptomatic smokers may cut down); and

(iv) the lack of data in the Doctors study on age of starting to smoke;

I would not regard this point as a major one. It would be valuable, however, to see additional analyses from other studies to try to confirm whether in fact the overall evidence does or does not indicate a rise in relative risk with increasing age.

The final, and most serious, apparent difficulty relates to the data on giving up smoking. Under a multistage hypothesis in which any stages are affected except the last, the incidence rate of lung cancer will continue to increase on giving up smoking, though the slope of the increase will depend dramatically on which stages are affected. As shown in section 3.4, the rise will be much greater if the first stage is affected than if the penultimate stage is affected. Even if both the first and penultimate stages are affected the rise may be only relatively modest for some considerable time, provided the penultimate stage is affected more than the first stage.

A decline in absolute risk can occur if the last stage is affected, but this will be immediate and not a gradual decline. Freedman and Navidi (1990) have claimed that the epidemiological evidence indicates that absolute risk of lung cancer declines on giving up smoking and that this is inconsistent with the predictions of the multistage model. Gaffney and Altshuler (1988) have also argued that the multistage model is inadequate because it cannot simultaneously fit the incidence in smokers and ex-smokers. They argue that the best fit to the data for continuing smokers predicts that excess incidence will greatly increase in ex-smokers whereas the data indicate no change or a decrease.

In interpreting this evidence a number of important points should be made:

(i) Freedman and Navidi, and Gaffney and Altshuler, pay little attention to the problems of bias caused by the non-representativeness of ex-smokers. Some studies, but not all, attempt to get round the bias due to some smokers giving up because of severe illness. If ignored, this might give the false impression that giving up smoking markedly increases risk of lung cancer in the short term. More difficult to adjust for is the bias in the reverse direction resulting from the likelihood that those who give up, because they have less inherent desire to smoke than those who continue, are more likely to have been smokers who have smoked in a way that predicts less risk regardless of whether they give up. They may have smoked less, inhaled less, smoked to a longer butt, smoked lower tar brands, etc., facts which are difficult, if not impossible, to adjust for completely.

(ii) The available data on risk in continuing smokers by age and number of cigarettes smoked do not actually permit reliable estimation of the relative effect of smoking on the first and penultimate stages to be made. Contrast, for example, Gaffney and Altshuler's best six-stage fit, based on the British Doctors data, which estimated the first stage effect to be almost three times stronger than that on the penultimate stage, with the work of Brown and Chu (1987) based on the Lubin study which estimated that the penultimate stage effect was about twice that on the first. While these estimates make different predictions about the pattern of risk on giving up smoking, neither should be relied upon. As regards the British

Doctors data, the absence of information on age at starting to smoke should particularly be noted, as taking it into account may have affected the predictions considerably.

(iii) Neither Freedman and Navidi, nor Gaffney and Altshuler, consider all the relevant data on ex-smoking (albeit some have appeared since their papers were published). Gaffney and Altshuler's analysis was based solely on the 20 year follow-up of the British Doctors data, which did not involve a large number of lung cancer deaths in ex-smokers. The "freezing" of the rate on stopping is clearly at best only an approximation. Doll (1978) in fact notes the data suggest a slight fall followed by an increase. Freedman and Navidi's analysis was based on two data sets for ex-smokers: the US Veterans data which appeared to show a slight decline in absolute risk on giving up smoking and the ACS CPS I data which appeared to show a more marked decline. Neither study, however, is based on a large number of lung cancer deaths in ex-smokers (169 in the Veterans, and 294 in the ACS study), and the numbers are particularly low as regards longer term ex-smokers. Thus the Veterans Study only has 21 deaths for ex-smokers who have given up for 20 years or more, while the ACS CPS I study only has 6 deaths for ex-smokers who have given up for 10 years or more (and this group remarkably shows a lower absolute risk than in nonsmokers - a fact that would not be explained by any model). More recent data, based on much larger numbers of lung cancer deaths in ex-smokers, show a very different pattern. Particularly noteworthy are the

case-control study of Lubin et al (1984) which involved almost 2000 lung cancer deaths in ex-smokers and the ACS CPS II prospective study (Halpern et al, 1993) which involved over 1000. The pattern of response in ex-smokers in the Lubin study was found by Brown and Chu (1987) to be well described by a multistage model, though the fact that the case-control study was age matched makes it impossible to determine trends in absolute risk from the time of giving up. The most interesting data set in this respect is that from the ACS CPS II study. As shown in Table 4, it is quite clear when one looks at trends in risk over a long period in time that risk does not decline or freeze, it clearly increases with age. Whether one considers absolute or excess risk, the increase in risk with increasing age in ex-smokers is clearly evident. It seems likely, though this has not formally been tested, that the pattern of risk in Table 4 could be fitted quite well by a multistage model. Certainly it would not fit the suggested alternative "two-stage model with clonal growth" of Gaffney and Altshuler (1988) which predicts constant excess risk in ex-smokers on giving up. The rise in risk between ages 69-73 and 74-80 in smokers giving up at age 60-64 from 409 to 607 per 100,000 per year is clearly vastly greater than the corresponding rise for lifelong nonsmokers from 31 to 39 per 100,000 per year (each of these rates being highly stable since they are based on about 100 lung cancer deaths).

Although a more certain evaluation could perhaps be reached by a further simultaneous detailed investigation of all the data, one must conclude that the multistage model remains a very useful one. There appears no obvious reason at this point in time why predictions based on it should not be quite reliable.

8. REFERENCES

Armitage P. A note on the time-homogeneous birth process. J R Statistic Soc Series B 1953;15:90-1.

Armitage P. Discussion on Doll's paper. J Royal Statistic Soc A 1971;134:155-6.

Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer 1954;8:1-12.

Berenblum I. Sequential aspects of chemical carcinogenesis. In: Becker FF (ed). Cancer. A comprehensive treatise, Vol 1. 2nd ed. Plenum Press, New York, 1982. 451-84.

Brown CC, Chu KC. Implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. JNCI 1983a;70:455-63.

Brown CC, Chu KC. A new method for the analysis of cohort studies: implication of the multistage theory of carcinogenesis applied to occupational arsenic exposure. Environ Health Perspec 1983b;50:293-308.

Brown CC, Chu KC. Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. J Chronic Dis 1987;40(Suppl 2):171-9.

Cook PJ, Doll R, Fellingham SA. A mathematical model for the age distribution of cancer in men. Int J Cancer 1969;4:93-112.

Crump K, Howe RB. The multistage model with a time-dependent dose pattern: application to carcinogenic risk assessment. Risk Analysis 1984;4:163-76.

Davies RF, Lee PN, Rothwell K. A study of the dose response of mouse skin to cigarette smoke condensate. Br J Cancer 1974;30:146-56.

Day NE, Brown CC. Multistage models and primary prevention of cancer. JNCI 1980;64:977-89.

Day NE. Epidemiological data and multistage carcinogenesis. In: Models, mechanisms and etiology of tumour promotion. IARC Scientific Publications No 56, Lyon, 1984:339-57.

Doll R. The age distribution of cancer; implications for models of carcinogenesis. J Royal Statistic Soc A 1971;134:133-55.

Doll R. An epidemiological perspective of the biology of cancer. Cancer Res 1978;38:3573-83.

Doll R, Gray R, Hafner B, Peto R. Mortality in relation to smoking: 22 years' observations on female British doctors. BMJ 1980;1:967-71.

Doll R, Peto R. Mortality in relation to smoking: 20 years' observation on male British doctors. BMJ 1976;2:1525-36.

Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. J Epidemiol Community Health 1978;32:303-13.

Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. JNCI 1981;66:1191-308.

Druckrey H. Quantitative aspects in chemical carcinogenesis. In: Truhaut R, editor. Potential carcinogenic hazards from drugs. UICC Monograph Series Vol 7, 1967:60-78.

Fisher JC, Holloman JH. A hypothesis for the origin of cancer foci. Cancer 1951;4:916-8.

Fisher RA, Tippett LHC. Limiting forms of the frequency distribution of the largest or smallest number of a sample. Proc Cambridge Phil Soc 1928;24:180-90.

Fréchet M. Sur la loi probabilité de l'écart maximum. Ann de la Soc polonaise de Math (Cracow) 1927;6:93-116.

Freedman DA, Navidi W. On the risk of lung cancer for ex-smokers. Technical Report No 135, University of California, 1987.

Freedman DA, Navidi W. Ex-smokers and the multistage model for lung cancer. Epidemiology 1990;1:21-9.

Gaffney M, Altshuler B. Examination of the role of cigarette smoke in lung carcinogenesis using multistage models. JNCI 1988;80:925-31.

Gumbel EJ. Statistics of extremes. Columbia University Press, New York, 1958.

Hakama M. Epidemiologic evidence for multi-stage theory of carcinogenesis. Int J Cancer 1971;7:557-64.

Hakulinen T, Pukkala E. Future incidence of lung cancer: forecasts based on hypothetical changes in the smoking habits of males. Int J Epidemiol 1981;10:233-40.

Halpern MT, Gillespie BW, Warner KW. Patterns of absolute risk of lung cancer mortality in former smokers. JNCI 1993;85:457-64.

Hammond EC. Smoking in relation to the death rates of one million men and women. In: Haenszel W, editor. Epidemiological approaches to the study of cancer and other chronic diseases. Natl Cancer Inst Monogr 1966; 19:127-204.

Hammond EC, Selikoff IJ, Seidman H. Asbestos exposure, cigarette smoking and death rates. Ann NY Acad Sci 1979;330:473-90.

Hecker E. Cocarcinogens of the tumour-promoter type as potential risk factors of cancer in men. A first complete experimental analysis of an etiological model situation and some of its consequences. In: Models, mechanisms and etiology of tumour promotion. IARC Scientific Publications No 56, Lyon, 1984:441-63.

Hegmann KT, Fraser AM, Keaney RP, Moser SE, Nilasena DS, Sedlars M et al. The effect of age at smoking initiation on lung cancer risk. Epidemiol 1993;4:444-8.

International Agency for Research on Cancer. Tobacco smoking. IARC Monographs on the evaluation of the carcinogenic risk of chemicals to humans. Volume 38, IARC, Lyon 1986.

Kahn H. The Dorn study of smoking and mortality among US veterans: report on 8½ years of observation. In: Haenszel W, editor. Epidemiological approaches to the study of cancer and other chronic diseases. Natl Cancer Inst Monogr 1966;19:1-126.

Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci. USA, 1971;68:820-3.

Lee PN. Final analysis - Experiment 1.1.1.9. The effect of stopping painting. Tobacco Research Council document M432, 1974.

Lee PN. Cigarette smoking and lung cancer. A new mathematical model. Tobacco Advisory Council document TA 1243, 1979.

Lee PN. A summary of evidence relating risk of lung cancer to type of cigarette smoked. Unpublished, 1993a.

Lee PN. Epidemiological studies relating family history of lung cancer to risk of the disease. Indoor Environment 1993b;2:129-42.

Lee PN, O'Neill JA. The effect both of time and dose applied on tumour incidence rates in benzopyrene skin painting experiments. Br J Cancer 1971;25:759-70.

Lee PN, Rothwell K, Whitehead JK. Fractionation of mouse skin carcinogens in cigarette smoke condensate. Br J Cancer 1977;35:730-742.

Lijinsky W. Life-span and cancer: the induction time of tumours in diverse animal species treated with nitrosodiethylamine. Carcinogenesis 1993;14:2373-5.

Likhachev A, Anisimov V, Montesano R, editors. Age-related factors in carcinogenesis (IARC Scientific Publication No 58) Lyon, International Agency for Research on Cancer, 1985.

Lubin JH, Blot WJ, Berrino F, Flamant R, Gillis CR, Kunze M, Schmahl D, Visco G. Modifying risk of developing lung cancer by changing habits of cigarette smoking. BMJ 1984;288:1953-56.

Major IR, Mole RH. Myeloid leukaemia in X-ray irradiated mice. Nature 1978;272:455-6.

Mazumdar S, Redmond CK, Costantino JP, Patwardhan RN, Zhou SYJ. Recent developments in the multistage modelling of cohort data for carcinogenic risk assessment. Environ Health Perspec 1991;90:271-7.

Moolgavkar SH. The multistage theory of carcinogenesis (Letter). Int J Cancer 1977;19:730.

Nordling CO. A new theory on the cancer-inducing mechanism. Br J Cancer 1953;1:68-72.

Parish SE. Exploiting animal tumour data using multistage models. D Phil Thesis, University of Oxford, 1981.

Passey RD. Some problems of lung cancer. Lancet 1962;ii:107-12.

Peto J. Early and late-stage carcinogenesis in mouse skin and man. In: Models, mechanisms and etiology of tumour promotion. IARC Scientific Publications No 56, Lyon, 1984:359-71.

Peto J, Seidman H, Selikoff IJ. Mesothelioma mortality in asbestos workers: Implications for models of carcinogenesis and risk assessment. Br J Cancer 1982;45:124-135.

Peto R. Epidemiology, multistage models and short-term mutagenicity tests. In: Hyatt H, Watson J, jWnston S, editors. Origins of human cancer. Cold Spring Harbor NY: Cold Spring Harbor Laboratory, 1977:1403-28.

Peto R, Parish SE, Gray RG. There is no such thing as ageing, and cancer is not related to it. IARC Scientific Publications No 58, Lyon, 1985:43-53.

Peto R, Roe FJC, Lee PN, Levy L, Clack J. Cancer and aging in mice and men. Br J Cancer 1975;32:411-26.

Peto R, Doll R. Comment on the letter by Moolgavkar. Int J Cancer 1977;19:731.

Pike MC, Doll R. Age at onset of lung cancer: its significance in relation to the effects of smoking. Lancet 1965;1:665-8.

Pike MC. A method of analysis of a certain class of experiments in carcinogenesis. Biometrics 1966;22:142-161.

Renan MJ. How many mutations are required for tumorigenesis? Implications from human cancer data. Molecular Carcinogenesis 1993;7:139-49.

Rogot E. Smoking and mortality among US veterans. J Chronic Dis 1974;27:189-203.

Seidman H. Age at exposure versus years of exposure. Natl Cancer Inst Monogr 1985;67:205-9.

Selikoff IJ, Hammond EC. Multiple risk factors in environmental cancer. In "Persons at high risk of cancer: An approach to cancer etiology and control" (ed J.F.Fraumeni, Jr.), Academic Press, New York 1975, 467-83.

Sellers TA, Bailey-Wilson JE, Elston RC, Wilson AF, Elston GZ, Ooi WL, Rothschild H. Evidence for Mendelian inheritance in the pathogenesis of lung cancer. J Natl Cancer Inst 1990;82:1272-9.

Sobue T, Yamaguchi N, Suzuki T, Fujimoto I, Matsuda M, Doli O et al. Lung cancer incidence rate for male ex-smokers according to age at cessation of smoking. Jpn J Cancer Res 1993;84:601-7.

Stenbäck F, Peto R, Shubik P. Initiation and promotion at different ages and doses in 2200 mice. II Decrease in promotion by TPA with ageing. Br J Cancer 1981;44:15-23.

Swartz JB. Use of a multistage model to predict time trends in smoking induced lung cancer. J Epidemiol Community Health 1992;46:311-5.

Szende B, Kendrey G, Lapis K, Lee PN, Roe FJC. Accuracy of admission and pre-autopsy clinical diagnosis in the light of autopsy findings: a study conducted in Budapest. Human and Experimental Toxicology 1994, in press.

Townsend J. Smoking and lung cancer. A cohort study of men and women in England and Wales 1935-1970. J Royal Statistic Soc A 1978;141:95-107.

US Surgeon General. Smoking and health. A Report of the Surgeon General. US Department of Health, Education and Welfare. Publication No (PHS) 79-50066, 1979.

US Surgeon General. The health consequences of smoking. Cancer. A report of the Surgeon General. US Department of Health and Human Services. Publication No (PHS) 82-50179, 1982.

US Surgeon General. Reducing the health consequences of smoking. 25 years of progress. A report of the Surgeon General. US Department of Health and Human Services. Publication No (CDC) 89-8411, 1989.

Whittemore AS. Effect of cigarette smoking in epidemiological studies of lung cancer. Stat Med 1988;7:223-8.

TABLE 1

Observed male lung cancer death rates per 100,000 per year (numbers of deaths) in relation to age, age of starting and number of cigarettes smoked (from Kahn, 1966)

	Age of starting to smoke			
Age	<15	15-19	20-24	25+
All cigarette smokers				
55-64	251 (70)	168 (293)	99 (133)	53 (30)
65-74	478 (65)	350 (259)	241 (138)	162 (70)
1-9 cigs/day				
55-64	NE* (1)	27 (5)	42 (6)	15 (2)
65-74	NE (2)	108 (7)	99 (8)	52 (5)
10-20 cigs/day				
55-64	156 (16)	118 (81)	78 (47)	43 (13)
65-74	321 (17)	322 (100)	186 (54)	152 (29)
31-39 cigs/day				
55-64	323 (32)	217 (133)	135 (55)	58 (10)
65-74	744 (30)	435 (89)	363 (49)	282 (25)
>39 cigs/day				
55-64	366 (15)	341 (49)	177 (14)	182 (3)
65-74	NE (12)	578 (32)	NE (16)	296 (6)

*NE: rate not estimated

TABLE 2

Fit of a fourth power law relationship of duration of smoking to risk of lung cancer (using data of Table 1 for all cigarette smokers)

Age of start	Age	Duration	Duration^4 (divided by 10^6)	Population (scaled)	Deaths observed	Deaths expected
25+	55-64	33	1.19	0.566	30	31.2
20-24	55-64	38	2.08	1.343	133	129.6
15-19	55-64	43	3.42	1.744	293	276.6
25+	65-74	43	3.42	0.432	70	68.5
<15	55-64	48	5.31	0.279	70	68.7
20-24	65-74	48	5.31	0.573	138	141.1
15-19	65-74	53	7.89	0.740	259	270.8
<15	65-74	58	11.32	0.136	65	71.4
Total					1058	1058.0

NB. Scaled population estimated by deaths/rate per 100,000 per year
Expected deaths calculated by multiplying population x duration^4 x scaling factor
Scaling factor = Σ observed deaths / Σ(population x duration^4).

TABLE 3

Dose relationships under various hypotheses

Hypothesis A - equal effects on stages 1 and 6

Dose (proportional to numbers of cigarettes smoked)	Stage effects α	Stage effects δ	Relative Risk at age 70-74	Linear fit
0	1	1	1176	1176
1	2	2	2665	3446
2	3	3	4466	5715
4	5	5	9005	10255
6	7	7	14794	14794
8	9	9	21833	19333
10	11	11	30122	23873

Hypothesis B - greater effect on stage 6 than stage 1

Dose	Stage effects α	Stage effects δ	Relative risk at age 70-74	Linear fit
0	1	1	1176	1176
1	1.25	3.875	4708	5270
2	1.5	6.75	8465	9363
4	2	12.5	16652	17550
6	2.5	18.25	25737	25737
8	3	24	35721	33924
10	3.5	29.75	46603	42111

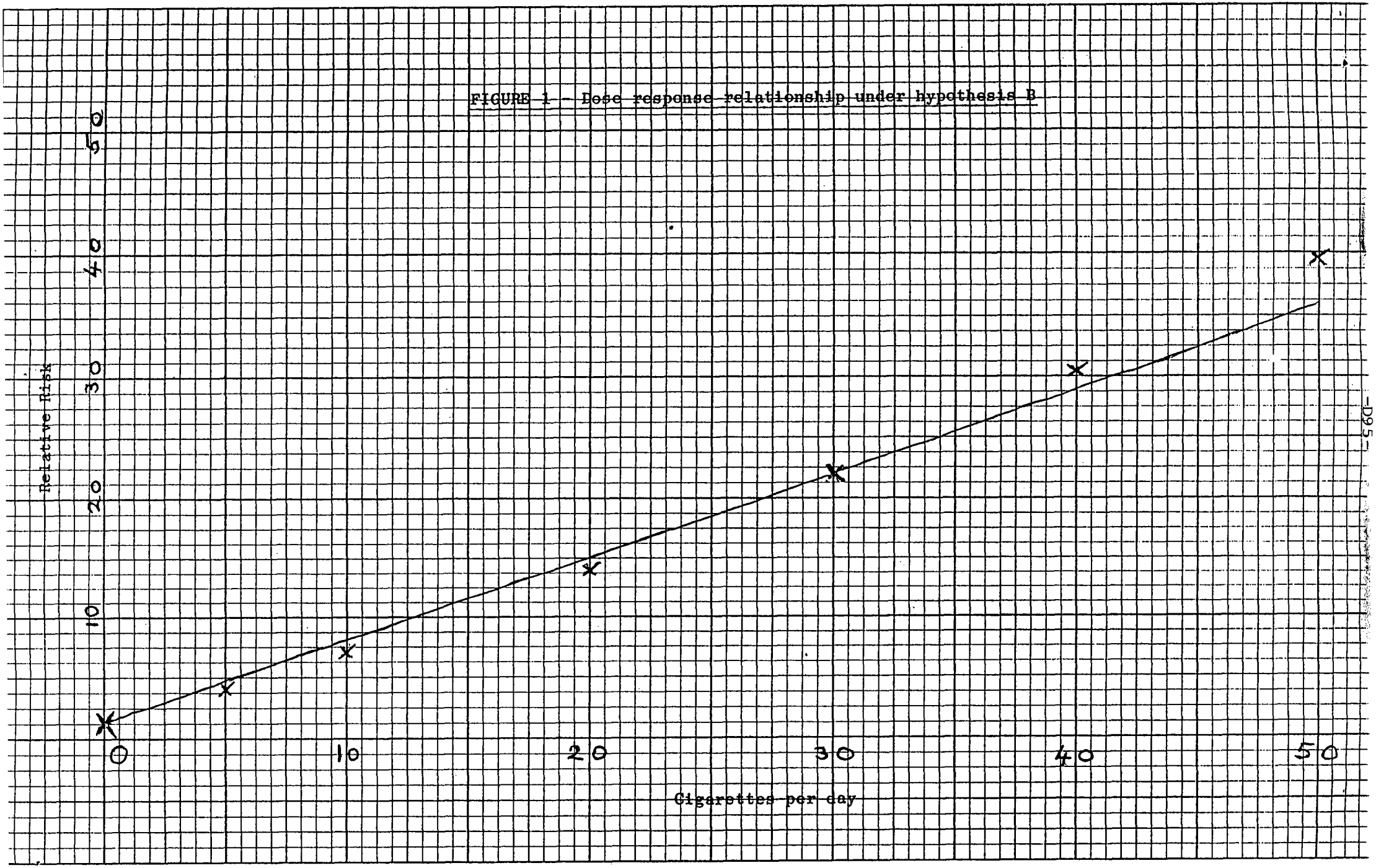
TABLE 4

Lung cancer incidence rates per 100,000 per year (numbers of deaths) in relation to age, and time of giving up smoking (from Halpern et al, 1993)

Smoking habits	Age							
	40-43	44-48	49-53	54-58	59-63	64-68	69-73	74-80
Never smoker	0.000* (0)	3.62 (9)	4.69 (20)	6.93 (33)	13.28 (61)	18.99 (75)	31.23 (91)	39.48 (93)
Current smoker	10.72 (5)	45.75 (62)	82.24 (195)	156.8 (398)	272.0 (592)	430.9 (622)	643.0 (518)	858.7 (332)
Former smoker Age at cessation								
30-39	-	7.73 (4)	18.46 (18)	27.70 (27)	19.29 (13)	57.39 (22)	68.49 (14)	42.76 (4)
40-49	-	-	-	52.21 (53)	73.59 (74)	106.8 (72)	109.2 (30)	114.4 (20)
50-54	-	-	-	-	134.8 (66)	133.8 (54)	170.9 (45)	241.5 (33)
55-59	-	-	-	-	-	244.0 (89)	270.5 (64)	353.6 (48)
60-64	-	-	-	-	-	-	409.2 (100)	607.4 (97)
65-69	-	-	-	-	-	-	-	724.8 (91)

*Based on 82,335 person years

FIGURE 1 - Dose response relationship under hypothesis B



APPENDIX E

10 year percentage change in US Observed Lung Cancer risk and in Predicted risk estimates using different smoking models and alternative data sources.

Notes.
1. Lung cancer rate (Observed, and Observed-background) is repeated on each page for convenience. See section 3.3.1 for definition of Background. See sections 3.3, 3.4 for definitions of risk estimates and smoking indicies.
2. Smoking model - S = Swartz, T = Townsend, blank indicates that result is independent of model. BASIC model has F = 15, N = 20, D = 0, K-1 = 4.5, L = 5, where:
   F = first year of smoking
   N = number of cigarettes per smoker per day
   D = drift
   K-1 = power in multistage calculations
   L = lag (years)
   Other models vary one of these parameters. See sections 3.2, 3.5 for more details
3. Data source is Harris, except for those marked INTSS.
4. INTSS uses BASIC model. (a) and (b) refers to method of extending cohorts, see section 7.3.3 (not relevant to 45-54, 1976-85).

Sex	Male								Female							
Age	45-54			55-64			65-74		45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate																
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Absolute risk estimates																
Swartz 1 Brit Docs																
S BASIC	9.3	-1.9	-13.3	19.5	5.7	-6.4	16.3	1.0	50.8	13.2	1.4	64.7	47.5	8.0	65.4	40.9
S F18	9.1	-2.0	-13.3	17.7	5.5	-6.4	14.7	0.9	48.4	12.3	0.7	63.5	45.3	7.2	64.2	39.0
S F21	8.4	-2.6	-13.3	16.0	5.0	-7.0	13.2	0.4	45.0	11.4	-0.2	60.9	42.0	6.4	61.7	36.1
S N30	10.1	-1.8	-14.0	21.6	6.2	-6.5	18.1	1.2	61.2	15.2	2.2	82.3	56.9	9.4	82.7	48.6
S N40	10.6	-1.6	-14.3	23.0	6.5	-6.6	19.2	1.3	68.8	16.6	2.8	94.9	63.4	10.4	94.8	53.7
S D005	9.2	-2.0	-13.2	19.0	5.6	-6.4	15.7	1.0	49.8	12.9	1.2	63.3	46.1	7.7	63.3	39.3
S INTSS(a)		-1.8	-10.4			-7.5				8.2	-4.3			9.5		
S INTSS(b)		0.5				-5.6				11.1				12.7		
Swartz 1 US Vets																
S BASIC	8.1	-2.0	-12.1	16.6	5.0	-6.0	13.8	0.8	39.1	10.7	0.7	46.2	36.6	6.3	46.7	31.6
S F18	8.0	-2.1	-12.0	15.3	4.8	-6.0	12.6	0.7	37.6	10.1	0.2	45.5	35.2	5.7	45.9	30.3
S F21	7.5	-2.5	-12.0	14.1	4.4	-6.4	11.4	0.4	35.4	9.5	-0.4	44.0	33.0	5.1	44.4	28.3
S N30	9.1	-2.0	-13.1	19.0	5.6	-6.4	15.9	1.0	48.9	12.8	1.2	61.6	45.8	7.7	62.3	39.4
S N40	9.7	-1.9	-13.7	20.6	6.0	-6.5	17.3	1.1	56.2	14.2	1.8	73.8	52.4	8.7	74.4	45.0
S D005	8.1	-2.0	-12.0	16.2	4.9	-6.0	13.3	0.8	38.5	10.5	0.5	45.3	35.6	6.0	45.3	30.4
S INTSS(a)		-1.8	-9.5			-6.9				5.1	-3.6			6.0		
S INTSS(b)		-0.2				-5.8				6.8				8.0		
Swartz 2 Brit Docs																
BASIC	10.1	0.5	-9.2	19.4	8.2	-2.3	17.6	6.0	50.9	14.7	4.0	67.0	48.5	11.3	70.6	44.6
F18	9.9	0.3	-9.6	18.5	8.1	-2.5	16.9	5.9	49.6	14.2	3.5	66.5	47.5	10.9	70.2	43.8
F21	9.4	-0.4	-10.1	17.6	7.6	-3.1	16.0	5.4	47.0	13.4	2.6	65.1	45.6	10.2	69.0	42.3
N30	10.5	0.5	-9.6	20.2	8.5	-2.3	18.2	6.2	57.4	15.8	4.3	78.7	53.0	12.0	80.4	47.9
N40	10.8	0.5	-9.8	20.7	8.7	-2.4	18.6	6.3	61.3	16.5	4.4	86.1	55.5	12.4	86.4	49.7
D005	10.1	0.5	-9.2	19.4	8.2	-2.3	17.6	6.0	50.9	14.7	4.0	67.0	48.5	11.3	70.6	44.6
INTSS(a)		0.3	-6.8			-2.9				2.3	-2.7			1.2		
INTSS(b)		0.6				-3.1				3.2				1.8		
Swartz 2 US Vets																
BASIC	9.0	0.4	-8.3	17.4	7.5	-2.1	16.1	5.6	38.9	12.2	3.4	47.7	39.4	9.8	53.0	37.7
F18	8.8	0.3	-8.6	16.6	7.4	-2.3	15.4	5.4	37.8	11.8	3.0	47.2	38.6	9.5	52.6	36.9
F21	8.3	-0.3	-9.0	15.7	6.9	-2.8	14.6	5.0	35.6	11.1	2.2	46.1	36.9	8.8	51.6	35.6
N30	9.7	0.5	-8.9	18.8	8.0	-2.2	17.1	5.9	46.6	13.8	3.8	59.7	45.3	10.8	64.2	42.3
N40	10.1	0.5	-9.3	19.5	8.3	-2.3	17.7	6.0	51.7	14.8	4.0	68.3	49.0	11.4	71.7	45.0
D005	9.0	0.4	-8.3	17.4	7.5	-2.1	16.1	5.6	38.9	12.2	3.4	47.7	39.4	9.8	53.0	37.7
INTSS(a)		0.3	-6.1			-2.6				2.0	-2.3			1.1		
INTSS(b)		0.6				-2.8				2.7				1.6		

Sex	Male								Female							
Age	45-54			55-64			65-74		45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate																
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Excess risk estimates																
Duration **k-1																
S BASIC	13.3	-0.1	-15.2	29.4	8.0	-6.0	24.0	1.9	128.7	26.2	8.6	166.0	101.5	16.1	151.0	77.8
S F18	13.7	0.2	-15.7	25.3	8.0	-6.0	20.8	1.7	119.9	24.1	6.9	162.3	95.3	14.4	147.9	73.3
S F21	12.6	-1.6	-16.4	21.4	6.9	-7.5	17.7	0.7	107.1	22.3	4.4	151.2	86.0	12.7	139.5	66.2
S K3	12.6	-0.1	-14.4	26.1	8.1	-5.4	21.8	3.0	106.4	23.5	7.1	149.0	85.4	14.6	136.1	67.5
S K6	13.8	-0.3	-15.4	32.6	7.9	-6.2	26.3	1.1	149.4	28.5	9.8	179.7	116.4	17.5	163.1	87.8
S LO	11.0	-3.5	-17.6	26.9	5.2	-7.8	20.5	-1.3	115.3	21.0	3.7	159.6	89.5	12.8	140.7	70.8
S D005	13.3	-0.1	-15.2	29.2	8.0	-6.0	23.8	1.9	127.6	26.1	8.6	165.0	100.4	16.0	149.9	76.9
T BASIC	14.4	2.9	-9.9	29.9	9.9	-3.4	23.7	3.5	129.2	29.9	13.3	158.8	106.7	22.2	145.1	85.3
T F18	14.6	2.5	-12.5	25.4	9.2	-4.9	19.5	1.9	120.4	27.8	10.9	155.1	100.4	20.0	142.0	80.0
T F21	13.1	-1.0	-16.5	20.8	6.6	-8.9	14.7	-0.8	107.6	26.0	6.9	144.1	90.9	17.2	133.6	71.5
T K3	13.1	1.6	-11.1	26.0	8.9	-4.0	21.0	3.7	106.8	25.7	9.7	144.0	88.4	17.9	130.8	71.0
T K6	15.4	4.0	-8.5	33.8	10.9	-2.4	26.8	3.7	150.0	33.2	16.2	170.7	123.6	26.0	157.1	99.6
T LO	12.6	-0.4	-13.0	26.9	7.1	-5.9	19.9	-0.2	117.5	26.6	8.4	152.3	96.8	17.9	136.6	76.2
S INTSS(a)		-0.9	-11.8			-7.5				40.9	-7.2			34.5		
S INTSS(b)		7.8				-1.0				61.7				49.7		
T INTSS(a)		0.1	-8.1			-6.8				25.0	-6.6			16.0		
T INTSS(b)		2.7				-6.4				37.8				23.6		
Multistage 1:0																
S BASIC	15.2	5.9	-4.2	32.0	14.0	4.5	29.1	12.9	129.2	30.4	16.6	158.8	107.6	26.6	146.9	93.7
S F18	15.6	6.4	-5.1	27.9	14.2	4.6	25.9	13.0	120.4	28.3	15.0	155.1	101.4	24.9	143.9	89.0
S F21	14.6	4.8	-6.6	24.1	13.3	3.1	22.9	12.1	107.6	26.7	12.6	144.1	92.0	23.4	135.7	81.7
S K3	14.0	4.5	-5.8	28.0	12.7	2.7	25.7	11.4	106.8	26.5	13.2	144.0	89.7	22.6	133.2	78.7
S K6	16.0	6.7	-3.2	35.7	14.9	5.6	32.1	13.9	150.0	33.5	19.0	170.7	124.2	29.6	158.3	107.4
S LO	14.6	5.3	-4.9	30.5	13.6	3.9	28.1	12.5	117.5	28.4	15.0	152.7	100.5	25.1	142.4	88.9
S D005	15.2	5.9	-4.2	32.0	14.0	4.5	29.1	12.9	129.2	30.4	16.6	158.8	107.6	26.6	146.9	93.7
T BASIC	15.2	5.9	-4.2	32.0	14.0	4.5	29.1	12.9	129.2	30.4	16.6	158.8	107.6	26.6	146.9	93.7
T F18	15.6	6.4	-5.1	27.9	14.2	4.6	25.9	13.0	120.4	28.3	15.0	155.1	101.4	24.9	143.9	89.0
T F21	14.6	4.8	-6.6	24.1	13.3	3.1	22.9	12.1	107.6	26.7	12.6	144.1	92.0	23.4	135.7	81.7
T K3	14.0	4.5	-5.8	28.0	12.7	2.7	25.7	11.4	106.8	26.5	13.2	144.0	89.7	22.6	133.2	78.7
T K6	16.0	6.7	-3.2	35.7	14.9	5.6	32.1	13.9	150.0	33.5	19.0	170.7	124.2	29.6	158.3	107.4
T LO	14.6	5.3	-4.9	30.5	13.6	3.9	28.1	12.5	117.5	28.4	15.0	152.7	100.5	25.1	142.4	88.9
S INTSS(a)		3.3	-2.0			3.0				21.1	-6.2			13.4		
S INTSS(b)		3.6				3.3				29.9				18.5		
T INTSS(a)		3.3	-2.0			3.0				21.1	-6.2			13.4		
T INTSS(b)		3.6				3.3				29.9				18.5		
Multistage 5:1																
S BASIC	11.8	0.0	-11.9	25.6	8.9	-3.1	22.7	5.9	84.3	19.7	5.3	129.6	75.7	14.7	123.3	66.5
S F18	11.6	-0.2	-12.6	22.8	8.8	-3.4	20.3	5.6	79.6	18.2	3.9	127.3	71.8	13.3	121.2	63.2
S F21	10.8	-1.5	-13.6	20.3	7.8	-4.7	18.0	4.6	72.9	16.7	2.1	121.8	65.9	11.8	116.2	58.1
S K3	12.3	1.0	-10.9	25.0	9.5	-2.3	22.1	6.6	89.1	21.3	7.1	132.6	76.1	15.7	123.6	65.1
S K6	11.0	-1.4	-13.3	25.0	8.0	-4.3	22.4	4.7	75.6	17.0	2.6	122.8	70.3	12.4	118.3	63.7
S LO	10.6	-1.8	-13.2	24.3	7.5	-3.8	21.0	4.6	80.4	17.1	3.1	126.9	71.5	13.2	119.3	63.9
S D005	11.8	0.0	-11.8	25.4	9.0	-3.0	22.6	6.0	83.5	19.5	5.2	128.7	74.9	14.6	122.2	65.9
T BASIC	12.1	0.8	-10.5	25.8	9.5	-2.3	22.6	6.2	85.0	20.5	6.4	128.4	77.0	16.4	122.3	68.9
T F18	11.9	0.3	-11.9	22.9	9.1	-3.1	19.9	5.5	80.2	18.9	4.8	126.1	73.0	14.8	120.2	65.3
T F21	10.9	-1.4	-13.6	20.1	7.7	-5.1	17.1	4.1	73.6	17.2	2.5	120.7	66.9	12.9	115.2	59.7
T K3	12.4	1.5	-9.8	24.9	9.8	-1.9	21.7	6.7	89.6	22.0	7.9	131.2	77.1	16.9	122.0	66.4
T K6	11.3	-0.6	-12.0	25.4	8.7	-3.4	22.7	5.3	76.2	17.6	3.6	122.0	71.5	14.1	118.0	66.6
T LO	11.1	-0.9	-12.0	24.4	8.1	-3.3	20.8	4.7	80.9	18.5	4.3	125.6	73.5	14.7	118.7	65.8
S INTSS(a)		-0.6	-9.1			-4.4				13.5	-5.5			12.5		
S INTSS(b)		1.6				-2.8				18.5				17.1		
T INTSS(a)		-0.4	-8.2			-4.3				10.5	-5.6			8.3		
T INTSS(b)		-0.1				-4.6				14.7				11.6		

Sex	Male								Female							
Age	45-54			55-64			65-74		45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate																
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Excess risk estimates (cont)																
Multistage 1:1																
S BASIC	10.1	-2.1	-13.8	20.8	6.7	-6.0	17.8	2.9	67.5	15.1	0.9	113.7	58.0	9.1	105.6	49.6
S F18	10.0	-2.3	-14.2	19.6	6.5	-6.2	16.6	2.7	65.9	14.6	0.4	112.9	56.5	8.5	104.7	48.2
S F21	9.6	-2.9	-14.6	18.6	6.1	-6.9	15.5	2.1	63.7	13.9	-0.4	111.0	54.3	7.8	102.8	46.0
S K3	10.9	-0.7	-12.2	21.8	7.9	-4.3	19.0	4.7	73.5	17.3	3.4	118.6	62.7	11.7	110.1	53.7
S K6	9.5	-3.3	-15.1	19.7	5.6	-7.5	16.4	1.2	63.3	13.4	-1.1	110.4	54.0	6.9	101.7	45.5
S L0	8.7	-4.3	-15.6	19.4	5.0	-7.0	16.0	1.2	62.6	12.0	-1.8	109.7	54.0	7.4	101.3	47.4
S D005	10.1	-2.1	-13.8	20.7	6.7	-6.0	17.6	2.9	67.2	15.0	0.9	113.3	57.6	9.1	105.0	49.1
T BASIC	10.2	-1.8	-13.3	20.9	7.0	-5.6	17.8	3.1	68.2	15.4	1.3	113.4	58.4	9.8	105.3	50.6
T F18	10.1	-2.1	-13.9	19.6	6.7	-6.1	16.5	2.7	66.6	14.8	0.6	112.6	56.9	9.1	104.5	49.0
T F21	9.6	-2.8	-14.7	18.5	6.0	-7.1	15.1	1.9	64.4	14.1	-0.3	110.7	54.6	8.2	102.6	46.6
T K3	11.0	-0.5	-11.7	21.8	8.1	-4.0	18.8	4.8	74.2	17.6	3.8	118.0	63.1	12.3	109.4	54.4
T K6	9.6	-3.0	-14.8	19.8	5.9	-7.1	16.6	1.5	64.0	13.5	-0.8	110.2	54.3	7.5	101.7	46.5
T L0	8.8	-4.0	-15.1	19.5	5.2	-6.8	16.0	1.3	62.8	12.4	-1.4	109.4	54.7	8.0	101.3	48.4
S INTSS(a)		-1.9	-11.0			-7.2				3.9	-4.6			4.0		
S INTSS(b)		-1.2				-7.1				5.3				5.4		
T INTSS(a)		-1.8	-10.7			-7.1				3.1	-4.6			2.5		
T INTSS(b)		-1.8				-7.8				4.3				3.6		
Multistage 1:2																
S BASIC	9.9	-2.5	-14.3	20.1	6.2	-6.8	16.8	1.9	66.0	14.6	0.3	112.2	56.0	8.2	103.6	47.0
S F18	9.8	-2.6	-14.6	19.1	6.1	-7.0	15.8	1.7	64.8	14.1	-0.1	111.6	54.8	7.7	102.9	45.9
S F21	9.5	-3.1	-14.9	18.3	5.7	-7.5	14.9	1.3	63.0	13.7	-0.7	110.1	53.0	7.2	101.4	44.1
S K3	10.6	-1.2	-12.9	21.2	7.4	-5.1	18.1	3.8	71.5	16.7	2.7	116.8	60.5	10.6	108.1	51.2
S K6	9.3	-3.5	-15.5	19.1	5.2	-8.1	15.4	0.3	62.4	13.1	-1.5	109.4	52.5	6.2	100.2	43.3
S L0	8.3	-4.9	-16.4	18.5	4.2	-8.1	14.7	-0.2	60.9	11.2	-2.7	108.0	51.7	6.2	98.9	44.5
S D005	9.9	-2.5	-14.3	19.9	6.2	-6.8	16.6	1.9	65.7	14.5	0.3	111.9	55.5	8.1	103.0	46.4
T BASIC	10.0	-2.2	-13.8	20.2	6.5	-6.4	16.8	2.2	66.7	14.8	0.7	111.9	56.4	8.9	103.4	48.0
T F18	9.9	-2.4	-14.3	19.2	6.2	-6.8	15.7	1.8	65.5	14.4	0.2	111.3	55.2	8.3	102.7	46.8
T F21	9.5	-3.0	-14.9	18.3	5.7	-7.6	14.6	1.1	63.7	13.8	-0.6	109.8	53.4	7.6	101.2	44.8
T K3	10.7	-0.9	-12.3	21.2	7.6	-4.8	18.0	3.9	72.3	17.0	3.1	116.3	61.0	11.3	107.4	52.0
T K6	9.4	-3.3	-15.1	19.2	5.5	-7.8	15.7	0.6	63.1	13.2	-1.2	109.2	52.9	6.8	100.2	44.4
T L0	8.5	-4.5	-15.8	18.6	4.5	-7.7	14.7	0.0	61.0	11.7	-2.2	107.7	52.5	6.9	99.0	45.5
S INTSS(a)		-2.2	-11.5			-8.0				3.2	-4.5			3.3		
S INTSS(b)		-1.5				-7.9				4.2				4.5		
T INTSS(a)		-2.1	-11.1			-7.9				2.4	-4.6			1.7		
T INTSS(b)		-2.0				-8.6				3.2				2.6		
Multistage 1:2E																
S BASIC	9.8	-2.8	-14.9	19.9	5.7	-7.6	16.2	0.8	66.0	14.4	0.0	112.4	55.7	7.6	103.6	45.9
S F18	9.7	-2.9	-15.0	19.0	5.7	-7.7	15.3	0.7	64.8	14.0	-0.4	111.8	54.6	7.2	102.9	44.9
S F21	9.4	-3.2	-15.2	18.2	5.4	-8.0	14.5	0.4	63.0	13.6	-1.0	110.3	52.8	6.7	101.5	43.3
S K3	10.5	-1.6	-13.7	20.9	6.9	-6.1	17.5	2.6	71.5	16.5	2.2	117.1	60.2	9.9	108.2	50.0
S K6	9.3	-3.7	-15.9	18.9	4.8	-8.7	14.9	-0.7	62.4	13.0	-1.7	109.5	52.4	5.8	100.2	42.5
S L0	8.1	-5.5	-17.2	18.2	3.5	-9.2	13.7	-1.9	60.8	10.9	-3.3	108.2	51.1	5.3	98.6	42.9
S D005	9.7	-2.8	-14.9	19.6	5.7	-7.6	15.9	0.7	65.6	14.3	-0.1	111.8	55.0	7.4	102.7	45.1
T BASIC	9.9	-2.3	-14.1	20.1	6.2	-7.0	16.3	1.2	66.7	14.8	0.6	111.9	56.4	8.6	103.3	47.5
T F18	9.8	-2.6	-14.6	19.0	5.9	-7.5	15.1	0.7	65.5	14.4	0.0	111.3	55.2	8.0	102.6	46.3
T F21	9.5	-3.2	-15.2	18.1	5.4	-8.2	14.0	0.1	63.7	13.8	-0.7	109.8	53.3	7.3	101.1	44.3
T K3	10.6	-1.2	-12.8	20.9	7.2	-5.6	17.3	2.9	72.3	17.0	2.8	116.3	60.9	10.8	107.2	51.2
T K6	9.4	-3.4	-15.3	19.2	5.3	-8.2	15.3	-0.2	63.1	13.2	-1.2	109.2	52.9	6.7	100.2	44.1
T L0	8.4	-4.9	-16.4	18.3	4.0	-8.7	13.7	-1.5	61.0	11.7	-2.5	107.7	52.3	6.4	98.6	44.6
S INTSS(a)		-2.4	-12.0			-8.8				3.6	-4.6			4.1		
S INTSS(b)		-1.3				-8.4				4.8				5.5		
T INTSS(a)		-2.3	-11.4			-8.7				2.4	-4.6			1.7		
T INTSS(b)		-2.2				-9.4				3.3				2.6		

Sex	Male								Female							
Age	45-54			55-64			65-74		45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate																
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Excess risk estimates (cont)																
Multistage 1:5																
S BASIC	9.7	-2.7	-14.7	19.6	5.8	-7.3	16.0	1.2	65.1	14.3	-0.1	111.3	54.7	7.6	102.3	45.2
S F18	9.7	-2.8	-14.8	18.8	5.8	-7.4	15.3	1.1	64.0	13.9	-0.4	110.8	53.7	7.2	101.8	44.3
S F21	9.4	-3.2	-15.0	18.1	5.5	-7.8	14.6	0.8	62.5	13.5	-0.9	109.6	52.3	6.8	100.5	42.9
S K3	10.4	-1.5	-13.3	20.7	7.0	-5.7	17.5	3.1	70.2	16.2	2.1	115.7	59.1	9.9	106.7	49.4
S K6	9.2	-3.7	-15.7	18.7	4.9	-8.6	14.8	-0.3	61.9	12.8	-1.7	108.9	51.6	5.8	99.3	41.9
S LO	8.1	-5.3	-16.9	17.9	3.7	-8.8	13.7	-1.2	59.8	10.8	-3.3	107.0	50.2	5.4	97.4	42.5
S D005	9.7	-2.8	-14.7	19.5	5.8	-7.3	15.8	1.2	64.8	14.2	-0.1	110.9	54.2	7.4	101.7	44.7
T BASIC	9.8	-2.4	-14.1	19.8	6.1	-6.9	16.1	1.5	65.8	14.5	0.3	111.0	55.2	8.3	102.1	46.4
T F18	9.7	-2.6	-14.5	18.9	5.9	-7.3	15.2	1.1	64.8	14.1	-0.1	110.5	54.2	7.8	101.5	45.3
T F21	9.4	-3.2	-15.1	18.1	5.5	-7.9	14.2	0.6	63.2	13.6	-0.8	109.3	52.6	7.2	100.3	43.6
T K3	10.5	-1.2	-12.7	20.7	7.2	-5.3	17.4	3.3	71.0	16.6	2.6	115.1	59.6	10.6	106.0	50.3
T K6	9.3	-3.5	-15.3	18.8	5.2	-8.2	15.0	0.0	62.5	13.0	-1.4	108.7	52.0	6.3	99.3	43.0
T LO	8.3	-4.9	-16.3	18.1	4.1	-8.4	13.8	-1.0	59.9	11.3	-2.7	106.6	51.0	6.2	97.5	43.6
S INTSS(a)		-2.4	-11.8			-8.5				2.7	-4.5			2.8		
S INTSS(b)		-1.6				-8.5				3.6				3.8		
T INTSS(a)		-2.3	-11.4			-8.4				1.8	-4.5			1.2		
T INTSS(b)		-2.2				-9.2				2.6				1.9		
Multistage 0:1																
S BASIC	9.3	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4	61.1	13.1	-1.1	107.0	49.5	6.0	96.4	39.5
S F18	9.2	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4	61.0	13.1	-1.2	106.9	49.5	6.0	96.4	39.5
S F21	9.2	-3.4	-15.1	17.7	5.1	-8.1	13.8	0.3	60.6	12.9	-1.4	106.8	49.3	5.9	96.3	39.4
S K3	9.8	-2.1	-13.4	18.8	6.4	-6.3	15.4	2.4	64.0	14.5	0.7	109.1	52.4	8.1	99.1	43.1
S K6	8.9	-4.0	-15.9	17.3	4.4	-9.1	12.9	-1.0	59.6	12.2	-2.3	106.2	48.1	4.7	95.3	37.6
S LO	7.5	-5.9	-17.2	15.9	3.0	-9.6	11.5	-2.1	55.0	9.4	-4.5	101.7	44.6	3.7	91.1	36.5
S D005	9.3	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4	61.1	13.1	-1.1	107.0	49.5	6.0	96.4	39.5
T BASIC	9.3	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4	61.8	13.1	-1.1	107.0	49.5	6.0	96.4	39.5
T F18	9.2	-3.2	-14.9	17.8	5.2	-8.0	13.9	0.4	61.7	13.1	-1.2	106.9	49.5	6.0	96.4	39.5
T F21	9.2	-3.4	-15.1	17.7	5.1	-8.1	13.8	0.3	61.3	12.9	-1.4	106.8	49.3	5.9	96.3	39.4
T K3	9.8	-2.1	-13.4	18.8	6.4	-6.3	15.4	2.4	64.9	14.5	0.7	109.1	52.4	8.1	99.1	43.1
T K6	8.9	-4.0	-15.9	17.3	4.4	-9.1	12.9	-1.0	60.3	12.2	-2.3	106.2	48.1	4.7	95.3	37.6
T LO	7.5	-5.9	-17.2	15.9	3.0	-9.6	11.5	-2.1	55.0	9.4	-4.5	101.7	44.6	3.7	91.1	36.5
S INTSS(a)		-2.6	-12.0			-9.1				-0.5	-4.1			-1.4		
S INTSS(b)		-2.7				-10.0				-0.7				-1.6		
T INTSS(a)		-2.6	-12.0			-9.1				-0.5	-4.1			-1.4		
T INTSS(b)		-2.7				-10.0				-0.7				-1.6		

Sex	Male								Female							
Age	45-54			55-64			65-74		45-54			55-64			65-74	
Period	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1956 1965	1966 1975	1976 1985	1966 1975	1976 1985
Lung cancer rate																
Observed	27.0	22.5	-9.3	31.5	19.8	7.2	30.1	9.4	93.4	87.4	23.5	50.2	124.4	56.1	95.1	97.8
Obs - 0.5*Background	28.7	23.6	-9.7	33.3	20.6	7.4	31.7	9.8	150.4	108.0	26.0	90.9	170.8	63.3	164.5	121.5
Obs - Background	30.6	24.7	-10.1	35.3	21.5	7.7	33.4	10.1	385.3	141.3	29.1	---	272.5	72.6	---	160.1
Smoking indices																
Av % smkrs lifetime																
BASIC	10.4	-0.6	-11.2	20.9	8.1	-3.4	18.6	5.7	68.5	16.6	3.6	114.0	58.7	11.9	105.6	51.3
F18	10.4	-0.9	-11.6	20.1	8.0	-3.7	17.8	5.5	66.9	16.0	2.9	113.1	57.5	11.4	104.9	50.4
F21	9.8	-1.7	-12.4	19.1	7.5	-4.5	17.0	5.0	64.1	15.2	1.8	111.0	55.5	10.6	103.3	48.7
LO	9.4	-2.1	-12.7	19.8	7.0	-4.4	17.4	4.7	63.4	14.1	1.4	109.8	55.0	10.3	102.1	49.0
INTSS(a)		-0.6	-7.6			-4.0				1.8	-2.5			0.8		
INTSS(b)		-0.4				-4.1				2.5				1.2		
Av % first 10 yrs																
BASIC	14.9	6.0	-4.5	33.8	14.8	5.9	33.4	14.6	121.9	29.0	17.0	178.6	120.5	28.8	178.7	118.7
F18	14.1	5.1	-5.7	26.9	14.0	5.0	26.7	13.9	102.3	26.0	13.8	159.2	101.3	25.8	158.8	99.9
F21	12.7	2.7	-7.1	22.6	12.6	2.6	22.4	12.4	86.4	22.7	10.1	144.3	85.6	22.6	143.7	84.6
LO	14.9	6.0	-4.5	33.8	14.8	5.9	33.4	14.6	121.9	29.0	17.0	178.6	120.5	28.8	178.7	118.7
INTSS(a)		2.9	-0.2			3.1				19.6	-3.3			19.0		
INTSS(b)		1.9				2.1				28.0				27.1		
Av % last 10 yrs																
BASIC	7.2	-6.1	-18.1	15.2	1.6	-12.2	9.2	-5.4	50.7	9.0	-5.2	95.7	40.0	0.5	84.6	28.8
F18	7.2	-6.1	-18.1	15.2	1.6	-12.2	9.2	-5.4	50.7	9.0	-5.2	95.7	40.0	0.5	84.6	28.8
F21	7.2	-6.1	-18.1	15.2	1.6	-12.2	9.2	-5.4	50.7	9.0	-5.2	95.7	40.0	0.5	84.6	28.8
LO	4.9	-9.6	-21.5	12.3	-1.7	-13.7	5.2	-9.4	45.2	4.5	-9.2	90.6	34.6	-1.6	77.6	26.1
INTSS(a)		-5.1	-15.8			-14.7				1.2	-7.3			-2.2		
INTSS(b)		-5.3				-15.8				1.0				-2.4		
% 20 yrs ago																
BASIC	13.8	3.5	-6.7	20.6	9.0	-2.5	16.9	5.2	97.4	25.0	12.5	118.3	58.6	14.1	101.7	44.7
F18	13.8	3.5	-6.7	20.6	9.0	-2.5	16.9	5.2	97.4	25.0	12.5	118.3	58.6	14.1	101.7	44.7
F21	13.7	3.9	-6.6	20.6	9.0	-2.5	16.9	5.2	101.4	25.2	12.6	118.3	58.6	14.1	101.7	44.7
LO	11.3	0.2	-9.2	19.1	7.3	-5.7	15.8	1.7	72.3	19.1	5.2	108.8	50.8	9.3	95.3	40.4
INTSS(a)		11.6	-1.6			-4.2				-15.1	12.6			12.7		
INTSS(b)		15.9				-4.9				-17.1				16.4		
% dur 30+ years																
S BASIC	14.5	-1.0	-13.9	23.0	8.5	-7.2	18.0	4.1	279.4	36.7	12.0	175.3	92.0	14.1	123.3	53.1
S F18	17.4	1.0	-9.0	22.8	7.2	-8.1	17.7	3.0	260.4	38.7	10.3	176.3	91.4	13.4	123.7	52.7
S F21	---	---	---	22.0	6.5	-8.9	16.9	2.0	---	---	---	171.5	94.2	12.3	123.4	51.7
S LO	10.9	-3.2	-20.0	19.0	5.8	-9.4	17.5	3.9	147.8	22.0	5.6	147.1	66.4	8.6	112.9	46.8
S D005	14.6	-1.1	-13.8	23.3	8.5	-7.2	18.2	4.0	279.6	36.7	12.0	175.9	92.9	14.2	124.7	54.2
T BASIC	17.1	8.3	-2.3	20.9	6.2	-8.1	14.6	3.0	279.7	44.8	24.8	167.5	97.0	19.2	114.7	42.9
T F18	21.9	12.1	-4.0	20.9	6.2	-8.1	14.6	3.0	261.6	47.8	24.7	167.5	97.0	19.2	114.7	42.9
T F21	---	---	---	20.3	5.3	-11.0	14.6	3.0	---	---	---	162.6	101.0	18.8	114.7	42.9
T LO	12.8	0.1	-14.5	14.9	3.2	-8.1	14.6	3.0	151.0	31.9	12.7	140.5	64.6	5.7	94.5	42.9
S INTSS(a)		-0.2	-7.5			-3.5				131.3	-6.1			10.3		
S INTSS(b)		22.6				1.7				327.6				15.3		
T INTSS(a)		3.4	-4.2			-15.7				94.6	-10.6			0.8		
T INTSS(b)		4.8				-14.1				217.2				4.7		

Appendix F

Can past prevalences of cigarette smokers be estimated retrospectively? Evidence from the UK Health and Lifestyle Survey

Authors: P N Lee and A Thornton

Date: 20.4.94

Harris (JNCI, 1983, 71, 473-479) estimated percentages of cigarette smokers among successive birth cohorts of men and women in the United States based on smoking histories of respondents to the 1978-80 Health Interview Surveys. In order to gain insight into the validity of this approach, we compared estimates of past percentages of smokers based on smoking histories given by respondents in the 1984/85 UK Health and Lifestyle Survey (HLS) with percentages of smokers reported in surveys carried out by Research Services for ITL from 1948 onwards.

The data from the Research Services surveys are those given in Tables 4.1.1 (men) and 4.1.2 (women) in "UK Smoking Statistics" edited by N Wald et al (Oxford University Press, 1988). They were supplied to the editors by the Tobacco Advisory Council (TAC). The tables give annual data on the percentage of men and women who smoke manufactured cigarettes by age for the years 1948-85. For the purposes of this report, only data for 1948, 1955, 1965, 1970, 1975, 1980 and 1985 were considered. Each survey concerned about 10,000 people.

The data from HLS were obtained on computer tape from Essex University Archive. Data were available on age of starting to smoke cigarettes for current and ex-smokers and on age of stopping for

ex-smokers. Assuming that smoking was continuous between these ages, or up to date of interview for current smokers, it was possible to estimate the percentage smoking 37, 30, 20, 15, 10, 5 and 0 years before interview. These data were taken to correspond to the seven time points considered for the TAC data (Tables F1, F2).

There were minor differences in method compared with Harris' method. Firstly, for current smokers, Harris took into account the most recent quit attempt, where smoking had stopped for at least a year. No equivalent information was available in HLS. Secondly, we based estimates only on respondents with complete information, whereas Harris made assumptions in order to include respondents with incomplete information.

Tables F3 (males) and F4 (females) compare the percentages of smokers as estimated from the HLS and as given by TAC. Also given are the numbers of subjects considered for the HLS and the difference, HLS - TAC, between the two percentages. For the age group 60+ percentages are not given for the years 1948, 1955 and 1965 as very few subjects of that age in those years would have survived to be surveyed in 1985/86. Percentages are also not given for the age group 16+ for the same three years as the HLS, since the HLS population would be so much younger than the TAC population at that time as to render comparison useless.

For males, the overall percentages of smokers are similar from the two surveys for 1975, 1980 and 1985, though for the individual age groups there are differences of up to about ±5%. For earlier years HLS

percentages tend to be lower. This is most evident for 1970 and particularly 1948, where all four age-specific percentages are lower by about 8-10%.

For females, with two minor exceptions, all percentages tend to be lower for HLS than for TAC. The difference is most marked for 1948, averaging about 10%. For other years, differences for age specific categories tend on average to be about 5%, with no obvious time trend.

There are a number of theoretical reasons why percentages might differ:

(i) Difference in definition of smoker. TAC includes only manufactured cigarette smokers, HLS all cigarette smokers. Thus, all other things being equal, HLS should give higher results, the difference relating to the percentage of the population who smoke handrolled cigarettes only. For women this percentage is miniscule and should not affect the comparison. For men, percentages of handrolled only smokers have, according to TAC data, long been about 4%, perhaps somewhat less than this for younger men and somewhat greater than this for older men.

(ii) Sampling error. Given random sampling, and given two observed percentages of smokers p_T and p_H for TAC and HLS, based on sample sizes N_T and N_H, one can estimate the 95% confidence limits of the difference in percentages by the formula

$$(p_H - p_T) \pm 1.96\sqrt{p_H(100 - p_H) / N_H + p_T(100 - p_T) / N_T}$$

For example, for $p_H = p_T = 50\%$ and $N_H = N_T = 400$, one would observe $0 \pm 6.9\%$. Reducing N_H and N_T would widen the limits, e.g. for $N_H = N_T = 200$ one would observe $0 \pm 9.8\%$. For lower or higher percentages the limits would narrow, though not much in the 30-70% range, e.g. $p_H = p_T = 30\%$, $N_H = N_T = 400$ gives $0 \pm 6.4\%$. For non-random sampling, e.g. stratified sampling as used by TAC, the confidence limits would be somewhat wider than indicated by the formula cited. Sampling error could well explain why for individual age groups in a particular year there is moderate fluctuation in the observed differences.

(iii) Survey methodology. No two surveys, conducted using different techniques, can be expected to give exactly the same results. The comparisons for females for current and recent years suggest that HLS pick up somewhat fewer smokers than TAC. The fact that HLS include handrolled cigarette smokers only and TAC do not, counterbalances this for males.

(iv) Biases due to mortality. While the TAC data are representative by age of the population in the year concerned, the HLS are not. Because survival decreases with increasing age, the average age of the HLS population considered for years before 1985 will be less than that for the TAC population. This should matter little, if at all, for the age groups 16-19, 20-24 and 25-34 where the age range is narrow and the survivorship good. It will be most important for the open-ended age groups 60+ and 65+, especially

for the earlier years. We have omitted presenting the results most affected by this, namely 60+ for 1948, 1955 and 1965. In theory, this bias may have some effect also for 60+ for 1970 and 1975, giving HLS percentages higher than expected (as frequency of smoking declines with old age), but there seems no evidence of this. Bias due to mortality for the age group 35-59 may also be relevant to some extent, although the relative invariance of percentage of smoking over this age range (except perhaps for women in the early years) should minimize this.

(v) Bias due to increased mortality in smokers. Systematic differences between the surveys should not matter greatly when comparing smoking experience in different cohorts, provided that these differences are reasonably consistent over time. One possible cause of an inconsistency over time is differential mortality of smokers and nonsmokers. As discussed in section 7.1, provided we limit attention to subjects aged up to 70 at survey, this bias should not be too bad. Where we are studying older subjects at interview in 1985/86 (e.g. 35-59 for 1948-1970 and 60+ for 1970 onward) some more important bias may have occurred, although in fact the data in Table 1 do not indicate any very large tendency for the difference to decline markedly with age. It is interesting to note that even in 1948, where the differences between HLS and TAC are more substantial, there is no obvious tendency for the difference to rise with increasing age, as would be expected if differential mortality were a major factor.

(vi) Bias due to inadequate recall of past smoking habits. This is an obvious theoretical possibility, but seems not to have been a major factor. It might have contributed to the rather larger differences seen for 1948 though other explanations are possible, including variation in TAC surveys - remember 1948 was the first survey and the methodology may have taken some time to stabilize.

General conclusions

While there are numerous theoretical sources of error, the actual magnitude of these seems not to be unacceptably large. Certainly, provided one limits attention to subjects aged 70 in 1985 one would expect that any attempt to compare patterns of smoking in different cohorts would come up with very similar answers, whether one used TAC or HLS data.

TABLE F1 - Prevalence of cigarette smoking estimated from the Health and Lifestyle Survey - Men

SMOKING BEHAVIOUR NOW

Age now		<16	16-24	25-34	35-49	50-64	65+	16+
Didn't smoke	n	-	338	444	651	598	511	2542
cigarettes	%	-	63.30	61.41	61.53	65.93	75.48	65.20
Did smoke cigarettes	n	-	196	279	407	309	166	1357
	%	-	36.70	38.59	38.47	34.07	24.52	34.80

SMOKING BEHAVIOUR FIVE YEARS AGO

Age five years ago		<16	16-24	25-34	35-49	50-64	65+	16+
Didn't smoke	n	179	375	413	504	498	317	2107
cigarettes	%	79.20	55.97	53.50	54.19	58.45	69.67	57.27
Did smoke cigarettes	n	47	295	359	426	354	138	1572
	%	20.80	44.03	46.50	45.81	41.55	30.33	42.73

SMOKING BEHAVIOUR TEN YEARS AGO

Age ten years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	565	159	175	356	763	295	1748
cigarettes	%	92.17	56.18	48.21	47.79	52.77	64.84	53.10
Did smoke cigarettes	n	48	124	188	389	683	160	1544
	%	7.83	43.82	51.79	52.21	47.23	35.16	46.90

SMOKING BEHAVIOUR FIFTEEN YEARS AGO

Age fifteen years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	890	166	181	283	649	138	1417
cigarettes	%	92.81	55.33	44.25	43.61	47.83	59.74	48.10
Did smoke cigarettes	n	69	134	228	366	708	93	1529
	%	7.19	44.67	55.75	56.39	52.17	40.26	51.90

SMOKING BEHAVIOUR TWENTY YEARS AGO

Age twenty years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1111	172	127	261	548	-	-
cigarettes	%	94.47	50.74	37.80	43.94	44.95	-	-
Did smoke cigarettes	n	65	167	209	333	671	-	-
	%	5.53	49.26	62.20	56.06	55.05	-	-

SMOKING BEHAVIOUR THIRTY YEARS AGO

Age thirty years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1121	142	111	202	289	-	-
cigarettes	%	95.49	57.49	39.50	32.17	42.88	-	-
Did smoke cigarettes	n	53	105	170	426	385	-	-
	%	4.51	42.51	60.50	67.83	57.12	-	-

SMOKING BEHAVIOUR THIRTY-SEVEN YEARS AGO

Age thirty-seven years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1031	111	109	167	142	-	-
cigarettes	%	95.46	47.23	35.50	32.12	40.00	-	-
Did smoke cigarettes	n	49	124	198	353	213	-	-
	%	4.54	52.77	64.50	67.88	60.00	-	-

TABLE F2 - Prevalence of cigarette smoking estimated from the Health and Lifestyle Survey - Women

SMOKING BEHAVIOUR NOW

Age now		<16	16-24	25-34	35-49	50-64	65+	16+
Didn't smoke	n	-	400	631	931	755	768	3485
cigarettes	%	-	64.10	64.85	66.79	64.97	82.58	68.56
Did smoke cigarettes	n	-	224	342	463	407	162	1598
	%	-	35.90	35.15	33.21	35.03	17.42	31.44

SMOKING BEHAVIOUR FIVE YEARS AGO

Age five years ago		<16	16-24	25-34	35-49	50-64	65+	16+
Didn't smoke	n	214	471	613	729	654	508	2975
cigarettes	%	83.59	56.21	58.05	58.32	60.33	82.74	61.44
Did smoke cigarettes	n	42	367	443	521	430	106	1867
	%	16.41	43.79	41.95	41.68	39.67	17.26	38.56

SMOKING BEHAVIOUR TEN YEARS AGO

Age ten years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	661	229	267	561	1054	489	2600
cigarettes	%	91.81	61.23	52.66	55.71	56.18	79.64	59.39
Did smoke cigarettes	n	59	145	240	446	822	125	1778
	%	8.19	38.77	47.34	44.29	43.82	20.36	40.61

SMOKING BEHAVIOUR FIFTEEN YEARS AGO

Age fifteen years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1142	239	284	454	984	283	2244
cigarettes	%	96.21	57.73	51.73	53.54	56.04	82.27	57.38
Did smoke cigarettes	n	45	175	265	394	772	61	1667
	%	3.79	42.27	48.27	46.46	43.96	17.73	42.62

SMOKING BEHAVIOUR TWENTY YEARS AGO

Age twenty years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1499	283	248	423	874	-	-
cigarettes	%	97.53	62.33	54.15	53.41	57.12	-	-
Did smoke cigarettes	n	38	171	210	369	656	-	-
	%	2.47	37.67	45.85	46.59	42.88	-	-

SMOKING BEHAVIOUR THIRTY YEARS AGO

Age thirty years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1571	234	234	376	621	-	-
cigarettes	%	98.81	74.52	58.21	49.34	67.43	-	-
Did smoke cigarettes	n	19	80	168	386	300	-	-
	%	1.19	25.48	41.79	50.66	32.57	-	-

SMOKING BEHAVIOUR THIRTY-SEVEN YEARS AGO

Age thirty-seven years ago		<16	16-19	20-24	25-34	35-59	60+	16+
Didn't smoke	n	1431	187	184	390	367	-	-
cigarettes	%	98.96	69.00	48.42	57.78	71.68	-	-
Did smoke cigarettes	n	15	84	196	285	145	-	-
	%	1.04	31.00	51.58	42.22	28.32	-	-

TABLE F3 - Comparison of percentage of cigarette smokers as recorded in TAC surveys and as estimated for the Health and Lifestyle Survey - men

Year	Source		Age group*					
			16-19	20-24	25-34	35-59	60+	16+
1948	TAC	%	61	74	76	70	-	-
	HLS	%	52.8	64.5	67.9	60.0	-	-
		(n)	(124)	(198)	(353)	(213)		
	difference in %		-8.2	-9.5	-8.1	-10.0		
1955	TAC	%	47	59	67	62	-	-
	HLS	%	42.5	60.5	67.8	57.1	-	-
		(n)	(105)	(170)	(426)	(385)		
	difference in %		-4.5	+1.5	+0.8	-4.9		
1965	TAC	%	50	63	56	56	-	-
	HLS	%	49.3	62.2	56.1	55.1	-	-
		(n)	(167)	(209)	(333)	(671)		
	difference in %		-0.7	-0.8	+0.1	-0.9		
1970	TAC	%	55	58	60	55	46	55
	HLS	%	44.7	55.8	56.4	52.2	40.3	51.9
		(n)	(134)	(228)	(366)	(708)	(93)	(1529)
	difference in %		-10.3	-2.2	-3.6	-2.8	-5.7	-3.1
1975	TAC	%	49	53	46	49	41	47
	HLS	%	43.8	51.8	52.2	47.2	35.2	46.9
		(n)	(124)	(188)	(389)	(683)	(160)	(1544)
	difference in %		-5.2	-1.2	+6.2	-1.8	-5.8	-0.1
			16-24	25-34	35-49	50-64	65+	16+
1980	TAC	%	44	47	43	43	28	42
	HLS	%	44.0	46.5	45.8	41.6	30.3	42.7
		(n)	(295)	(359)	(426)	(354)	(138)	(1572)
	difference in %		0.0	-0.5	+2.8	-1.4	+2.3	+0.7
1985	TAC	%	42	40	36	29	24	35
	HLS	%	36.7	38.6	38.5	34.1	24.5	34.8
		(n)	(196)	(279)	(407)	(309)	(166)	(1357)
	difference in %		-5.3	-1.4	+2.5	+5.1	+0.5	-0.2

*All groups for the HLS are based on the age the respondents would have been in the years considered.

TABLE F4 - Comparison of percentage of cigarette smokers as recorded in TAC surveys and as estimated for the Health and Lifestyle Survey - women

Year	Source		Age group*					
			16-19	20-24	25-34	35-59	60+	16+
1948	TAC	%	43	54	52	41	-	-
	HLS	%	31.0	51.6	42.2	28.3	-	-
		(n)	(84)	(196)	(285)	(145)		
	difference in %		-12.0	-2.4	-9.8	-12.7		
1955	TAC	%	26	39	51	41	-	-
	HLS	%	25.5	41.8	50.7	32.6	-	-
		(n)	(80)	(168)	(386)	(300)		
	difference in %		-0.5	+2.8	-0.3	-8.4		
1965	TAC	%	40	51	50	50	-	-
	HLS	%	37.7	45.9	46.6	42.9	-	-
		(n)	(171)	(210)	(369)	(656)		
	difference in %		-2.3	-5.1	-3.4	-7.1		
1970	TAC	%	52	54	51	50	26	44
	HLS	%	42.3	48.3	46.5	44.0	17.7	42.6
		(n)	(175)	(265)	(394)	(772)	(61)	(1667)
	difference in %		-9.7	-5.7	-4.5	-6.0	-8.3	-1.4
1975	TAC	%	46	53	49	49	27	43
	HLS	%	38.8	47.3	44.3	43.8	20.4	40.6
		(n)	(145)	(240)	(446)	(822)	(125)	(1778)
	difference in %		-7.2	-5.7	-4.7	-5.2	-6.6	-2.4
			16-24	25-34	35-49	50-64	65+	16+
1980	TAC	%	40	45	44	46	21	39
	HLS	%	43.8	42.0	41.7	39.7	17.3	38.6
		(n)	(367)	(443)	(521)	(430)	(106)	(1867)
	difference in %		+3.8	-3.0	-2.3	-6.3	-3.7	-0.4
1985	TAC	%	40	42	38	37	19	34
	HLS	%	35.9	35.2	33.2	35.0	17.4	31.4
		(n)	(224)	(342)	(463)	(407)	(162)	(1598)
	difference in %		-4.1	-6.8	-4.8	-2.0	-1.6	-2.6

*All groups for the HLS are based on the age the respondents would have been in the years considered.

Appendix G

Trends in lung cancer in nonsmokers

Author: P N Lee

Date: 7.4.94

It has been suggested by a number of authors that factors other than smoking are playing an increasing role in the aetiology of lung cancer. In theory, one of the most direct methods of obtaining evidence on this would be to study trends over time in the risk of lung cancer among lifelong nonsmokers. In practice, there are a number of reasons why it is quite difficult to obtain such evidence.

Firstly, it should be realized that national mortality statistics, which give voluminous data on risk of disease by cause, age, sex, country and year, do not give data broken down by smoking habits. This is because they are based on death certificates, where smoking habits are not recorded. Estimates of risk of lung cancer in nonsmokers can only be obtained from prospective epidemiological studies (case-control studies can only determine relative, not absolute, risk). Such studies have to be very large indeed to get reliable results, given the rarity of lung cancer in nonsmokers. For example, 20 years' observations on 34,440 male British doctors (Doll and Peto, 1976) only yielded 10 lung cancer deaths in nonsmokers, far too few to determine any time trend reliably. There are only a very limited number of studies which have the potential to produce useful data.

Based on the two American Cancer Society (ACS) Cancer Prevention Studies (CPS), each of over a million men and women, the first starting in 1959 with follow-up for 12 years, the second starting in 1982 with follow-up for four years, Garfinkel and Silverberg (1990) compared age-standardized lung cancer death rates in four four-year periods. As shown in Table 1, there was no real evidence of a time trend, with, in each sex, rates quite comparable in the four periods. A similar conclusion can be reached from results of an earlier analysis (data also shown in the table) based on partly incomplete follow-up (US Surgeon-General, 1989).

There are some difficulties in interpreting directly from these data that no increase has occurred:

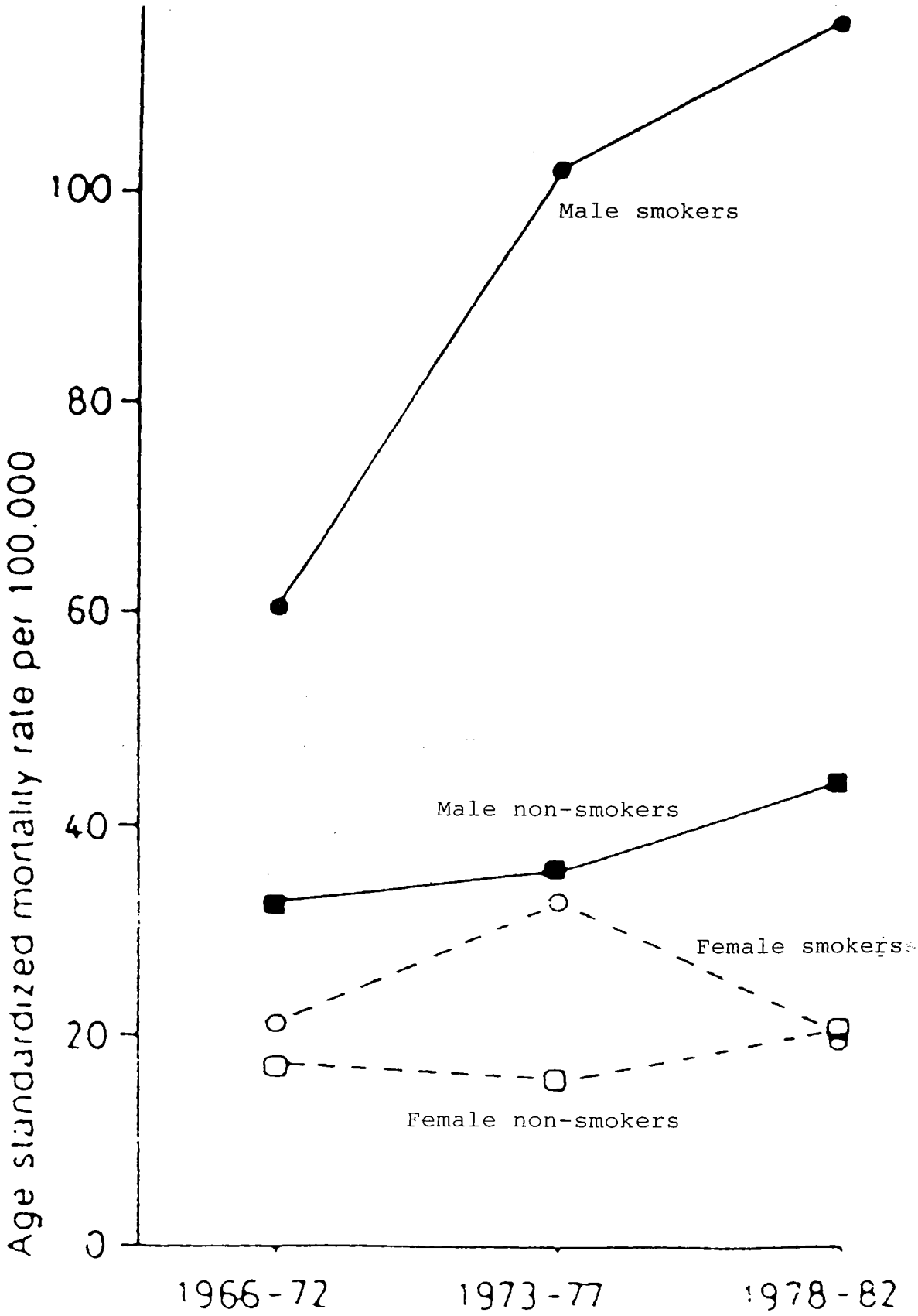
(a) Sampling variation does not exclude the possibility of a modest true increase having occurred.

(b) The populations studied are known to be unrepresentative of the US population at large, being virtually wholly white, much more educated and affluent than average, and much less likely to work in occupations that incur a high risk of lung cancer.

(c) The diagnoses are based on death certificates which are known to be unreliable. In the absence of autopsy, which infrequently occurs, clinical diagnosis of lung cancer has been shown to be inaccurate, with evidence (Feinstein and Wells, 1974) of a particular problem in nonsmokers. It is not clear, however, what effect such inaccuracy should have on trends.

Another US study which has been studied for trends in nonsmokers' lung cancer death rates is the US Veterans' Study, in which over a quarter of a million US veterans were interviewed in 1954 or 1957 and followed-up for up to 16 years. Doll and Peto (1981) presented results of an analysis (see Table 2) which again showed no evidence of any significant trends over time. Considering these data, and also those from CPS-I, Doll and Peto remained "unconvinced that any material trends in true lung cancer death rates among American non-smokers have occurred in recent decades", though they noted that "some such increases should be expected if the effects of passive smoking reported by Hirayama (1981) and Trichopoulos et al (1981) are confirmed".

A third large prospective study which has provided some data on trends in lung cancer is the Japanese study of Hirayama in which over a quarter of a million Japanese men and women, interviewed in 1965, were followed up for 17 years. In his book, Hirayama (1990) presented a graph (reproduced below) showing trends in age-standardized lung cancer rates over three periods, 1966-72, 1973-77, and 1978-82. In both sexes a slight increase is seen in nonsmokers' lung cancer rates over the period, but Hirayama makes no statement as to statistical significance. Given further data presented in the same book (see Table 3) showing inconsistent time trends in nonsmokers in different age groups, it appears the increases are probably not significant. One must have considerable reservations about the validity of these analyses, since they do not show the expected rise in risk with age, and because of a number of other study weaknesses discussed elsewhere (Lee, 1992).


Age standardized mortality rate per 100.000
100
80
60
40
20
0
Male smokers
Male non-smokers
Female smokers
Female non-smokers
1966-72
1973-77
1978-82

There have been a number of other attempts to try to gain information on trends in lung cancer among nonsmokers.

Enstrom (1979) presented a paper claiming that lung cancer mortality among persons who never smoked cigarettes rose substantially between 1914 and 1968. Though he concluded that most of the relative increase that occurred before 1935 was probably due to changes in diagnostic criteria, he considered that real increases had occurred since 1935, and that factors other than cigarette smoking had had a significant effect on the mortality rate from this disease. In order to obtain data on trends in this period he used four sources of information:

1914: Data for 24 states on overall lung cancer rates, it being assumed that the data were representative of the US and that they would have been unaffected by smoking at that time, i.e. they could be assumed to be nonsmokers' rates.

1935: National data on overall lung cancer rates, it being assumed that for those aged 65 or over, nonsmokers had the same rates as the total population;

1958: Data from the 1958-59 National Mortality Survey, which combined information from a nationally representative 10% sample of all deaths in the US, for whom data on smoking were obtained by a questionnaire sent to the family informant, and a representative sample of the living population, who were asked questions *inter alia* on smoking.

1966-68: Similarly to the 1958 data.

The main results from Enstrom's analysis are summarized in Table 4. Although they shown a markedly increasing trend, there are two major problems in inferring any true increase in lung cancer rates. The first, noted by Enstrom, is that substantial improvements in diagnosis had occurred. Certainly it is well known that in 1914 the ability to detect lung cancer in-life was very limited. The second major problem is that the smoking data collected in 1958 and 1966-68 came from proxies. Given a proportion of respondents would never have known the full life history of the decedent, it is likely, as pointed out by Doll and Peto (1981), that some of the so-called lifelong nonsmokers were in fact ex-smokers. As the risk of lung cancer in ex-smokers was increasing with time, correlated with the increasing likelihood of having smoked for longer periods of time, this inclusion of ex-smokers might have caused an apparent increase in risk among men and women reported to be smokers when no true increase in fact existed. In support of this argument, Doll and Peto pointed out that age-adjusted lung cancer death rates in nonsmokers in the 1966-68 National Mortality Survey were actually 80% higher than seen in CPS-I (1960-72). However, it must be pointed out that it is not clear whether the whole of this excess is due to more true ex-smokers being included since, as noted above, the CPS-I population is unrepresentative in many ways.

Enstrom (1979) also included a comparison (reproduced in Table 5) of lung cancer rates in men who had never smoked in the US Veterans Study and in the ACS CPS-I study, referable to the period 1954-63, and in active Mormons in California, referable to the period 1968-75. Although death rates in the Mormons were about twice as high as those in the other

groups, Doll and Peto (1981) point out that this is not actually evidence that nonsmoker death rates increased at all between 1960 and the early 1970's, the reason being that about one-third of active Mormons in California are actually ex-smokers and not all lifelong never smokers, as would be necessary for a valid comparison. It is also far from clear that the populations of the three studies are comparable in respect of many variables other than smoking.

Mori and Sakai (1984) carried out a study involving all 15,367 cases autopsied over the period 1936 to 1978 in the Department of Pathology at the University of Tokyo. From the clinical history abstracts attached to the autopsy protocol 6610 cases, 4269 men and 2341 women, were selected who were aged 20 or over and who had cigarette smoking history available. As shown in Table 6, there was a striking tendency for age adjusted incidence of lung cancer to rise among nonsmokers, with risk rising significantly ($p<0.05$) in both sexes. In interpreting this finding, a number of points have to be considered:

(i) Since these were all autopsy cases, improvements in diagnosis can effectively be excluded as an explanation for the increase.

(ii) There was a striking increase in average age of the cases over the study period, but age adjustment should have accounted for this.

(iii) It is unclear how representative the autopsied population is of all deaths. The autopsy rate is known to be very low in Japan.

(iv) Smoking data taken from clinical notes may be seriously inaccurate. The probability of cigarette smoking history being available for a lung cancer case might have increased dramatically. At the beginning of the study lung cancer was not known to be associated

with smoking, but at the end it would be difficult to imagine a suspect lung cancer case not being asked about his smoking habits.

(v) Lung cancer rates have risen very steeply in Japan since the war, much more so than in Western countries. Hirayama (1981) presented a graph showing a 10-fold increase between 1947 and 1978, whereas Hirayama (1984) reported smoker/nonsmoker relative risks much lower than this. This suggests a major effect of factors other than smoking in Japan.

(vi) Mori and Sakai themselves felt their results indicated that factors such as atmospheric pollution, heavy metals, asbestos, diesel exhaust, and urbanization were possibly as important or more important than cigarette smoking.

Stevens and Moolgavkar (1984) carried out a statistical analysis relating age-specific data on trends in male lung cancer deaths in England and Wales over the period 1941-45 to 1971-75 to UK data on the annual percentage of smokers and an estimated cumulative constant tar cigarette consumption by age and birth cohort. They fitted a model in which risk was estimated as a product of terms representing effects of age, cigarette consumption and period of death. Their model explained more than 99% of the observed variation in death rates. One conclusion of their model was that lung cancer rates among nonsmokers had been declining continuously since 1951-55 (see Table 7), a decline they attribute to reductions in smoke and SO_2 pollution. Although Lee, Fry and Forey (1990) also concluded, by means of a rather different approach, that there had been some decline in lung cancer rates in young men and women that cannot be attributed to cigarette smoking, Stevens and

Moolgavkar's paper is weak in that the function they fit to account for effects of cigarette smoking is totally implausible, implying inter alia that a smoker aged 75 who smoked two packs a day would have 7000 times the risk of lung cancer of a smoker aged 75 who did not smoke. Clearly the form of the function used to fit cigarette smoking effects may have a dramatic effect on conclusions regarding nonsmokers.

Another indirect attempt to estimate trends in nonsmokers' death rates is the truly dismal paper by Axelson et al (1990). They correctly pointed out that, given the lung cancer rate for the total population (L), the proportion of the population who have ever smoked (S), and the relative risk of lung cancer for ever smokers compared to never smokers (R), one can easily estimate the lung cancer rate for never smokers. Using estimates of L, S and R for Japan, Italy and the US at various time points they then concluded that there has been a positive time trend in each country in rates for never smokers. An obvious major flaw in their analysis is that they assumed R does not vary over time when there is good evidence that it has increased substantially. (Compare, for example, the estimates of R=2.69 for 1959-65 and R=11.94 for 1982-86 given in the 1989 Surgeon-General's Report based on the two American Cancer Society Cancer Prevention Studies). This on its own is sufficient to totally invalidate their analysis, but there are a number of other weaknesses too, including failure to study age-specific rates, failure to consider possible effects of smoking habit misclassification on the estimates of R, and assuming that lung cancer rates can be accurately estimated simply on the basis of the percentage of smokers 20 years earlier. At one point in their paper they did consider the possibility

that increased duration of smoking might have biased their analysis but they dismissed this on the basis of results of Garfinkel and Stellman (1988) which they interpreted as showing only a weak effect of duration. However, their interpretation is totally erroneous, based on a false comparison of two standardized mortality rates with different bases. The whole paper, which is extremely superficial, can be considered worthless.

A better indirect attempt to estimate trends in nonsmokers' death rates was made by Forastière et al (1993). Based on smoking habit surveys conducted in Italy in 1957, 1965, 1980 and 1986-87 and national estimates of lung cancer mortality rates for 1956-58, 1965-67, 1980-82 and 1987-89, the authors estimated lung cancer death rates in nonsmokers based on four different models:

Model 1 - Relative risks for smokers and ex-smokers constant over the period (10 and 4 for males; 4 and 1.6 for females)

Model 2 - Relative risks for smokers and ex-smokers depend on the average number of cigarettes smoked per day, but not on duration of smoking

Model 3 - Relative risks for smokers and ex-smokers depend on a function given by Whittemore (1988) in which excess risk is a product of duration of smoking and packs per day

Model 4 - Relative risks for smokers and ex-smokers depend on a "multistage" function fitted by Whittemore (1988) to data for British doctors.

As shown in Table 8, all models in both sexes showed a consistent rise over the period studied. The authors reported that the rises were evident in analysis by separate age group and claimed that in sensitivity analysis (using Model 4) the conclusions were similar even after taking account of possible underestimation of smoking, different assumed values of age of starting to smoke (data for 1957 and 1965 were not available and had to be estimated), and different assumed values of the parameters in Whittemore's "multistage" function.

Though suggestive that, as the authors conclude, "factors other than smoking play an important role in causing lung cancer in Italy", one must have reservations for a number of reasons. Firstly, the results involving Model 1 and Model 2 are likely to be irrelevant since they do not take duration of smoking into account at all. Secondly, the functions used in Model 3 and Model 4, and the assumed data for age of starting to smoke in 1956-58 and 1965-67, may not have taken duration of smoking properly into account. Observed trends over time in smokers' relative risk reported elsewhere (see comments on the Axelson et al paper) have been much greater than those fitted here from Model 4 (rising from 7.2 to 13.1 in males and from 2.6 to 4.0 in females between 1956-58 and 1987-89), which may be indicative of poor fit of the model or use of inappropriate data. Also it should be noted that Whittemore's Model 4 for the risk at age t in smokers starting at age t_0 and stopping at age t_1 is not actually multistage at all. (Ignoring the lag period of five years) she uses a function of the form

$$R = At^k + B(t_1-t_0)^k + C(t_1^k-t_0^k) + D(t_1-t_0)^k$$

TABLE 1: Trends in lung cancer rates (per 100,000 per year) in US nonsmokers (ACS data)

		Male	Female
From Garfinkel and Silverberg (1990)[1]			
1960-64	CPS-I	14.6	11.7
1965-68		16.6	12.4
1969-72		16.7	12.2
1982-86	CPS-II	15.4	12.1
From US Surgeon-General (1989)[2]			
1959-65	CPS-I	15.5(12.5-19.3)	10.3(8.9-11.9)
1982-86	CPS-II	13.6(10.8-17.0)	11.4(9.8-13.3)

[1]Rates standardized to the age distribution of the US population in 1970.

[2]Rates standardized to the age distribution of the US population in 1965; death rates for CPS-II corrected for delayed ascertainment of cause of death, all death certificates not having been received at the time the analysis was conducted; numbers in parentheses are 95% confidence intervals.

TABLE 2: Trends in lung cancer rates in male US nonsmokers (US veterans' data)

Years since entry to study[1]	Lung cancers		
	Observed	Expected[2]	Ratio
1	6	6.5	0.9
2,3,4	24	23.6	1.0
5,6,7	31	30.9	1.0
8,9,10	40	39.2	1.0
11,12,13	41	43.9	0.9
14,15,16	35	33.0	1.0
Total	177	177.0	1.0

[1]There were two samples of veterans, one interviewed in early 1954, one in early 1957.

[2]Expected assuming there is no trend over time in lung cancer rate.

TABLE 3: Trends in lung cancer rates (per 100,000 per year) in male Japanese nonsmokers (Hirayama data)

Period	Age group			
	55-59	60-64	65-69	70-74
1966-72	7	15	28	51
1973-77	43	24	49	72
1978-82	0	37	13	48

TABLE 4: Trends in US lung cancer rates (per 100,000 per year) in nonsmokers (Enstrom data)

Sex	Year	Smoking[1]	Age group 55-64	65-74	75-84	35-84[2]
Male	1914	NSC	3.0	2.6	1.2	1.6(148)
	1935	NSC	-	26.7	23.3	-
	1958	NS	12.7	25.0	55.0	10.8(80)
	1958	NSC	14.8	33.7	69.7	13.3(80)
	1966-68	NSC	32.2	65.6	89.9	22.8(108)
Female	1914	NS	2.2	2.2	1.5	1.3(124)
	1935	NS	9.8	14.5	14.5	-
	1958-9	NS	10.4	21.0	34.0	8.3(456)
	1966-68	NS	11.4	19.6	38.8	8.3(123)

[1]NS = never smoked, NSC = never smoked cigarettes

[2]Age adjusted to the 1960 US population, numbers of deaths in parentheses

TABLE 5: Comparison of lung cancer rates (per 100,000 per year) in three groups of white males (Enstrom data)

Study population	Year	Age group 55-64	65-74	75-84	35-84[1]
US Veterans					
Never smoked or occasionally only	1954-62	10	32	50	9.4(78)
Never smoked cigarettes	1954-62	12	38	60	12.7(156)
ACS CPS-I					
Never smoked regularly	1960-63	15	15	44	10.4(49)
Never smoked cigarettes	1960-63	18	29	56	13.4(104)
US Veterans + ACS CPS-I combined					
Never smoked	1954-63	12	26	45	10.8(127)
Never smoked cigarettes	1954-63	14	35	57	13.1(260)
Active Mormons					
All	1968-75	28	54	145	24.5(63)

[1]Age adjusted to the 1960 US population, numbers of deaths in parentheses.

TABLE 6: Trends in lung cancer incidence[1] among autopsied men and women in Tokyo (Mori and Sakai, 1984)

Period	Men	Women	Total
1936-45	0.2%	1.2%	0.8%
1946-55	1.8%	1.6%	2.0%
1959-68	3.2%	3.9%	4.0%
1969-78	6.0%	4.2%	4.7%
Trend p	<0.05	<0.05	<0.02

[1]Age adjusted.

TABLE 7: Trends in estimated lung cancer death rate (per 100,000 per year) among British male nonsmokers aged 35-84 (from Moolgavkar and Stevens)

Year	Lung cancer rate
1941-45	14.9
1946-50	17.8
1951-55	19.3
1956-60	18.8
1961-65	14.0
1966-70	12.0
1971-75	8.6

TABLE 8: Estimated trends in lung cancer rates (per 100,000 per year) in Italy (Forastière data)

Model (see text)	Sex	Years			
		1956-58	1965-67	1980-82	1987-89
1. (constant RRs)	Male	3.2	6.0	12.4	15.8
	Female	4.6	6.1	7.2	8.2
2. (dose)	Male	4.1	7.8	12.9	16.6
	Female	4.5	6.1	5.6	6.3
3. (dose and duration, packs-function)	Male	3.3	6.0	9.3	10.6
	Female	5.1	6.8	7.1	7.5
4. (dose and duration, multistage)	Male	4.4	7.9	11.8	12.3
	Female	5.1	6.9	7.4	8.1

Appendix H

TABLE H1
Estimates of prevalence of smoking in Italy, from La Vecchia et al.

Calendar Year	Cohort 1895	1905	1915	1925	1935	1945	1955	1965
Male								
1910	6.3	10.4	15.6	14.3	15.1	13.8	17.8	12.7
1920	48.9	57.3	59.9	64.9	59.8	59.8	55.3	
1930	55.6	58.9	63.2	68.3	61.2	57.3		
1940	53.7	55.2	60.6	62.6	55.9			
1950	43.6	53.3	54.5	52.2				
1960	38.4	44.6	41.1					
1970	34.0	30.1						
1980	18.4							
Female								
1910	0.2	0.3	0.9	1.0	1.7	2.3	5.9	8.0
1920	1.6	2.5	5.1	8.0	11.4	20.3	32.1	
1930	2.4	3.8	7.2	11.3	15.7	25.2		
1940	2.6	4.5	7.8	12.1	17.4			
1950	2.4	4.5	8.3	12.5				
1960	2.0	4.1	6.8					
1970	1.5	3.0						
1980	1.1							

TABLE H2

Estimates of prevalence of smoking in Norway, from Ronneberg et al.

Age	Cohort 1890-1894	1895-1899	1900-1904	1905-1909	1910-1914	1915-1919	1920-1924	1925-1929	1930-1934	1935-1939	1940-1944	1945-1949	1950-1954	1955-1959	1960-1964	1965-1969	1970-1974
Male																	
15-19	37	40	43	44	47	46	49	55	57	57	42	35	36	37	30	29	20
20-24	53	57	62	63	67	70	75	78	76	70	57	52	51	43	37	35	
25-29	56	61	66	67	71	75	76	76	71	61	60	54	45	42	37		
30-34	57	61	68	70	74	77	75	73	66	61	52	47	46	44			
35-39	56	61	68	71	74	75	73	67	61	54	48	45	45				
40-44	57	62	69	71	73	73	67	61	52	41	46	43					
45-49	57	62	69	70	68	65	61	61	42	43	40						
50-54	57	63	68	66	61	61	57	46	44	42							
55-59	57	62	59	56	58	50	45	45	37								
60-64	57	52	53	53	52	40	42	36									
65-69	46	45	47	43	38	38	30										
70-74	39	41	32	39	34	37											
Female																	
15-19	1	1	1	2	4	5	8	13	23	28	27	26	35	36	33	26	25
20-24	2	3	4	8	11	15	26	37	42	40	39	52	47	43	40	35	
25-29	3	4	7	11	17	24	36	40	40	39	52	46	44	44	41		
30-34	3	5	10	15	23	31	37	37	38	47	39	41	43	43			
35-39	4	7	13	20	27	35	34	38	43	42	37	39	44				
40-44	6	10	16	24	30	30	34	38	37	33	37	39					
45-49	7	12	19	26	25	27	34	35	36	36	36						
50-54	9	15	21	21	22	29	32	31	33	30							
55-59	10	15	15	18	24	24	23	26	32								
60-64	10	10	13	19	19	23	24	24									
65-69	4	8	13	11	12	12	18										
70-74	5	7	7	11	10	15											

Office on Smoking and Health

Fact Sheet
Epidemiology Branch

In 1987, the Office on Smoking and Health formed an Epidemiology Branch to enhance research activities relating to tobacco use. The Branch is involved in a variety of activities to determine tobacco use patterns in the United States. It conducts new scientific studies and surveys, analyzes existing data sources, and provides technical and scientific assistance to researchers, health departments, and other health professionals.

The main functions of the Epidemiology Branch include the following:

- Undertaking studies to determine tobacco use patterns and to identify barriers that may be slowing down the reduction in smoking prevalence.
- Disseminating the results of studies in a manner that assists in establishing a public agenda against tobacco use.
- Coordinating and maintaining computer tapes of national data containing smoking information which can be used as the basis for additional studies.
- Providing scientific support for the annual Surgeon General's reports on smoking and health.
- Providing advice and research blueprints to smoking researchers working at the State and local levels to assist in evaluating interventions to reduce smoking prevalence and environmental tobacco smoke exposure.

The data sets that are available for analysis include: 12 National Health Interview Surveys since 1965; 4 Current Population Surveys conducted since 1966; yearly Behavioral Risk Factor Surveillance System surveys since 1981; 5 Adult Use of Tobacco Surveys since 1964; 6 Teenage Tobacco Surveys since 1968; yearly High School Senior Surveys since 1975; and 9 National Institute on Drug Abuse Household Surveys since 1971. In addition, there are a number of other special data sets such as the series of National Health and Nutrition Examination Surveys.

Gary A. Giovino, Ph.D.
Acting Chief
Epidemiology Branch
Office on Smoking and Health
(301) 443-0620
FTS 443-0620

10/90