April, 1989

EX-SMOKERS AND THE MULTISTAGE MODEL
FOR LUNG CANCER

by

David A. Freedman
Statistics Department
University of California
Berkeley, Calif 94720
415-642-2781

and

William C. Navidi
Mathematics Department
University of Southern California
Los Angeles, Calif 90089-1113
213-743-8810

Technical Report No. 135
Statistics Department
University of California
Berkeley, Ca 94720

To appear in
EPIDEMIOLOGY, vol 1, 1990

Research partially supported
by NSF Grant DMS 86-01634
Abstract

Most versions of the multistage model predict that when persons stop smoking, their excess risk for lung cancer will continue to increase. Discussions of the model usually indicate that the excess risk stabilizes. The data show that the risk declines. Implications for models of carcinogenesis are discussed.

Acknowledgments

We would like to thank J Cairns, L Garfinkel, S Moolgavkar, D Petitti, J Peto, M Pike, J Robins, S Swan, D Thomas, and A Whittemore for their help, and especially K Rothman for editorial support. The American Cancer Society and the National Cancer Institute kindly provided data. Our research was partially supported by NSF Grant DMS 86-01634.
Current mathematical models for carcinogenesis\textsuperscript{1} imply that when persons stop smoking, their excess risk for lung cancer will continue to increase. Within the field of risk assessment, however, there seems to be general agreement that the excess risk stabilizes on cessation of exposure\textsuperscript{2-6}. We report here on data from three of the main cohort studies, and find that the excess risk declines; indeed, even the absolute risk (background + excess) seems to decline for about 20 years after cessation of smoking.

All the empirical studies we have identified, which correlate the risk in ex-smokers with time since quitting, confirm the finding that excess risk declines. We begin by reviewing the cohort studies. Hammond\textsuperscript{7} uses direct standardization on age, and concludes that after 10 years, the lung cancer risk for ex-smokers in the 25-state ACS cohort is virtually the same as that for non-smokers. Hammond & Horn\textsuperscript{8} report on the 9-state ACS cohort, and Kahn\textsuperscript{9} on the veterans. Data from Cederlof et al\textsuperscript{10} are compatible with present results, although the numbers are very small. In a large cohort of female ex-smokers, Garfinkel and Stelman\textsuperscript{11} show that relative risk declines rapidly after cessation of smoking, but stays above one.
Next, case-control studies. In a recent very large study, Lubin et al.\textsuperscript{12} find decreasing excess risk: "The subjects who had smoked for one to 19 years returned to the risk level of lifetime non-smokers within five to 10 years after stopping smoking." Other studies\textsuperscript{13-15} report similar results. We have found no other studies that report on the risk for ex-smokers by years since quit. The IARC\textsuperscript{4} reviews recent literature on the effects of smoking; for reviews of the earlier literature, see the US Public Health Service\textsuperscript{16-17}.

Declining excess risk is contrary to the predictions of the Armitage-Doll multistage model\textsuperscript{1}. The decline is compatible with pathology results\textsuperscript{18}, and could be explained if the body is able to repair the lesions caused by smoking. For another discussion focused on the British doctors, see Gaffney and Altshuler\textsuperscript{19}; they too find declining excess risk after cessation, note the inconsistency with the Armitage-Doll model, and suggest using alternative models. Also see Moolgavkar, Dewanji and Luebeck\textsuperscript{19a}. 
Materials and methods

The three cohorts under consideration are as follows:

a) The British doctors. Doll & Peto report on smoking and lung cancer in their seminal cohort study of British doctors. The data quality is considered to be excellent; dose was ascertained on three separate questionnaires. Information on age at start of smoking, however, has not been published; following Doll & Peto, this value is imputed as 22.5 years (including some allowance for the time from malignancy to death). Although the study lasted 20 years with about 34,000 subjects, the number of events--lung cancer cases--is relatively small. There is some deficit in events at the highest ages, and at the highest dose. Dose is measured in cigarettes per day. Doll & Peto report data on non-smokers and current smokers, selecting only subjects who smoked at a nearly constant rate; only 215 events out of 571 are kept. The published data on ex-smokers are not in usable form, and the unpublished data do not appear to be available. Summary data on ex-smokers, however, have been published.
b) The Dorn veterans. Kahn⁹ reports on Dorn's cohort study of about 300,000 American veterans, begun in 1954; also see Rogot²⁴ and Whittemore²⁵. The data used here come from a tape supplied in 1981 by the National Cancer Institute under the Freedom of Information Act. This tape combines the 1954 and 1957 cohorts, and reports on follow-up through 1969; furthermore, the data have been edited by NCI personnel. Data on the tape therefore differ from published tables. Dose was ascertained only once, and there is some deficit in events at the highest ages. On the positive side, this data set is quite large (1266 events); it has information on age at start of smoking; it has ex-smokers; the risk for current smokers increases with dose.

c) The ACS (American Cancer Society) volunteers. This study is described by Hammond⁷. L. Garfinkel of the ACS provided a table of person-years and events for current smokers over the period July, 1960 to June, 1965, by age, age at start of smoking, dose, and sex. (The table differs in some respects from published data.) Data were also provided for non-smokers and ex-smokers. The data quality seems good. The risk for smokers goes up with dose; there is some deficit in events beyond age 79, compared with rates in the age group 75-79.
For the veterans and the ACS cohorts, the risk from lung cancer for ex-smokers can be tabulated as a function of years since quitting. The risk declines. Of course, persons who quit long ago may be lighter smokers, and may have been younger when they quit. These confounding factors can be adjusted for, by direct standardization; on the whole, the risk still declines. (Age is standardized as of time of quitting; standardizing age as of time of entry to the study would in principle answer a different question, about the impact of duration of smoking on risk.)

For the British doctors, Doll\textsuperscript{21} compared the risk of ex-smokers with the risk for current smokers and non-smokers, adjusting for age and dose at time of quitting. A similar analysis will be presented here for the veterans and ACS cohorts. The observed number of events for ex-smokers will be compared with the number expected--

(i) if absolute risk stabilizes at time of quitting;

(ii) if excess risk stabilizes at time of quitting, but background risk continues to increase with age, as for non-smokers.

The observed number will also be compared with the number expected for lifelong non-smokers of the same age.
The multistage model is used as the basis for computing expected numbers. In effect, using the model trades variance for bias. (As a basis for testing the model, the procedure can hardly be objected to.) The expected rates are projected from the continuing smokers or non-smokers to ex-smokers. This approach has two advantages: it uses the bulk of the data, and it tests the ability of the model to extrapolate from one set of circumstances to another. If the model does have a biological basis, the test seems a fair one.
Results

As shown in Table 1 for the veterans cohort, the crude risk of lung cancer among ex-smokers declines steadily from time of quitting.

----- Table 1 about here -----

Persons in the different lines of Table 1 differ by dose and age at time of quitting, as shown in the first two columns of Table 2. For example, those who quit 5-9 years before baseline are lighter smokers than those who quit 1-4 years before, and were younger when they stopped smoking (although the differences are small).

----- Table 2 about here -----

The third column in Table 2 shows directly standardized risks; age is standardized as of the time when exposure stopped. The standardized risk declines fairly steadily as a function of years since quit, except for a perhaps artifactual spike at 30-34 years.

----- Table 3 about here -----

Tables 3 and 4 repeat the analysis for the ACS cohort. Crude risks are shown in Table 3, and decline rapidly. Table 4 shows the standardized risks, and these still decline steadily, although not as rapidly.

----- Table 4 about here ----
For the comparison of observeds and expecteds, Figure 1 shows results on the British doctors (Doll\textsuperscript{21}, Wynder & Hecht\textsuperscript{26} and the US Public Health Service\textsuperscript{17}). The middle curve plots the risk for ex-smokers, relative to their risk at time of quitting. (The top curve shows the risk for continuing smokers; the bottom, for non-smokers: Doll & Hill's figure\textsuperscript{22} is based on less data, but may be clearer.)

----- Figure 1 about here ----- 

The ratio is plotted as a function of years since quit. It declines at first, then rises; but remains below 1.00 until 20 years after quitting. (Of course, risk increases with age, and that may explain the late rise in the curve; more on this below.) The figure is drawn in a logarithmic scale, so the risk 5 years afterward is about 50\% of the risk at time of quitting.

----- Table 5 -----
Table 5 presents a similar analysis for the veterans cohort, comparing the observed number of events for ex-smokers with expected numbers. In column 1 of Table 5, the "expecteds" are computed on the basis of risk at time of quitting; this computation takes into account dose, age at start, and age at quit. In effect, absolute risk-- not just excess risk-- is frozen at the time of quitting.

Although it stays close to one, the ratio O/E declines for about 20 years, and then starts rising-- in fair agreement with the middle line in Figure 1. By 15-24 years after quitting, the absolute risk is below what it was at the time of quitting. Since background risk increases, excess risk must be decreasing.

The late increase in O/E is probably due to age, which has not been factored into the E's for column 1. In column 2, excess risk is frozen at the time of quitting and background risk from aging is allowed to increase. (This calculation happens to give the risk predicted by the best multistage model fitted to the data on current smokers and non-smokers, as shown in the Appendix.) Thus, column 2 adjusts for dose, age at start, age at quit, and current age. The ratio of observed to expected drops quite steadily. Indeed, after 15 years, the excess risk drops below its value at time of quitting.
Column 3 shows the ratio of observed to background—the expecteds are for non-smokers of the same age. This calculation adjusts for current age only. The ratio of observed to expected falls dramatically. Beyond 25 years, the ex-smokers look like never-smokers (although the number of events is rather small).

An example may clarify the meaning of the E's. Take someone who starts smoking at age 19, smokes 30 cigarettes a day until age 42, then quits. His empirical risk at (say) age 57 is compared with three theoretical risks, in the three columns of Table 5; it is convenient to explain the columns in the order 1, 3, 2.

Column 1: The risk for persons aged 42 who smoked 30 cigarettes per day from age 19 onwards.

Column 3: The risk for persons aged 57 who never smoked.

Column 2: The risk equals for non-smokers as in column 3, plus the excess risk for the smokers as in column 1.

(Explicit formulas for the expecteds are in the Appendix.)

Table 6 repeats the analysis for the ACS males; in all three columns, O/E drops sharply. (In the last line of the table, there are only 6 events, so the results cannot be taken too literally.)

------ Table 6 about here ------
Discussion

The Armitage-Doll\textsuperscript{1} multistage model says in essence that a cell progresses to malignancy through the states of a Markov chain. This model is often used in cancer risk assessment; see the Environmental Protection Agency\textsuperscript{27}. And it is often cited in discussions of the biological mechanisms of cancer; see Borszonyi et al\textsuperscript{28}. The model makes a strong qualitative prediction about the impact of ceasing exposure: excess risk will continue to increase (see the Appendix). Analysts who use the model agree that the excess risk of lung cancer freezes when persons stop smoking. The data show, however, that the excess risk drops. This finding provides fairly strong evidence against the model; other evidence is summarized in the Appendix.

The problems created for the model by ex-smokers seem to be well known; the issue is mentioned, for example, by Doll & Peto\textsuperscript{20}. The resolution attempted in Brown & Chu\textsuperscript{2}, much as we respect the authors, is not satisfying. The multistage model in that paper is fitted not to data but to output from logistic regressions, which are themselves inconsistent with the multistage model; the parameters of the fitted models are allowed to depend on dose; age and duration of smoking are treated as extra parameters-- constant across subjects-- and estimated, even though these data are available.
If risk of cancer at other sites drops after cessation of exposure, that would support the present findings on lung cancer. With smoking and bladder cancer, it seems to be generally accepted that risk drops when exposure ends\(^4\). For oral and pharyngeal cancer, Blot et al\(^9\) demonstrate a drop in risk when exposure ends. (There is also some supporting evidence from molecular biology\(^30\).)

Animal experiments also give some evidence; however, results depend on the test system. For skin painting, the risk drops after cessation of exposure\(^31-32\); likewise, some liver lesions can be repaired\(^33\). In the other direction, when dosing with 2-AAF stops, the risk of liver cancer continues to increase\(^34\). The suggestion is that for some human carcinogens, excess risk will decline after cessation of exposure; for others, excess risk will stabilize or increase.

The quality of the data we use here must be considered. In Figure 1 as well as the tables, years since quit includes time before follow-up starts. This approach may be questioned, in view of reporting errors in the data. However, if such errors are mainly random, the real decline in risk is even steeper. If there is a component of error systematically related to years since quit, an estimated decline rate may be biased, but an increase seems unlikely.
If errors at baseline are related to the cancer outcomes later, the sign becomes doubtful (likewise for reporting errors in confounders). For example, in the veterans cohort, the persons who quit shortly before baseline might be more likely to start smoking again, a phenomenon that could explain the higher risk for these ex-smokers. Such explanations, of course, are less tenable for the British doctors (where smoking status was determined three times) or the ACS cohort (with only 5 years of followup). Moreover, much of cancer epidemiology is based on retrospective determination of exposure, and it is hard to see why the data used here are weaker than other data routinely used in epidemiologic research or in risk assessment.

Narrower issues of data quality might be mentioned. The first line in Table 1 may be artifactually high, if people quit because they have cancer. The possibility of a similar artifact in line 1 of Table 3 is reduced since the ACS volunteers were instructed to recruit only healthy subjects--and apparently succeeded, at least for the 9-state study8. The last line in Tables 1 and 2 may be low in part from the deficit in events among older persons. Other ad hoc arguments, however, would have to be invoked to explain away other lines in these tables.

The decline in crude risks for ex-smokers reported in Tables 1 and 3 cannot be explained by the multistage model: see column 2 of Table 5 and Table 8. Nor are these data explained by the theory that excess risk freezes on cessation of smoking (column 2 of Tables 5 and 6). The data are, of course, explicable on the hypothesis that excess risk declines when smoking stops.
Declining excess risk is not compatible with the versions of the multistage model considered here. One modification of the model might be to incorporate a long and variable waiting time from malignancy to clinical endpoint; for a review of the evidence on this idea, see De Vita et al\textsuperscript{135}. Another patch might be to have rates of progression through the stages vary from person to person; for some evidence on this hypothesis, see Peto et al\textsuperscript{136}. A similar (but speculative) option is to use individual frailty parameters.

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. (This idea is incorporated into the model used by Gaffney & Altshuler\textsuperscript{19}.) A repair mechanism would explain the decline in excess risk after cessation of smoking: excess risk due to the possibility that certain lesions become cancerous would decline as such lesions were repaired.

Repair mechanisms are not compatible with the multistage model in standard form (which assumes a fixed order of progression through the stages). But repair is compatible with the autopsy results in Auerbach et al\textsuperscript{18}, who conclude

"Therefore, it seems virtually certain that the number of epithelial changes (particularly cells with atypical nuclei) decreases after cessation of cigarette smoking."

They go on to suggest that

"exposure to tobacco smoke may alter the local environment in bronchial epithelium in such a way as to favor the survival and reproduction of cells with atypical nuclei once such cells have been produced." [citation omitted]
Technical Appendix

The multistage model

In essence, the model says that a normal cell goes through a definite sequence of stages until it becomes cancerous. The order of the progression is fixed, so repair processes are excluded. Absent carcinogenic exposure (at zero dose), waiting times in the various stages are assumed to be independent, exponential random variables. So there is a background rate of progression through each stage, which may be different for the different stages. For details, see Whittemore & Keller\textsuperscript{37}. The discussion here follows Freedman & Navidi\textsuperscript{36-39}.

An animal or human organ is a collection of cells, and fails (gets cancer) when the first cell in the collection fails. Thus, the failure time for the organ is the minimum of the failure times of its component cells. (In some versions of the model, there is a waiting period from malignancy to detection.) Different cells are assumed to be independent with identically distributed failure times.
The next assumption: If a subject is exposed to a carcinogen like tobacco smoke, the rate of progression through the various stages increases in proportion to dose; the constant of proportionality depends on the stage. For the "insensitive" stages, this constant is zero; for the "sensitive" stages, the constant of proportionality is positive. The time spent in each stage is assumed to be much longer, on average, than the lifetime of the subject. The rates of progression through the various stages are assumed to be the same for all cells and all subjects. For modeling lung cancer, it is conventional to take the 1st and n-1st stages as the sensitive ones.

With a final assumption, independence of competing risks, the model can be used to generate a likelihood function for data; parameters such as the number of stages can be estimated by maximum likelihood. Then the adequacy of the fit can be assessed by a chi-squared test.

Equations (1-2-3) show the hazard function for non-smokers, current smokers, and ex-smokers: the dose is constant from start of smoking $T_0$ to end of smoking $T_1$. Current age is denoted $t$. There are $n$ stages, with the 1st and n-1st allowed to be sensitive. The coefficients $A,B,C,D$ are built up from the rates of progression through the various stages; they must be non-negative, and satisfy the constraint $AD=BC$. (The argument is given in Whittemore & Keller or Freedman & Navidi.)
(1) $A t^{n-1}$

(2) $A t^{n-1} + B \text{dose} \ (t^{n-1}-T_0^{n-1})$
   $+ C \text{dose} \ (t-T_0)^{n-1} + D \text{dose}^2 \ (t-T_0)^{n-1}$

(3) $A t^{n-1} + B \text{dose} \ (T_1^{n-1}-T_0^{n-1})$
   $+ C \text{dose} \ [(t-T_0)^{n-1}-(t-T_1)^{n-1}] + D \text{dose}^2 (T_1-T_0)^{n-1}$

Table 7 shows results for the three cohorts discussed here; with the British doctors and the veterans, the model is fitted on data for the current smokers and the non-smokers, and there are 21 degrees of freedom in the chi-squared. The ACS data are for current smokers only, and there are 13 degrees of freedom.\textsuperscript{38-39} The dose is in cigarettes per day.

----- Table 7 about here -----

For all three cohorts, a 6-stage model fits best by the chi-squared criterion, although 5 stages have been suggested before. The sensitivity of the stages differs across the cohorts: 1st and 5th for the doctors, 5th for the veterans, 1st for the ACS males. (In the ACS cohort, the 5th stage can be constrained to be insensitive with practically no change in the chi-squared statistic.)
For the British doctors, the model fits well. However, P<5/100,000 for the veterans; for the ACS males, P=1%. In the latter two cohorts, residuals from the model tend to be positive for intermediate ages and dose groups, but negative at the extremes. With the veterans data, the first stage is estimated as insensitive, and the model predicts little change in risk from varying age at start of smoking; that prediction is contradicted by the data, and is another indication of problems with the model.38-39

In the ACS cohort, the data on non-smokers arrived after we had fit the model on the current smokers using formula (2). As a cross-validation test, we predicted the risk for the non-smokers using formula (1). The prediction was 500 ± 60 with 99 observed (P<1/1,000,000). The model cannot extrapolate from smokers to non-smokers even though there are a substantial number of events in the dose group 1-9 cigarettes per day.

Of course, lack of fit in Table 7 might be due to problems with the data as well as the model, and the data sets are quite large. However, the failure in cross-validation must also be taken into account, and the disagreement among the cohorts as to which stages are sensitive.
The data on ex-smokers provide another test of the multistage model. Indeed, the model makes a strong qualitative prediction about behavior of risk when exposure stops. If the next-to-last stage only is sensitive, $B$ is positive and $C=D=0$. Then excess risk freezes when exposure stops: mathematically, the excess risk is

$$B \text{ dose } (T_{i}^{n-1}-T_{o}^{n-1})$$

This function does not depend on age $t$.

If the first stage is sensitive, then $C>0$; and excess risk increases even after exposure ends. Indeed, the excess risk in equation (3) is the sum of the $B$-, $C$-, and $D$-terms. The $B$- and $D$-terms do not involve age, and stabilize when exposure ends. However, the $C$-term is

$$C \text{ dose } [(t-T_{o})^{n-1}-(t-T_{i})^{n-1}]$$

This function increases with age, denoted by $t$ in the formula.

In sum, freezing of excess risk when exposure stops is incompatible with early-stage sensitivity. Nevertheless, early-stage sensitivity is needed for the model to predict a strong response of risk to age at start of smoking. On this point, there is a real conflict for the multistage theory.
Generally, whatever the sensitive stages may be, excess risk in the model cannot decline after smoking stops.\textsuperscript{42} Last-stage sensitivity is a special case: for example, with the first and last stage sensitive, risk drops immediately on cessation of smoking to that implied by a model with only the first stage sensitive. (This discontinuity is the usual argument against such models.) However, after cessation, the excess risk would start rising quite sharply with time, rather than falling.

The British doctors data on current smokers can be fit by having the first and last stage sensitive, rather than first and second-to-last. But this change would not really help with the ex-smokers, because it would not produce a continuing decline in excess risk. Likewise for the ACS data (Table 6).

The veterans current smokers can be fit with only the last stage being sensitive, just as well (or as badly) as with the penultimate stage. But then predicted risk for ex-smokers reverts to background immediately on cessation of smoking; column 3 of Table 5 shows that this is not a good option.
Using the model to compute expecteds

In Tables 5 and 6, the E's for column 1 are obtained from equation (2), by setting current age \( t \) equal to age at quitting \( T_1 \). The E's for column 2 are obtained by adding the term \( A(t^n-1-T_1^n-1) \), which represents the increase in background risk from time of quitting \( (T_1) \) to the present \( (t) \). Column 3 is computed from equation (1); for the ACS cohort, \( A \) is re-estimated from the non-smokers only, as 1.8.

For the veterans, \( C=D=0 \), as shown in Table 7; and the multistage model does predict the freezing of excess risk. So column 2 in Table 5 in fact uses the E's from the multistage model, providing a direct test of the model. To illustrate the procedure for computing E's, column 1 in Table 8 shows risks per 100,000 person-years for a non-smoker at various ages \( t \), computed from equation (1). Column 2 shows the risks for a continuing smoker at the same ages, computed from equation (2); for this illustration, age at start is \( T_0=19 \) and the dose is \( d=30 \) cigarettes per day. Column 3 shows the risks for an ex-smoker, computed from equation (3); age at start is \( T_0=19 \), age at quit is \( T_1=42 \), and the dose was \( d=30 \) cigarettes per day.

----- Table 8 around here ----
In column 1 of Table 5, this ex-smoker would have an expected risk at age 57 of 41/100,000, namely, his risk at time of quitting. In column 2, his expected risk at age 57 would be 11 + (41-2) = 50/100,000. Here, 11 is the risk for a non-smoker at age 57; and 41-2 is the subject's excess risk at time of quitting. In column 3, the expected risk at age 57 is 11, the risk for a non-smoker. As ex-smokers are aged through the study until they die, they contribute person-years and expecteds to the various lines of Table 5, in the usual way.

For the ACS cohort, C>0; so the E's underlying column 2 in Table 6 grow more slowly than the E's computed from equation (3): the E's for column 2 of Table 6 have excess risk frozen at quitting, while the E's from equation (3) incorporate increasing excess risk. A direct test of the model is provided in Table 8, which compares the O's with the E's from the model. Age at start was not ascertained for the ACS ex-smokers; we have imputed this several ways in Table 9; the value 19 years was the one used in Table 6. In any case, the pattern of decline in O/E is not much affected by this choice.

----- Table 9 about here -----
References


33. Farber, E and Sarma, DSR. Hepatocarcinogenesis: a dynamic cellular perspective. Laboratory Investigation 1987; 56: 4-22


Table 1. The veterans ex-smokers: observed lung cancer deaths. The observed number of cases per 100,000 person years declines steadily.

<table>
<thead>
<tr>
<th>years since quit</th>
<th>person-years</th>
<th>observed</th>
<th>crude rate per 100,000 py</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>15,693</td>
<td>26</td>
<td>166</td>
</tr>
<tr>
<td>5-9</td>
<td>33,633</td>
<td>45</td>
<td>134</td>
</tr>
<tr>
<td>10-14</td>
<td>41,786</td>
<td>52</td>
<td>124</td>
</tr>
<tr>
<td>15-19</td>
<td>35,008</td>
<td>25</td>
<td>71</td>
</tr>
<tr>
<td>20-24</td>
<td>27,878</td>
<td>11</td>
<td>39</td>
</tr>
<tr>
<td>25-29</td>
<td>21,844</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>30-34</td>
<td>13,426</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>35-39</td>
<td>4,212</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes. Persons who quit following doctor's orders" are excluded from the data. "Years since quit" is at recruitment into the study (1954 or 1957).
Table 2.  The veterans ex-smokers, directly standardized on dose and age at quitting (column 4).  Risk per 100,000 person years.  Reference group consists of all ex-smokers.  Columns 1, 2 and 3 show the weighted average dose, age at start, and age at quit (by person years) before standardization.  Dose is in cigarettes per day.

<table>
<thead>
<tr>
<th>Years since quit</th>
<th>Average dose</th>
<th>Average age at start</th>
<th>Average age at quitting</th>
<th>Standardized risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>23</td>
<td>21</td>
<td>53</td>
<td>87</td>
</tr>
<tr>
<td>5-9</td>
<td>22</td>
<td>21</td>
<td>52</td>
<td>98</td>
</tr>
<tr>
<td>10-14</td>
<td>22</td>
<td>21</td>
<td>50</td>
<td>88</td>
</tr>
<tr>
<td>15-19</td>
<td>21</td>
<td>21</td>
<td>49</td>
<td>74</td>
</tr>
<tr>
<td>20-24</td>
<td>19</td>
<td>20</td>
<td>43</td>
<td>48</td>
</tr>
<tr>
<td>25-29</td>
<td>17</td>
<td>20</td>
<td>41</td>
<td>16</td>
</tr>
<tr>
<td>30-34</td>
<td>15</td>
<td>20</td>
<td>39</td>
<td>520</td>
</tr>
<tr>
<td>35+</td>
<td>14</td>
<td>19</td>
<td>38</td>
<td>0</td>
</tr>
</tbody>
</table>

Notes.  The high risk for 30-34 stems from one event in a cell with 5.5 person-years, and may therefore be an artifact.
Table 3. Observed absolute risk for ACS male ex-smokers (number of lung cancer deaths per 100,000 person years) by years since quit smoking.

<table>
<thead>
<tr>
<th>years since quit</th>
<th>person-years</th>
<th>number of events</th>
<th>crude rate per 100,000 py</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>42,053</td>
<td>69</td>
<td>164</td>
</tr>
<tr>
<td>1-4</td>
<td>97,469</td>
<td>111</td>
<td>114</td>
</tr>
<tr>
<td>5-9</td>
<td>201,655</td>
<td>108</td>
<td>54</td>
</tr>
<tr>
<td>10+</td>
<td>66,566</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

Note. "Years since quit" is at the beginning of the study, in 1959.
Table 4. The ACS male ex-smokers, directly standardized on dose and age at quitting (column 3). Risk per 100,000 persons years. Reference group consists of those who quit 1-4 years before baseline. Columns 1 and 2 show the weighted average dose and age at quit (by person years) before standardization. Does is in cigarettes per day.

<table>
<thead>
<tr>
<th>years since quit</th>
<th>average dose</th>
<th>average age at quitting</th>
<th>standardized risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>29</td>
<td>52</td>
<td>158</td>
</tr>
<tr>
<td>1-4</td>
<td>28</td>
<td>52</td>
<td>114</td>
</tr>
<tr>
<td>5-9</td>
<td>27</td>
<td>47</td>
<td>83</td>
</tr>
<tr>
<td>10+</td>
<td>22</td>
<td>44</td>
<td>53</td>
</tr>
</tbody>
</table>

Notes. Line 1 is a special case; the age distribution for ex-smokers in this group is about the same as for the reference group, so the only adjustment was to reduce the rate by about 3%, compensating for the slightly heavier smoking (29 vs 28 cigarettes per day).
Figure 1. The British doctors. "Relationship between the incidence of bronchial carcinoma and time since cigarette smoking was stopped, compared with the relationship in continuing smokers and non-smokers."
The non-smokers and continuing smokers are matched to ex-smokers by age distribution (and dose at time of quitting, for ex-smokers).

Source: Redrawn from a figure in the Surgeon-General's report; also see Wynder & Hecht. The original figure seems to be from Doll; the caption and axes labels are quoted from that source.
Table 5. The veterans ex-smokers: the ratio of observed to expected lung cancer deaths. In the first column, the expected numbers are computed from the risk at time of quitting smoking. In the second column, the risk is computed on the basis of current age, with excess risk frozen from the time of quitting onwards. In the third column, the risk is computed from non-smokers of the same age.

<table>
<thead>
<tr>
<th>years since quit</th>
<th>E from risk at time of quitting</th>
<th>E from current age &amp; excess risk at quitting</th>
<th>E for non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>1.3</td>
<td>1.3</td>
<td>13.7</td>
</tr>
<tr>
<td>5-9</td>
<td>1.2</td>
<td>1.1</td>
<td>8.5</td>
</tr>
<tr>
<td>10-14</td>
<td>1.2</td>
<td>1.1</td>
<td>6.1</td>
</tr>
<tr>
<td>15-19</td>
<td>.9</td>
<td>.8</td>
<td>3.0</td>
</tr>
<tr>
<td>20-24</td>
<td>.9</td>
<td>.6</td>
<td>1.7</td>
</tr>
<tr>
<td>25-29</td>
<td>1.2</td>
<td>.6</td>
<td>1.0</td>
</tr>
<tr>
<td>30-34</td>
<td>1.6</td>
<td>.6</td>
<td>.9</td>
</tr>
<tr>
<td>35-39</td>
<td>.0</td>
<td>.0</td>
<td>.0</td>
</tr>
</tbody>
</table>

Notes. Years since quit is given in 5-year intervals; truncated midpoints are used in the calculation. For example, a person who quit 5-9 years before 1954 is assumed to have quit in 1947. The multistage model is used to smooth the risks.
Table 6. The ACS male ex-smokers: the ratio of observed to expected lung cancer deaths. In the first column, the expected numbers are computed from the risk at time of quitting smoking. In the second column, the risk is computed on the basis of current age, with excess risk frozen from the time of quitting onwards. In the third column, the risk is computed from non-smokers of the same age.

<table>
<thead>
<tr>
<th>Years since quit</th>
<th>E from risk at time &amp; excess risk of quitting</th>
<th>E for non-smokers</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>1.6</td>
<td>12.8</td>
</tr>
<tr>
<td>1-4</td>
<td>1.2</td>
<td>7.8</td>
</tr>
<tr>
<td>5-9</td>
<td>.9</td>
<td>3.5</td>
</tr>
<tr>
<td>10+</td>
<td>.3</td>
<td>.4</td>
</tr>
</tbody>
</table>

Notes. Age was reported at the beginning of the study, in 1959. The study period was 1960-65. By convention, a subject with e.g. age 45-49 in 1959 is taken as having age 47+4 in the study period. If this person quit 5-9 years before the start of the study, age at quit would be taken as 47-7; 10+ is taken as 15. The risks for non-smokers are smoothed separately, due to heterogeneity in the data.
Table 7. Results for the 3 cohorts: the best-fitting model has 6 stages; stages 1 and 5 are allowed to be sensitive. Standard errors are shown below the estimates; the estimates and standard errors should be divided by 103. Estimates reported as .00 were exactly 0; estimates constrained to 0 are marked by an asterisk.

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>chi-sq</th>
<th>comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.50</td>
<td>.45</td>
<td>1.16</td>
<td>.35</td>
<td>20</td>
<td>British doctors; fits; 1st &amp; 5th sensitive; 21 degrees of freedom</td>
</tr>
<tr>
<td>.54</td>
<td>.22</td>
<td>.88</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.75</td>
<td>1.01</td>
<td>.00</td>
<td>.00</td>
<td>50</td>
<td>Veterans; best fit, P=5/100,000; 5th stage only sensitive; 21 degrees of freedom</td>
</tr>
<tr>
<td>.13</td>
<td>.072</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.02</td>
<td>.028</td>
<td>2.52</td>
<td>.0077</td>
<td>28</td>
<td>ACS males; best fit, P=1%; 1st &amp; 5th sensitive; 13 degrees of freedom</td>
</tr>
<tr>
<td>1.07</td>
<td>.047</td>
<td>.73</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8.59</td>
<td>*</td>
<td>3.00</td>
<td>*</td>
<td>28</td>
<td>ACS males P=1%; 1st stage only sensitive</td>
</tr>
<tr>
<td>1.01</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 8. Expected risks (per 100,000 person years) of lung cancer at various ages for a non-smoker, a continuing smoker, and an ex-smoker, computed from the 6-stage model fitted to the veterans cohort. Age at start of smoking is 19, age at quit is 42, and dose is 30 cigarettes per day.

<table>
<thead>
<tr>
<th>Age</th>
<th>Non-smoker</th>
<th>Continuing smoker</th>
<th>Ex-smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>2</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td>47</td>
<td>4</td>
<td>73</td>
<td>43</td>
</tr>
<tr>
<td>52</td>
<td>7</td>
<td>121</td>
<td>46</td>
</tr>
<tr>
<td>57</td>
<td>11</td>
<td>192</td>
<td>49</td>
</tr>
<tr>
<td>62</td>
<td>16</td>
<td>293</td>
<td>55</td>
</tr>
<tr>
<td>67</td>
<td>24</td>
<td>432</td>
<td>62</td>
</tr>
<tr>
<td>72</td>
<td>34</td>
<td>619</td>
<td>73</td>
</tr>
</tbody>
</table>
Table 9. The ACS male ex-smokers: the ratio of observed to expected lung cancer deaths, expecteds being computed from the multistage model.

<table>
<thead>
<tr>
<th>years since quit</th>
<th>17</th>
<th>19</th>
<th>21</th>
<th>23</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;1</td>
<td>.95</td>
<td>1.08</td>
<td>1.22</td>
<td>1.36</td>
<td>1.51</td>
</tr>
<tr>
<td>1-4</td>
<td>.58</td>
<td>.65</td>
<td>.74</td>
<td>.82</td>
<td>.91</td>
</tr>
<tr>
<td>5-9</td>
<td>.26</td>
<td>.30</td>
<td>.33</td>
<td>.37</td>
<td>.41</td>
</tr>
<tr>
<td>10+</td>
<td>.04</td>
<td>.04</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Notes: 10+ is taken as 15. In a few cases, imputed age at start exceeds age at quit; risk is then computed at background rates.
TECHNICAL REPORTS
Statistics Department
University of California, Berkeley


17. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.


43. Bickel, P. J. and Yahav, J. A. (1985). On estimating the number of unseen species: how many executions were there?


56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.


64. O’SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.


71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.


77. O’SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.


90. Dabrowska, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
95. CANCELLED
100. Le Cam, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.
114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
131. Same as No. 140
134. CANCELLED
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
172. ADLER, R.J. and EPSTEIN, R. (September 1988). Intersection local times for infinite systems of planar brownian motions and for the brownian density process.


179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.


209. BREIMAN, L. (July 1989). Fitting additive models to regression data.


Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
University of California
Berkeley, California 94720

Cost: $1 per copy.