

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 carcinogenesis¹ 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 exposure²⁻⁶. 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. Hammond⁷ 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 & Horn⁸ report on the 9-state ACS cohort, and Kahn⁹ on the veterans. Data from Cederlof et al¹⁰ are compatible with present results, although the numbers are very small. In a large cohort of female ex-smokers, Garfinkel and Stelman¹¹ 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 al¹² 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 studies¹³⁻¹⁵ report similar results. We have found no other studies that report on the risk for ex-smokers by years since quit. The IARC⁴ reviews recent literature on the effects of smoking; for reviews of the earlier literature, see the US Public Health Service¹⁶⁻¹⁷.

Declining excess risk is contrary to the predictions of the Armitage-Doll multistage model¹. The decline is compatible with pathology results¹⁸, 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 Altshuler¹⁹; 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^{19a}.

Materials and methods

The three cohorts under consideration are as follows:

a) The British doctors. Doll & Peto²⁰ 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 & Peto²⁰ 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 published²¹⁻²³.

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, Doll²¹ 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 (Doll²¹, Wynder & Hecht²⁶ and the US Public Health Service¹⁷). 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 figure²² 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-Doll¹ 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 Agency²⁷. And it is often cited in discussions of the biological mechanisms of cancer; see Borszonyi et al²⁸. 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 & Peto²⁹. The resolution attempted in Brown & Chu², 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 ends⁴. For oral and pharyngeal cancer, Blot et al²⁹ demonstrate a drop in risk when exposure ends. (There is also some supporting evidence from molecular biology³⁰.)

Animal experiments also give some evidence; however, results depend on the test system. For skin painting, the risk drops after cessation of exposure³¹⁻³²; likewise, some liver lesions can be repaired³³. In the other direction, when dosing with 2-AAF stops, the risk of liver cancer continues to increase³⁴. 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 al³⁵. 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 al³⁶. 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 & Altshuler¹⁹.) 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 al¹⁸, 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 & Keller³⁷. The discussion here follows Freedman & Navidi³⁸⁻³⁹.

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 subject⁴⁰. 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^{2,3,20,41,42}.

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 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 & Keller³⁷ or Freedman & Navidi³⁸⁻³⁹.)

$$(1) \quad A t^{n-1}$$

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

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

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.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

$$B \text{ dose } (T_1^{n-1} - T_0^{n-1})$$

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

$$C \text{ dose } [(t-T_0)^{n-1} - (t-T_1)^{n-1}]$$

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.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 $A(t^{a-1} - T_1^{a-1})$, 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

1. Armitage P and Doll R. Stochastic models for carcinogenesis. In LeCam L and Neyman J, eds. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol IV. Berkeley: University of California Press, 1961, pp19-38.
2. Brown CC & Chu K. 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: 171S-179S.
3. Day NE. Epidemiological data and multistage carcinogenesis. In Borszonymi M, Day NE, Lapis K and Yamasaki H, eds. Models, Mechanisms and Etiology of Tumour Promotion. Lyon, France: International Agency for Research on Cancer, Sci Pub No 56, 1984, pp339-358; esp p349.
4. International Agency for Research on Cancer. Tobacco Smoking. Lyon, France: Mon 38, 1986. See esp. p208 on effect of age at start, p. 210 on lung cancer, and p. 258 on bladder cancer.
5. Peto J. Early- and late-stage carcinogenesis in mouse skin and in man. In Borszonymi M, Day NE, Lapis K and Yamasaki H, eds. Models, Mechanisms and Etiology of Tumour Promotion. Lyon, France: International Agency for Research on Cancer, Sci Pub No 56, 1984, pp359-372; esp p362.
6. Peto R. Epidemiology, multistage models, and short-term mutagenicity tests. In Hiatt HH, Watson JD & Winsten JA, eds. Origins of Human Cancer, Book C. Cold Spring Harbor, 1977, esp p1415.
7. Hammond EC. Smoking in relation to the death rates of one million men and women. In Haenszel W, ed. Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases. Washington, DC: NCI Mon 19, 1966.
8. Hammond EC & Horn D. Smoking & death rates-- report on forty-four months of follow-up of 187,766 men. JAMA 1958; 166: 1294-1308.
9. Kahn HA. The Dorn study of smoking and mortality among US veterans: report on eight and one-half years of observation. In Haenszel W, ed. Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases. Washington, DC: NCI Mon 19. (1966)

10. Cederlof R, Friberg L, Hrubec Z & Lorch U. The relationship of smoking and some social covariables in mortality and cancer morbidity: a ten year follow-up on a probability sample of 55,000 Swedish subjects age 18-69, Parts 1 and 2. Stockholm: Department of Environmental Hygiene, The Karolinska Institute, 1975.
11. Garfinkel L and Stellman SD. Smoking and lung cancer in women: findings in a prospective study. Cancer Research; 48: 6951-5.
12. 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. Br Med J 1984; 288: 1953-6.
13. Wicken AJ. Environmental and personal factors in lung cancer and bronchitis mortality in Northern Ireland 1960-2. London: Tobacco Research Council, 1966.
14. Wynder EL, Mabuchi K & Beattie Jr, EJ. The epidemiology of lung cancer. Recent trends. JAMA; 1970: 213 2221-8.
15. Schwartz D & Denoix PF. L'enquete francaise sur l'etiologie du cancer broncho-pulmonaire. Role du tabac. Sem Hop Paris 1957; 33: 3630-43.
16. US Public Health Service. Smoking and Health. Report of the advisory committee to the Surgeon General of the Public Health Service. Department of Health, Education and Welfare. Washington, DC, 1964.
17. US Public Health Service. Smoking and Health. Report of the advisory committee to the Surgeon General of the Public Health Service. Department of Health, Education and Welfare. Washington, DC, 1979.
18. Auerbach O, Stout AP, Hammond EC, & Garfinkel L. Bronchial epithelium in former smokers. N Engl J Med 1962; 267: 119-25.
19. Gaffney M & Altshuler B. Examination of the role of cigarette smoke in lung carcinogenesis using multistage models J Natl Cancer Inst 1988; 80: 925-31
- 19a. Moolgavkar, SH, Dewanji, A, & Luebeck, G. Cigarette smoking and lung cancer: reanalysis of the British Doctors' data. J Natl Cancer Inst 1989; 81: 415-20
20. Doll R & Peto R. Cigarette smoking and bronchial carcinoma among regular smokers and lifelong non-smokers. J Epidem and Comm Health 1978; 32: 303-13.
21. Doll R. Practical steps towards the prevention of bronchial carcinoma. Scot Med J 1970; 15: 433-47.

22. Doll R & Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. Br Med J 1964; 1: 1399-1410.
23. Doll R & Peto R. Mortality in relation to smoking: 20 years' observations on male British Doctors. Br Med J 1976; 2: 1525-36
24. Rogot, E. Smoking and mortality among US veterans. J Chronic Dis 1974; 27: 189-203.
25. Whittemore A . Effect of cigarette smoking in epidemiological studies of lung cancer. Statistics in Medecine 1988; 7: 223-238.
26. Wynder EL & Hecht S, eds. Lung Cancer. Geneva: UICC Technical Report Series, Vol 25, 1976.
27. US Environmental Protection Agency. Guidelines for carcinogenic risk assessment. Federal Register 1986; 51: 33992-34003.
28. Borszonyi M, Day NE, Lapis K and Yamasaki, H, eds. Models, Mechanisms and Etiology of Tumour Promotion. Lyon, France: International Agency for Research on Cancer, Sci Pub No 56, 1984.
29. WJ Blot, JK McLaughlin, DM Winn, DF Austin, RS Greenberg, S Preston-Martin, L Bernstein, JB Schoenber, A Stemhagen, JF Fraumeni, Jr. Smoking and drinking in relation oral and pharyngeal cancer. Cancer Research 1988; 48: 3282-7
30. DH Phillips, A Hewer, CN Martin, RC Garner & MM King. Correlation of DNA adduct levels in human lung with cigarette smoking. Nature 1988; 336: 790-2
31. Hennings H, Shores R, Wenk ML, Spangler EF, Tarone R, and Yuspa SH. Malignant conversion of mouse skin tumours is increased by tumour initiators and unaffected by tumour promoters. Nature 1983; 304: 67-68
32. McKenzie I & Rous P. The experimental disclosure of latent neoplastic changes in tarred skin. J Exp Med 1941; 73: 391-416
33. Farber, E and Sarma, DSR. Hepatocarcinogenesis: a dynamic cellular perspective. Laboratory Investigation 1987; 56: 4-22
34. Littlefield NA, Farmer JH, Gaylor DW and Sheldon WG. Effects of dose and time in a long-term, low dose carcinogenic study. J Environ Pathol Toxicol 1980; 3: 17-34.

35. DeVita Jr VT, Hellman S, & Rosenberg SA. Cancer: Principles & Practice of Oncology. 2nd ed. Philadelphia: Lippincott, 1985.
36. Peto R, Parish SE & Gray RG. There is no such thing as ageing, and cancer is not related to it. In Likhachev A, Ansimov V & Montesano R, eds. Age-Related Factors in Carcinogenesis. Lyon, France: International Agency for Research on Cancer, Sci Pub No 58, 1986.
37. Whittemore A & Keller J. Quantitative theories of carcinogenesis. SIAM Review 1978; 20: 1-30.
38. Freedman DA & Navidi W. On the multistage model for carcinogenesis. Technical report No 97, Department of Statistics, University of California, Berkeley, 1987.
39. Freedman DA & Navidi W. Multistage models for carcinogenesis. To appear in Environ Health Perspectives, 1989, vol 81.
40. Moolgavkar SH. The multistage theory of carcinogenesis and the age distribution of cancer in man. J Natl Cancer Inst 1978; 61: 49-52.
41. Breslow NE and Day NE. Statistical Methods in Cancer Research. Volume II-- The Design and Analysis of Cohort Studies. Lyon, France: The International Agency for Research on Cancer, Sci Publ No 82, 1987, esp secs 6.3-4.
42. Day NE & Brown C. Multistage models and primary prevention of cancer. J Natl Cancer Inst 1980; 64: 977-89.

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

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

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.

years since quit	average dose	average age at start	average age at quitting	standardized risk
0-4	23	21	53	87
5-9	22	21	52	98
10-14	22	21	50	88
15-19	21	21	49	74
20-24	19	20	43	48
25-29	17	20	41	16
30-34	15	20	39	520
35+	14	19	38	0

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.

years since quit	person- years	number of events	crude rate per 100,000 py
<1	42,053	69	164
1-4	97,469	111	114
5-9	201,655	108	54
10+	66,566	6	9

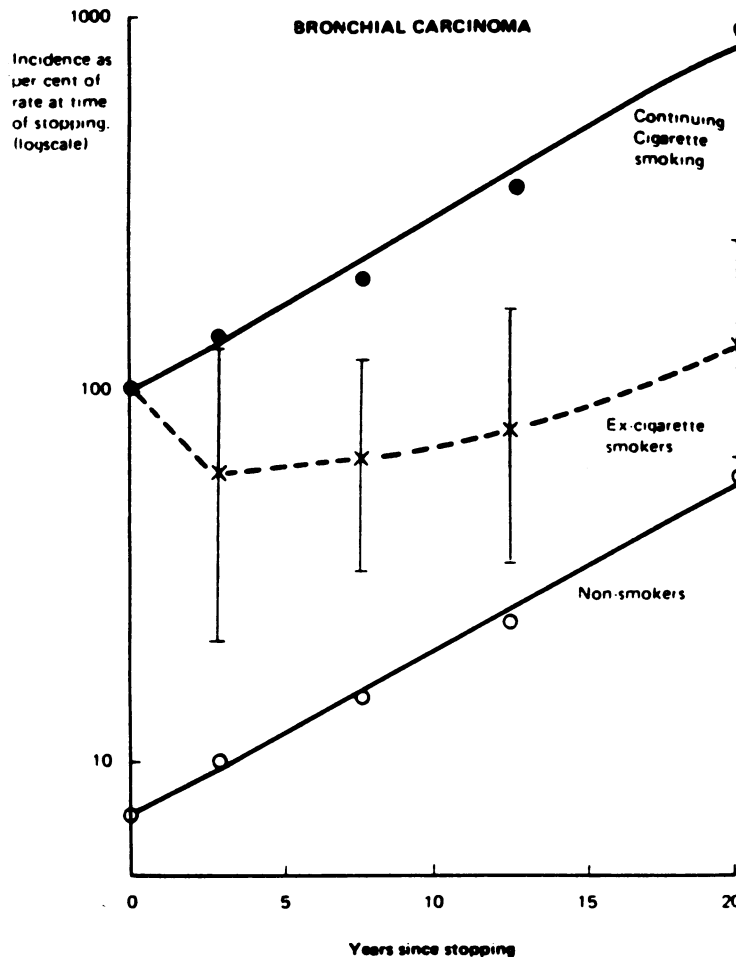
Note. "Years since quit" is at the beginning of
the study, in 1959.

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.

years since quit	average dose	average age at quitting	standardized risk
<1	29	52	158
1-4	28	52	114
5-9	27	47	83
10+	22	44	53

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 report¹⁷; also see Wynder & Hecht²⁶. The original figure seems to be from Doll²¹; 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.

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

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.

years since quit	Observed/Expected		
	E from risk at time of quitting	E from current age & excess risk at quitting	E for non-smokers
<1	1.6	1.3	12.8
1-4	1.2	.9	7.8
5-9	.9	.5	3.5
10+	.3	.08	.4

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.

A	B	C	D	chi-sq	comment
1.50 .54	.45 .22	1.16 .88	.35	20	British doctors; fits; 1st & 5th sensitive; 21 degrees of freedom
1.75 .13	1.01 .072	.00 .02	.00	50	Veterans; best fit, P=5/100,000; 5th stage only sensitive; 21 degrees of freedom
9.02 1.07	.028 .047	2.52 .73	.0077	28	ACS males; best fit, P=1%; 1st & 5th sensitive; 13 degrees of freedom
8.59 1.01	* 	3.00 .23	*	28	ACS males P=1%; 1st stage only sensitive

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.

Age	Non-smoker	Continuing smoker	Ex-smoker
42	2	41	41
47	4	73	43
52	7	121	46
57	11	192	49
62	16	293	55
67	24	432	62
72	34	619	73

Table 9. The ACS male ex-smokers: the ratio of observed to expected lung cancer deaths, expecteds being computed from the multistage model.

years since quit	Observed/Expected, when imputed age at start is				
	17	19	21	23	25
<1	.95	1.08	1.22	1.36	1.51
1-4	.58	.65	.74	.82	.91
5-9	.26	.30	.33	.37	.41
10+	.04	.04	.05	.05	.05

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

1. BREIMAN, L. and FREEDMAN, D. (Nov. 1981, revised Feb. 1982). How many variables should be entered in a regression equation? Jour. Amer. Statist. Assoc., March 1983, 78, No. 381, 131-136.
2. BRILLINGER, D. R. (Jan. 1982). Some contrasting examples of the time and frequency domain approaches to time series analysis. Time Series Methods in Hydrosiences, (A. H. El-Shaarawi and S. R. Esterby, eds.) Elsevier Scientific Publishing Co., Amsterdam, 1982, pp. 1-15.
3. DOKSUM, K. A. (Jan. 1982). On the performance of estimates in proportional hazard and log-linear models. Survival Analysis, (John Crowley and Richard A. Johnson, eds.) IMS Lecture Notes - Monograph Series, (Shanti S. Gupta, series ed.) 1982, 74-84.
4. BICKEL, P. J. and BREIMAN, L. (Feb. 1982). Sums of functions of nearest neighbor distances, moment bounds, limit theorems and a goodness of fit test. Ann. Prob., Feb. 1982, 11, No. 1, 185-214.
5. BRILLINGER, D. R. and TUKEY, J. W. (March 1982). Spectrum estimation and system identification relying on a Fourier transform. The Collected Works of J. W. Tukey, vol. 2, Wadsworth, 1985, 1001-1141.
6. BERAN, R. (May 1982). Jackknife approximation to bootstrap estimates. Ann. Statist., March 1984, 12 No. 1, 101-118.
7. BICKEL, P. J. and FREEDMAN, D. A. (June 1982). Bootstrapping regression models with many parameters. Lehmann Festschrift, (P. J. Bickel, K. Doksum and J. L. Hodges, Jr., eds.) Wadsworth Press, Belmont, 1983, 28-48.
8. BICKEL, P. J. and COLLINS, J. (March 1982). Minimizing Fisher information over mixtures of distributions. Sankhyā, 1983, 45, Series A, Pt. 1, 1-19.
9. BREIMAN, L. and FRIEDMAN, J. (July 1982). Estimating optimal transformations for multiple regression and correlation.
10. FREEDMAN, D. A. and PETERS, S. (July 1982, revised Aug. 1983). Bootstrapping a regression equation: some empirical results. JASA, 1984, 79, 97-106.
11. EATON, M. L. and FREEDMAN, D. A. (Sept. 1982). A remark on adjusting for covariates in multiple regression.
12. BICKEL, P. J. (April 1982). Minimax estimation of the mean of a mean of a normal distribution subject to doing well at a point. Recent Advances in Statistics, Academic Press, 1983.
14. FREEDMAN, D. A., ROTHENBERG, T. and SUTCH, R. (Oct. 1982). A review of a residential energy end use model.
15. BRILLINGER, D. and PREISLER, H. (Nov. 1982). Maximum likelihood estimation in a latent variable problem. Studies in Econometrics, Time Series, and Multivariate Statistics, (eds. S. Karlin, T. Amemiya, L. A. Goodman). Academic Press, New York, 1983, pp. 31-65.
16. BICKEL, P. J. (Nov. 1982). Robust regression based on infinitesimal neighborhoods. Ann. Statist., Dec. 1984, 12, 1349-1368.
17. DRAPER, D. C. (Feb. 1983). Rank-based robust analysis of linear models. I. Exposition and review. Statistical Science, 1988, Vol.3 No. 2 239-271.
18. DRAPER, D. C. (Feb 1983). Rank-based robust inference in regression models with several observations per cell.
19. FREEDMAN, D. A. and FIENBERG, S. (Feb. 1983, revised April 1983). Statistics and the scientific method, Comments on and reactions to Freedman, A rejoinder to Fienberg's comments. Springer New York 1985 Cohort Analysis in Social Research, (W. M. Mason and S. E. Fienberg, eds.).
20. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Jan. 1984). Using the bootstrap to evaluate forecasting equations. J. of Forecasting, 1985, Vol. 4, 251-262.
21. FREEDMAN, D. A. and PETERS, S. C. (March 1983, revised Aug. 1983). Bootstrapping an econometric model: some empirical results. JBES, 1985, 2, 150-158.
22. FREEDMAN, D. A. (March 1983). Structural-equation models: a case study.
23. DAGGETT, R. S. and FREEDMAN, D. (April 1983, revised Sept. 1983). Econometrics and the law: a case study in the proof of antitrust damages. Proc. of the Berkeley Conference, in honor of Jerzy Neyman and Jack Kiefer. Vol I pp. 123-172. (L. Le Cam, R. Olshen eds.) Wadsworth, 1985.

24. DOKSUM, K. and YANDELL, B. (April 1983). Tests for exponentiality. Handbook of Statistics, (P. R. Krishnaiah and P. K. Sen, eds.) 4, 1984, 579-611.
25. FREEDMAN, D. A. (May 1983). Comments on a paper by Markus.
26. FREEDMAN, D. (Oct. 1983, revised March 1984). On bootstrapping two-stage least-squares estimates in stationary linear models. Ann. Statist., 1984, 12, 827-842.
27. DOKSUM, K. A. (Dec. 1983). An extension of partial likelihood methods for proportional hazard models to general transformation models. Ann. Statist., 1987, 15, 325-345.
28. BICKEL, P. J., GOETZE, F. and VAN ZWET, W. R. (Jan. 1984). A simple analysis of third order efficiency of estimate Proc. of the Neyman-Kiefer Conference, (L. Le Cam, ed.) Wadsworth, 1985.
29. BICKEL, P. J. and FREEDMAN, D. A. Asymptotic normality and the bootstrap in stratified sampling. Ann. Statist. 12 470-482.
30. FREEDMAN, D. A. (Jan. 1984). The mean vs. the median: a case study in 4-R Act litigation. JBES, 1985 Vol 3 pp. 1-13.
31. STONE, C. J. (Feb. 1984). An asymptotically optimal window selection rule for kernel density estimates. Ann. Statist., Dec. 1984, 12, 1285-1297.
32. BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.
33. STONE, C. J. (Oct. 1984). Additive regression and other nonparametric models. Ann. Statist., 1985, 13, 689-705.
34. STONE, C. J. (June 1984). An asymptotically optimal histogram selection rule. Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer (L. Le Cam and R. A. Olshen, eds.), II, 513-520.
35. FREEDMAN, D. A. and NAVIDI, W. C. (Sept. 1984, revised Jan. 1985). Regression models for adjusting the 1980 Census. Statistical Science, Feb 1986, Vol. 1, No. 1, 3-39.
36. FREEDMAN, D. A. (Sept. 1984, revised Nov. 1984). De Finetti's theorem in continuous time.
37. DIACONIS, P. and FREEDMAN, D. (Oct. 1984). An elementary proof of Stirling's formula. Amer. Math Monthly, Feb 1986, Vol. 93, No. 2, 123-125.
38. LE CAM, L. (Nov. 1984). Sur l'approximation de familles de mesures par des familles Gaussiennes. Ann. Inst. Henri Poincaré, 1985, 21, 225-287.
39. DIACONIS, P. and FREEDMAN, D. A. (Nov. 1984). A note on weak star uniformities.
40. BREIMAN, L. and IHAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.
41. STONE, C. J. (Jan. 1985). The dimensionality reduction principle for generalized additive models.
42. LE CAM, L. (Jan. 1985). On the normal approximation for sums of independent variables.
43. BICKEL, P. J. and YAHAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?
44. BRILLINGER, D. R. (1985). The natural variability of vital rates and associated statistics. Biometrics, to appear.
45. BRILLINGER, D. R. (1985). Fourier inference: some methods for the analysis of array and nonGaussian series data. Water Resources Bulletin, 1985, 21, 743-756.
46. BREIMAN, L. and STONE, C. J. (1985). Broad spectrum estimates and confidence intervals for tail quantiles.
47. DABROWSKA, D. M. and DOKSUM, K. A. (1985, revised March 1987). Partial likelihood in transformation models with censored data. Scandinavian J. Statist., 1988, 15, 1-23.
48. HAYCOCK, K. A. and BRILLINGER, D. R. (November 1985). LIBDRB: A subroutine library for elementary time series analysis.
49. BRILLINGER, D. R. (October 1985). Fitting cosines: some procedures and some physical examples. Joshi Festschrift, 1986. D. Reidel.
50. BRILLINGER, D. R. (November 1985). What do seismology and neurophysiology have in common? - Statistics! Comptes Rendus Math. Rep. Acad. Sci. Canada, January, 1986.
51. COX, D. D. and O'SULLIVAN, F. (October 1985). Analysis of penalized likelihood-type estimators with application to generalized smoothing in Sobolev Spaces.

52. O'SULLIVAN, F. (November 1985). A practical perspective on ill-posed inverse problems: A review with some new developments. To appear in Journal of Statistical Science.
53. LE CAM, L. and YANG, G. L. (November 1985, revised March 1987). On the preservation of local asymptotic normality under information loss.
54. BLACKWELL, D. (November 1985). Approximate normality of large products.
55. FREEDMAN, D. A. (June 1987). As others see us: A case study in path analysis. Journal of Educational Statistics. 12, 101-128.
56. LE CAM, L. and YANG, G. L. (January 1986). Replaced by No. 68.
57. LE CAM, L. (February 1986). On the Bernstein - von Mises theorem.
58. O'SULLIVAN, F. (January 1986). Estimation of Densities and Hazards by the Method of Penalized likelihood.
59. ALDOUS, D. and DIACONIS, P. (February 1986). Strong Uniform Times and Finite Random Walks.
60. ALDOUS, D. (March 1986). On the Markov Chain simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.
61. CHENG, C-S. (April 1986). An Optimization Problem with Applications to Optimal Design Theory.
62. CHENG, C-S., MAJUMDAR, D., STUFKEN, J. & TURE, T. E. (May 1986, revised Jan 1987). Optimal step type design for comparing test treatments with a control.
63. CHENG, C-S. (May 1986, revised Jan. 1987). An Application of the Kiefer-Wolfowitz Equivalence Theorem.
64. O'SULLIVAN, F. (May 1986). Nonparametric Estimation in the Cox Proportional Hazards Model.
65. ALDOUS, D. (JUNE 1986). Finite-Time Implications of Relaxation Times for Stochastically Monotone Processes.
66. PITMAN, J. (JULY 1986, revised November 1986). Stationary Excursions.
67. DABROWSKA, D. and DOKSUM, K. (July 1986, revised November 1986). Estimates and confidence intervals for median and mean life in the proportional hazard model with censored data. Biometrika, 1987, 74, 799-808.
68. LE CAM, L. and YANG, G.L. (July 1986). Distinguished Statistics, Loss of information and a theorem of Robert B. Davies (Fourth edition).
69. STONE, C.J. (July 1986). Asymptotic properties of logspline density estimation.
71. BICKEL, P.J. and YAHAV, J.A. (July 1986). Richardson Extrapolation and the Bootstrap.
72. LEHMANN, E.L. (July 1986). Statistics - an overview.
73. STONE, C.J. (August 1986). A nonparametric framework for statistical modelling.
74. BIANE, PH. and YOR, M. (August 1986). A relation between Lévy's stochastic area formula, Legendre polynomial, and some continued fractions of Gauss.
75. LEHMANN, E.L. (August 1986, revised July 1987). Comparing Location Experiments.
76. O'SULLIVAN, F. (September 1986). Relative risk estimation.
77. O'SULLIVAN, F. (September 1986). Deconvolution of episodic hormone data.
78. PITMAN, J. & YOR, M. (September 1987). Further asymptotic laws of planar Brownian motion.
79. FREEDMAN, D.A. & ZEISEL, H. (November 1986). From mouse to man: The quantitative assessment of cancer risks. To appear in Statistical Science.
80. BRILLINGER, D.R. (October 1986). Maximum likelihood analysis of spike trains of interacting nerve cells.
81. DABROWSKA, D.M. (November 1986). Nonparametric regression with censored survival time data.
82. DOKSUM, K.J. and LO, A.Y. (Nov 1986, revised Aug 1988). Consistent and robust Bayes Procedures for Location based on Partial Information.
83. DABROWSKA, D.M., DOKSUM, K.A. and MIURA, R. (November 1986). Rank estimates in a class of semiparametric two-sample models.

84. BRILLINGER, D. (December 1986). Some statistical methods for random process data from seismology and neurophysiology.
85. DIACONIS, P. and FREEDMAN, D. (December 1986). A dozen de Finetti-style results in search of a theory. Ann. Inst. Henri Poincaré, 1987, 23, 397-423.
86. DABROWSKA, D.M. (January 1987). Uniform consistency of nearest neighbour and kernel conditional Kaplan - Meier estimates.
87. FREEDMAN, D.A., NAVIDI, W. and PETERS, S.C. (February 1987). On the impact of variable selection in fitting regression equations.
88. ALDOUS, D. (February 1987, revised April 1987). Hashing with linear probing, under non-uniform probabilities.
89. DABROWSKA, D.M. and DOKSUM, K.A. (March 1987, revised January 1988). Estimating and testing in a two sample generalized odds rate model. J. Amer. Statist. Assoc., 1988, 83, 744-749.
90. DABROWSKA, D.M. (March 1987). Rank tests for matched pair experiments with censored data.
91. DIACONIS, P and FREEDMAN, D.A. (April 1988). Conditional limit theorems for exponential families and finite versions of de Finetti's theorem. To appear in the Journal of Applied Probability.
92. DABROWSKA, D.M. (April 1987, revised September 1987). Kaplan-Meier estimate on the plane.
- 92a. ALDOUS, D. (April 1987). The Harmonic mean formula for probabilities of Unions: Applications to sparse random graphs.
93. DABROWSKA, D.M. (June 1987, revised Feb 1988). Nonparametric quantile regression with censored data.
94. DONOHO, D.L. & STARK, P.B. (June 1987). Uncertainty principles and signal recovery.
95. CANCELLED
96. BRILLINGER, D.R. (June 1987). Some examples of the statistical analysis of seismological data. To appear in Proceedings, Centennial Anniversary Symposium, Seismographic Stations, University of California, Berkeley.
97. FREEDMAN, D.A. and NAVIDI, W. (June 1987). On the multi-stage model for carcinogenesis. To appear in Environmental Health Perspectives.
98. O'SULLIVAN, F. and WONG, T. (June 1987). Determining a function diffusion coefficient in the heat equation.
99. O'SULLIVAN, F. (June 1987). Constrained non-linear regularization with application to some system identification problems.
100. LE CAM, L. (July 1987, revised Nov 1987). On the standard asymptotic confidence ellipsoids of Wald.
101. DONOHO, D.L. and LIU, R.C. (July 1987). Pathologies of some minimum distance estimators. Annals of Statistics, June, 1988.
102. BRILLINGER, D.R., DOWNING, K.H. and GLAESER, R.M. (July 1987). Some statistical aspects of low-dose electron imaging of crystals.
103. LE CAM, L. (August 1987). Harald Cramér and sums of independent random variables.
104. DONOHO, A.W., DONOHO, D.L. and GASKO, M. (August 1987). Macspin: Dynamic graphics on a desktop computer. IEEE Computer Graphics and applications, June, 1988.
105. DONOHO, D.L. and LIU, R.C. (August 1987). On minimax estimation of linear functionals.
106. DABROWSKA, D.M. (August 1987). Kaplan-Meier estimate on the plane: weak convergence, LIL and the bootstrap.
107. CHENG, C-S. (Aug 1987, revised Oct 1988). Some orthogonal main-effect plans for asymmetrical factorials.
108. CHENG, C-S. and JACROUX, M. (August 1987). On the construction of trend-free run orders of two-level factorial designs.
109. KLASS, M.J. (August 1987). Maximizing $E \max_{1 \leq k \leq n} S_k^+ / ES_n^+$: A prophet inequality for sums of I.I.D. mean zero variates.
110. DONOHO, D.L. and LIU, R.C. (August 1987). The "automatic" robustness of minimum distance functionals. Annals of Statistics, June, 1988.
111. BICKEL, P.J. and GHOSH, J.K. (August 1987, revised June 1988). A decomposition for the likelihood ratio statistic and the Bartlett correction — a Bayesian argument.

112. BURDZY, K., PITMAN, J.W. and YOR, M. (September 1987). Some asymptotic laws for crossings and excursions.
113. ADHIKARI, A. and PITMAN, J. (September 1987). The shortest planar arc of width 1.
114. RITOV, Y. (September 1987). Estimation in a linear regression model with censored data.
115. BICKEL, P.J. and RITOV, Y. (Sept. 1987, revised Aug 1988). Large sample theory of estimation in biased sampling regression models I.
116. RITOV, Y. and BICKEL, P.J. (Sept. 1987, revised Aug. 1988). Achieving information bounds in non and semiparametric models.
117. RITOV, Y. (October 1987). On the convergence of a maximal correlation algorithm with alternating projections.
118. ALDOUS, D.J. (October 1987). Meeting times for independent Markov chains.
119. HESSE, C.H. (October 1987). An asymptotic expansion for the mean of the passage-time distribution of integrated Brownian Motion.
120. DONOHO, D. and LIU, R. (Oct. 1987, revised Mar. 1988, Oct. 1988). Geometrizing rates of convergence, II.
121. BRILLINGER, D.R. (October 1987). Estimating the chances of large earthquakes by radiocarbon dating and statistical modelling. Statistics a Guide to the Unknown, pp. 249-260 (Eds. J.M. Tanur et al.) Wadsworth, Pacific Grove.
122. ALDOUS, D., FLANNERY, B. and PALACIOS, J.L. (November 1987). Two applications of urn processes: The fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains.
123. DONOHO, D.L., MACGIBBON, B. and LIU, R.C. (Nov. 1987, revised July 1988). Minimax risk for hyperrectangles.
124. ALDOUS, D. (November 1987). Stopping times and tightness II.
125. HESSE, C.H. (November 1987). The present state of a stochastic model for sedimentation.
126. DALANG, R.C. (December 1987, revised June 1988). Optimal stopping of two-parameter processes on nonstandard probability spaces.
127. Same as No. 133.
128. DONOHO, D. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean II.
129. SMITH, D.L. (December 1987). Exponential bounds in Vapnik-Červonenkis classes of index 1.
130. STONE, C.J. (Nov. 1987, revised Sept. 1988). Uniform error bounds involving logspline models.
131. Same as No. 140
132. HESSE, C.H. (Dec. 1987, revised June 1989). A Bahadur - Type representation for empirical quantiles of a large class of stationary, possibly infinite - variance, linear processes
133. DONOHO, D.L. and GASKO, M. (December 1987). Multivariate generalizations of the median and trimmed mean, I.
134. CANCELLED
135. FREEDMAN, D.A. and NAVIDI, W. (December 1987). On the risk of lung cancer for ex-smokers.
136. LE CAM, L. (January 1988). On some stochastic models of the effects of radiation on cell survival.
137. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the uniform consistency of Bayes estimates for multinomial probabilities.
- 137a. DONOHO, D.L. and LIU, R.C. (1987). Geometrizing rates of convergence, I.
138. DONOHO, D.L. and LIU, R.C. (January 1988). Geometrizing rates of convergence, III.
139. BERAN, R. (January 1988). Refining simultaneous confidence sets.
140. HESSE, C.H. (December 1987). Numerical and statistical aspects of neural networks.
141. BRILLINGER, D.R. (Jan. 1988). Two reports on trend analysis: a) An elementary trend analysis of Rio negro levels at Manaus, 1903-1985. b) Consistent detection of a monotonic trend superposed on a stationary time series.
142. DONOHO, D.L. (Jan. 1985, revised Jan. 1988). One-sided inference about functionals of a density.

143. DALANG, R.C. (Feb. 1988, revised Nov. 1988). Randomization in the two-armed bandit problem.
144. DABROWSKA, D.M., DOKSUM, K.A. and SONG, J.K. (February 1988). Graphical comparisons of cumulative hazards for two populations.
145. ALDOUS, D.J. (February 1988). Lower bounds for covering times for reversible Markov Chains and random walks on graphs.
146. BICKEL, P.J. and RITOV, Y. (Feb.1988, revised August 1988). Estimating integrated squared density derivatives.
147. STARK, P.B. (March 1988). Strict bounds and applications.
148. DONOHO, D.L. and STARK, P.B. (March 1988). Rearrangements and smoothing.
149. NOLAN, D. (Sept. 1989). On min-max majority and deepest points.
150. SEILLIER, F. (March 1988). Sequential probability forecasts and the probability integral transform.
151. NOLAN, D. (Mar. 1988, revised May 1989). Asymptotics for multivariate trimming.
152. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On a theorem of Kuchler and Lauritzen.
153. DIACONIS, P. and FREEDMAN, D.A. (April 1988). On the problem of types.
154. DOKSUM, K.A. and GASKO, M. (May 1988). On the correspondence between models in binary regression analysis and survival analysis.
155. LEHMANN, E.L. (May 1988). Jerzy Neyman, 1894-1981.
156. ALDOUS, D.J. (May 1988). Stein's method in a two-dimensional coverage problem.
157. FAN, J. (June 1988). On the optimal rates of convergence for nonparametric deconvolution problem.
158. DABROWSKA, D. (June 1988). Signed-rank tests for censored matched pairs.
159. BERAN, R.J. and MILLAR, P.W. (June 1988). Multivariate symmetry models.
160. BERAN, R.J. and MILLAR, P.W. (June 1988). Tests of fit for logistic models.
161. BREIMAN, L. and PETERS, S. (June 1988). Comparing automatic bivariate smoothers (A public service enterprise).
162. FAN, J. (June 1988). Optimal global rates of convergence for nonparametric deconvolution problem.
163. DIACONIS, P. and FREEDMAN, D.A. (June 1988). A singular measure which is locally uniform. (Revised by Tech Report No. 180).
164. BICKEL, P.J. and KRIEGER, A.M. (July 1988). Confidence bands for a distribution function using the bootstrap.
165. HESSE, C.H. (July 1988). New methods in the analysis of economic time series I.
166. FAN, JIANQING (July 1988). Nonparametric estimation of quadratic functionals in Gaussian white noise.
167. BREIMAN, L., STONE, C.J. and KOOPERBERG, C. (August 1988). Confidence bounds for extreme quantiles.
168. LE CAM, L. (August 1988). Maximum likelihood an introduction.
169. BREIMAN, L. (Aug.1988, revised Feb. 1989). Submodel selection and evaluation in regression I. The X-fixed case and little bootstrap.
170. LE CAM, L. (September 1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge.
171. STONE, C.J. (September 1988). Large-sample inference for logspline models.
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.
173. MILLAR, P.W. (October 1988). Optimal estimation in the non-parametric multiplicative intensity model.
174. YOR, M. (October 1988). Interwinings of Bessel processes.
175. ROJO, J. (October 1988). On the concept of tail-heaviness.

176. ABRAHAM, D.M. and RIZZARDI, F. (September 1988). BLSS - The Berkeley interactive statistical system: An overview.
177. MILLAR, P.W. (October 1988). Gamma-funnels in the domain of a probability, with statistical implications.
178. DONOHO, D.L. and LIU, R.C. (October 1988). Hardest one-dimensional subproblems.
179. DONOHO, D.L. and STARK, P.B. (October 1988). Recovery of sparse signal when the low frequency information is missing.
180. FREEDMAN, D.A. and PITMAN, J.A. (Nov. 1988). A measure which is singular and uniformly locally uniform. (Revision of Tech Report No. 163).
181. DOKSUM, K.A. and HOYLAND, ARNLJOT (Nov. 1988, revised Jan. 1989, Aug. '89). Models for variable stress accelerated life testing experiments based on Wiener processes and the inverse Gaussian distribution.
182. DALANG, R.C., MORTON, A. and WILLINGER, W. (November 1988). Equivalent martingale measures and no-arbitrage in stochastic securities market models.
183. BERAN, R. (November 1988). Calibrating prediction regions.
184. BARLOW, M.T., PITMAN, J. and YOR, M. (Feb. 1989). On Walsh's Brownian Motions.
185. DALANG, R.C. and WALSH, J.B. (Dec. 1988). Almost-equivalence of the germ-field Markov property and the sharp Markov property of the Brownian sheet.
186. HESSE, C.H. (Dec. 1988). Level-Crossing of integrated Ornstein-Uhlenbeck processes
187. NEVEU, J. and PITMAN, J.W. (Feb. 1989). Renewal property of the extrema and tree property of the excursion of a one-dimensional brownian motion.
188. NEVEU, J. and PITMAN, J.W. (Feb. 1989). The branching process in a brownian excursion.
189. PITMAN, J.W. and YOR, M. (Mar. 1989). Some extensions of the arcsine law.
190. STARK, P.B. (Dec. 1988). Duality and discretization in linear inverse problems.
191. LEHMANN, E.L. and SCHOLZ, F.W. (Jan. 1989). Ancillarity.
192. PEMANTLE, R. (Feb. 1989). A time-dependent version of Pólya's urn.
193. PEMANTLE, R. (Feb. 1989). Nonconvergence to unstable points in urn models and stochastic approximations.
194. PEMANTLE, R. (Feb. 1989, revised May 1989). When are touchpoints limits for generalized Pólya urns.
195. PEMANTLE, R. (Feb. 1989). Random walk in a random environment and first-passage percolation on trees.
196. BARLOW, M., PITMAN, J. and YOR, M. (Feb. 1989). Une extension multidimensionnelle de la loi de l'arc sinus.
197. BREIMAN, L. and SPECTOR, P. (Mar. 1989). Submodel selection and evaluation in regression — the X-random case.
198. BREIMAN, L., TSUR, Y. and ZEMEL, A. (Mar. 1989). A simple estimation procedure for censored regression models with known error distribution.
199. BRILLINGER, D.R. (Mar. 1989). Two papers on bilinear systems: a) A study of second- and third-order spectral procedures and maximum likelihood identification of a bilinear system. b) Some statistical aspects of NMR spectroscopy, *Actas del 2º congreso latinoamericano de probabilidad y estadística matemática*, Caracas, 1985.
200. BRILLINGER, D.R. (Mar. 1989). Two papers on higher-order spectra: a) Parameter estimation for nonGaussian processes via second and third order spectra with an application to some endocrine data. b) Some history of the study of higher-order moments and spectra.
201. DE LA PENA, V. and KLASS, M.J. (April 1989). L bounds for quadratic forms of independent random variables.
202. FREEDMAN, D.A. and NAVIDI, W.C. (April 1989). Testing the independence of competing risks.
203. TERDIK, G. (May 1989). Bilinear state space realization for polynomial stochastic systems.
204. DONOHO, D.L. and JOHNSTONE, I.M. (May 1989). Minimax risk over l_p -Balls.
205. PEMANTLE, R., PROPP, J. and ULLMAN, D. (May 1989). On tensor powers of integer programs.

- 206. MILASEVIC, P. and NOLAN, D. (May 1989). Estimation on the sphere: A geometric approach.
- 207. SPEED, T.P. and YU, B. (July 1989, rev. Oct. 1989). Stochastic complexity and model selection: normal regression.
- 208. DUBINS, L.E. (June 1989). A group decision device: Its pareto-like optimality.
- 209. BREIMAN, L. (July 1989). Fitting additive models to regression data.
- 210. PEMANTLE, R. (July 1989). Vertex-reinforced random walk.
- 211. LE CAM, L. (August 1989). On measurability and convergence in distribution.
- 212. FELDMAN, R.E. (July 1989). Autoregressive processes and first-hit probabilities for randomized random walks.
- 213. DONOHO, D.L., JOHNSTONE, I.M., HOCH, J.C. and STERN, A.S. (August 1989). Maximum entropy and the nearly black object.
- 214. DONOHO, D.L. (August, 1989). Statistical estimation and optimal recovery.
- 215. STONE, C. (August, 1989). Asymptotics for doubly-flexible logspline response models.
- 216. NOLAN, D. (August 1989). The excess mass ellipsoid.
- 217. FREEDMAN, D.A. (August 1989). Statistical models and shoe leather.
- 218. BICKEL, P.J., NAIR, V.N. and WANG, P.C.C. (August 1989). Nonparametric inference under biased sampling from a finite population.
- 219. DALANG, R.C. and WALSH, J.B. (September 1989). The sharp Markov property of the brownian sheet and related processes.
- 220. BURMAN, P. and NOLAN, D. (Oct. 1989). Location-adaptive density estimation and nearest-neighbor distance.

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.