

T H E **B** F I L E S

Case studies of bias in real life epidemiologic studies

Bias File 2. Should we stop drinking coffee? The story of coffee and pancreatic cancer

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Department of Epidemiology, Biostatistics & Occupational Health

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca & jay.kaufman@mcgill.ca



THIS CASE STUDY CAN BE FREELY USED FOR EDUCATIONAL PURPOSES WITH DUE CREDIT

Bias File 2. Should we stop drinking coffee? The story of coffee and pancreatic cancer

The story



Brian MacMahon (1923 - 2007) was a British-American epidemiologist who chaired the Department of Epidemiology at Harvard from 1958 until 1988. In 1981, he published a paper in the *New England Journal of Medicine*, a case-control study on coffee drinking and pancreatic cancer [MacMahon B, *et al.* 1981]. The study concluded that "coffee use might account for a substantial proportion of the cases of this disease in the United States." According to some reports, after this study came out, MacMahon stopped drinking coffee and replaced coffee with tea in his office. This publication provoked a storm of protest from coffee drinkers and industry groups, with coverage in the *New York Times*, *Time* magazine and *Newsweek*. Subsequent studies, including one by MacMahon's group, failed to confirm the association. So, what went wrong and why?

The study

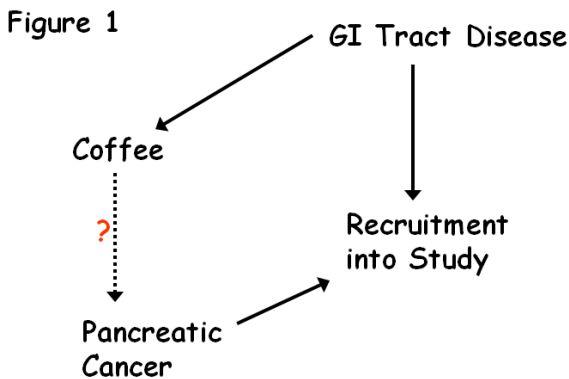
From the original abstract:

We questioned 369 patients with histologically proved cancer of the pancreas and 644 control patients about their use of tobacco, alcohol, tea, and coffee. There was a weak positive association between pancreatic cancer and cigarette smoking, but we found no association with use of cigars, pipe tobacco, alcoholic beverages, or tea. A strong association between coffee consumption and pancreatic cancer was evident in both sexes. The association was not affected by controlling for cigarette use. For the sexes combined, there was a significant dose-response relation (P approximately 0.001); after adjustment for cigarette smoking, the relative risk associated with drinking up to two cups of coffee per day was 1.8 (95% confidence limits, 1.0 to 3.0), and that with three or more cups per day was 2.7 (1.6 to 4.7). This association should be evaluated with other data; if it reflects a causal relation between coffee drinking and pancreatic cancer, coffee use might account for a substantial proportion of the cases of this disease in the United States.

The bias

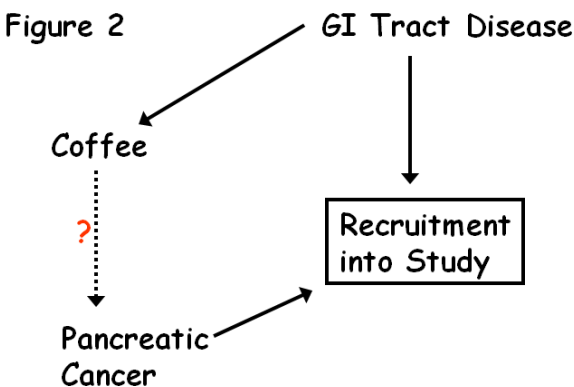
The MacMahon study had several problems and several experts have debated these in various journals, but a widely recognized bias was related to control selection. A nice, easy to read explanation can be found in the Gordis text [Gordis L, 2009], but a 1981 paper by Feinstein drew attention to this problem). Controls in the MacMahon study were selected from a group of patients hospitalized by the same physicians who had diagnosed and hospitalized the cases' disease. The idea was to make the selection process of cases and controls similar. It was also logistically easier to get controls using this method. However, as the exposure factor was coffee drinking, it turned out that patients seen by the physicians who diagnosed pancreatic cancer often had gastrointestinal disorders and were thus advised not to drink coffee (or had chosen to reduce coffee drinking by themselves). So, this led to the selection of controls with higher prevalence of gastrointestinal disorders, and these controls had an unusually low odds of exposure (coffee intake). These in turn may have led to a spurious positive association between coffee intake and pancreatic cancer that could not be subsequently confirmed.

This problem can be explored further using causal diagrams. Since the study used a case-control design, cases were sampled from the source population with higher frequency than the controls, which is



represented by the directed arrow between "pancreatic cancer" and "recruitment into study" in Figure 1. However, controls were selected by being hospitalized by the same doctors who treated the cases. If they were not hospitalized for pancreatic cancer, they must have been hospitalized for some other disease, which gave them a higher representation of GI tract disease than observed in the source population. Patients with GI tract disease may have been discouraged from drinking coffee, which gave controls a lower prevalence of exposure than seen in the source population. This is shown in

Figure 1 as a directed arc from "GI tract disease" to coffee and to "recruitment into study"



Collider stratification bias occurs when one conditions (in the design or the analysis) on a common child of two parents. In this case, restricting the observations to people recruited into the study (Figure 2) changes the correlation structure so that it is no longer the same as in the source population. Specifically, pancreatic cancer and GI tract diseases may be uncorrelated in the general population. However, among patients hospitalized by the doctors who had admitted patients with pancreatic cancer, the ones who didn't have pancreatic disease were more likely to have

something else: a GI tract disease. Therefore, restriction to the population of the doctors who hospitalized the cases induces a negative correlation between these two diseases in the data set.

Figure 3 (those included in the case-control study)

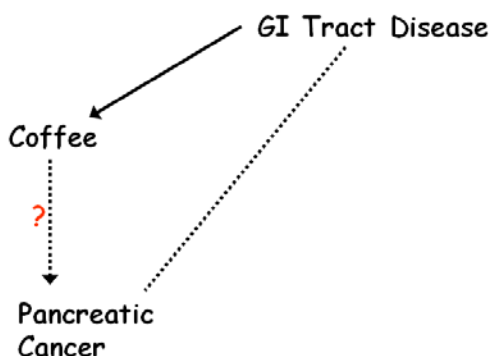


Figure 3 shows a graph of the data for the study population, as opposed to the source population. Restriction to the subjects recruited from the hospital has created a correlation between GI tract disease and pancreatic cancer. Since GI tract disease lowers exposure, an unblocked backdoor path is now opened, which leads to confounding of the estimated exposure effect (shown with a dashed line and a question mark). Specifically, since the induced correlation is negative, and the effect of GI tract disease on coffee is negative, the exposure estimate

for coffee on pancreatic cancer will be biased upward (Vander Stoep et al 1999).

The lesson

Control selection is a critical element of case-control studies, and even the best among us can make erroneous choices. Considerable thought needs to go into this critical step in study design. As Rothman et al. emphasize in their textbook (*Modern Epidemiology*, 2008), the two important rules for control selection are:

1. Controls should be selected from the same population - the source population (i.e. study base) - that gives rise to the study cases. If this rule cannot be followed, there needs to be solid evidence that the population supplying controls has an exposure distribution identical to that of the population that is the source of cases, which is a very stringent demand that is rarely demonstrable.
2. Within strata of factors that will be used for stratification in the analysis, controls should be selected independently of their exposure status, in that the sampling rate for controls should not vary with exposure.

A more general concern than the issue of control selection in case-control studies is the problem of selection bias (Hernán et al 2004). Whenever the epidemiologist conditions statistically (e.g. by stratification, exclusion or adjustment) on a factor affected by exposure and affected by outcome, a spurious correlation will occur in the study data-set that does not reflect an association in the real world from which the data were drawn. If there is already a non-null association between exposure and outcome, it can be shifted upwards or downwards by this form of bias.

Sources and suggested readings*

1. MacMahon B, Yen S, Trichopoulos D *et al.* Coffee and cancer of the pancreas. *N Engl J Med* 1981;304: 630–633.
2. Schmeck HM. Critics say coffee study was flawed. *New York Times*, June 30, 1981.
3. Gordis L. *Epidemiology*. Saunders, 2008.
4. Feinstein A *et al.* Coffee and Pancreatic Cancer. The Problems of Etiologic Science and Epidemiologic Case-Control Research. *JAMA* 1981;246:957-961.
5. Rothman K, Greenland S, Lash T. *Modern epidemiology*. Lippincott Williams & Wilkins, 3rd edition, 2008.
6. Coffee and Pancreatic Cancer. An Interview With Brian MacMahon. *EpiMonitor*, April/May, 1981.
7. Vander Stoep A, Beresford SA, Weiss NS. A didactic device for teaching epidemiology students how to anticipate the effect of a third factor on an exposure-outcome relation. *Am J Epidemiol.* 1999 Jul 15;150(2):221.
8. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004 Sep;15(5):615-25.

Image credit: *Epidemiology*: July 2004 - Volume 15 - Issue 4 - pp 504-508

*From this readings list, the most relevant papers are enclosed.

COFFEE AND CANCER OF THE PANCREAS

BRIAN MACMAHON, M.D., STELLA YEN, M.D., DIMITRIOS TRICHOPOULOS, M.D., KENNETH WARREN, M.D.,
AND GEORGE NARDI, M.D.

Abstract We questioned 369 patients with histologically proved cancer of the pancreas and 644 control patients about their use of tobacco, alcohol, tea, and coffee. There was a weak positive association between pancreatic cancer and cigarette smoking, but we found no association with use of cigars, pipe tobacco, alcoholic beverages, or tea. A strong association between coffee consumption and pancreatic cancer was evident in both sexes. The association was not affected by controlling for cigarette use. For the sexes combined, there was a significant dose-re-

sponse relation ($P \sim 0.001$); after adjustment for cigarette smoking, the relative risk associated with drinking up to two cups of coffee per day was 1.8 (95 per cent confidence limits, 1.0 to 3.0), and that with three or more cups per day was 2.7 (1.6 to 4.7). This association should be evaluated with other data; if it reflects a causal relation between coffee drinking and pancreatic cancer, coffee use might account for a substantial proportion of the cases of this disease in the United States. (N Engl J Med. 1981; 304:630-3.)

OVER the past few decades, cancer of the pancreas has emerged as one of the most important neoplasias in human beings. It now accounts for approximately 20,000 deaths annually in the United States. Causative factors have been sought in several previous studies, but only cigarette smoking has emerged as a consistent, though relatively weak, exogenous risk factor. We report the results of a study that was planned to reevaluate the relation of this disease to smoking and to examine the role of alcohol consumption as a possible confounding variable. Data were also obtained on intake of tea and coffee — factors that have not been adequately investigated in this disease.

METHODS

We conducted a case-control interview study. The cases were patients with histologic diagnoses of cancer of the exocrine pancreas who were in any of 11 large hospitals in the Boston metropolitan area and Rhode Island between October 1974 and August 1979. Patients with tumors of the islet cells, periampullary duodenal mucosa, or ampulla of Vater were not included. We identified 578 patients and interviewed 405 of them. Twenty patients died and 35 were discharged before an interview could be arranged; 78 were too sick to be interviewed, 14 had language difficulties, and 26 refused the interview. Also excluded from the analysis were eight nonwhite patients, four residents of countries other than the United States, eight patients older than 79 years, and 16 patients whose interview information was judged by the interviewer to be of questionable reliability. The analysis is based on data from the remaining 369 patients.

To assemble a control series, the interviewers also attempted to question all other patients who were under the care of the same physician in the same hospital at the time of an interview with a patient with pancreatic cancer. Either before the interview (if the information was known) or afterward, patients with diseases of the pancreas or hepatobiliary tract or diseases known to be associated with smoking or alcohol consumption were excluded. The principal diagnostic categories excluded (in addition to diseases of the biliary tract or pancreas) were cardiovascular disease, diabetes mellitus, respiratory or bladder cancer, and peptic ulcer. From a total of 1118 eligible patients, we interviewed 700; nine died and 131 were discharged before the interview, 179 were too ill, 26 had language problems, and 73 refused. After exclusion of 17 nonwhites, five foreign residents, four persons older than 79 years, and 30 persons

whose interviews were judged to be unreliable, the control series used for the analysis consisted of 644 patients. Minor differences between tables in the stated numbers of cases and controls resulted from absence of specific items being analyzed in a few questionnaires.

The control series was composed of two principal diagnostic groups: 273 patients with cancer other than cancers of the pancreas and biliary tract, respiratory tract, or bladder and 371 patients with other disorders. Of the control patients with cancer, the tumor was in the breast in 65 patients, colon in 60, rectum in 25, stomach in 24, small intestine in nine, ovary in eight, prostate in eight, and cervix in seven; there were also 16 with melanoma and 15 with lymphoma. No other cancer was found in more than four subjects. Diagnoses in the controls without cancer were of a wide variety, although because of the nature of the practices of many of the physicians who were responsible for patients with cancer of the pancreas, patients with gastroenterologic conditions were probably overrepresented in relation to a general hospital population. The principal diagnoses were hernia in 70 patients; colitis, enteritis, or diverticulitis in 41; bowel obstruction, adhesions, or fistula in 26; gastritis in 17; other gastroenterologic conditions in 47; benign tumors in 29; varicose veins or phlebitis in 21; genitourinary disorders in 20; neurologic disorders in 20; gynecologic disorders in 16; and other conditions in 64.

In the analyses, the patients with pancreatic cancer were compared with the control patients with cancer and independently with the control group without cancer. The findings were quite similar, and only the results with the combined control group are presented here.

Several questions in the interview probed the duration and intensity of smoking of cigarettes, cigars, and pipes. Questions on alcoholic beverages asked about the frequency of use before the onset of illness, the age span over which such use occurred, and the type of beverage used most frequently. The questions on tea and coffee were limited to the number of cups consumed in a typical day before the current illness was evident.

Tests of significance and estimates of adjusted relative risks and their confidence limits were derived with the method of Mantel and Haenszel¹ and its extension.² The data were stratified by age in 10-year groups and by sex where appropriate. All confidence limits are 95 per cent intervals. Most analyses were performed with the calculator programs developed by Rothman and Boice.³

RESULTS

Tobacco

There was no difference between cases and controls in the use of cigars or pipe tobacco. Among men, the relative risk associated with use of cigars (with non-smokers as the referent group) was 1.0 (confidence interval, 0.7 to 1.4), and that with use of a pipe was 1.0 (confidence interval, 0.7 to 1.4).

From the Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave., Boston, MA 02115, where reprint requests should be addressed to Dr. MacMahon.

Supported by a grant (5 P01 CA 06373) from the National Cancer Institute.

Table 1. Distribution of Cases and Controls According to Cigarette-Smoking Habits and Estimates of Risk Ratios.

SEX	CATEGORY	NEVER SMOKED	EX-SMOKERS	CURRENT SMOKERS		TOTAL*
				<1 PACK/DAY	>1 PACK/DAY	
Men	Cases (no.)	40	99	22	57	218
	Controls (no.)	74	122	35	75	306
	Adjusted relative risk †	1.0	1.4	1.1	1.4	1.4
	95% confidence interval	—	0.9-2.3	0.5-2.2	0.9-2.4	0.9-2.2
Women	Cases (no.)	62	41	20	26	149
	Controls (no.)	160	86	36	55	337
	Adjusted relative risk †	1.0	1.3	1.5	1.6	1.5
	95% confidence interval	—	0.8-2.2	0.8-2.8	0.9-2.9	1.0-2.2

*Adjusted relative risks and 95 per cent confidence intervals in this column are for consumers of any amount (including ex-consumers) as compared with nonconsumers.
 †Mantel-Haenszel estimates of risk ratios, adjusted over categories of age in decades. In all comparisons, the referent category was subjects who had never smoked. Chi-square (Mantel extension) with equally spaced scores, adjusted over age in decades: 1.2 for men, 4.1 for women.

The data on use of cigarettes are shown in Table 1. There was a weak positive association. Although only the data for women showed a significant dose-response relation, the estimate of the relative risk associated with smoking at any time for both sexes combined was 1.4; the difference from the referent risk was significant (confidence interval, 1.1 to 1.9).

Alcohol

Table 2 shows a comparison of use of alcoholic beverages by cases and by controls. No notable or significant association appeared. The combined estimate of relative risk associated with drinking at any time was 0.9, with a confidence interval of 0.6 to 1.3, and that associated with regular drinking was 0.8 (confidence interval, 0.5 to 1.3).

No difference between cases and controls was found in the statements about the type of alcoholic beverage used most frequently (data not shown).

Table 2. Distribution of Cases and Controls According to Alcohol-Drinking Habits and Estimates of Risk Ratios.

SEX	CATEGORY	ALCOHOL DRINKING			TOTAL
		NONE	OCCASIONAL	REGULAR	
Men	Cases (no.)	16	113	89	218
	Controls (no.)	27	157	123	307
	Adjusted relative risk *	1.0	1.3	1.3	1.3
	95% confidence interval	—	0.7-2.6	0.6-2.6	0.7-2.5
Women	Cases (no.)	33	99	17	149
	Controls (no.)	59	221	57	337
	Adjusted relative risk *	1.0	0.8	0.5	0.8
	95% confidence interval	—	0.5-1.3	0.3-1.1	0.5-1.3

*Chi-square (Mantel extension) with equally spaced scores, adjusted over age in decades: 0.2 for men, 2.7 for women. All data are analyzed as in Table 1.

Tea

The tea consumption of cases and controls is shown in Table 3. A slight inverse association appeared in both sexes, but it was not significant in either.

Coffee

An unexpected association of pancreatic cancer with coffee consumption was evident (Table 4). Among men, each category of coffee consumption had a statistically significant excess risk as compared with that of nondrinkers of coffee, but the dose-response relation was flat. Among women, both categories of consumers of three or more cups per day had significantly elevated risks, and the dose-response relation (as measured by the Mantel test) was highly significant (P<0.001). For the sexes combined, with a simultaneous adjustment for sex and age, the trend was also highly significant (chi-square, 11.0), and the adjusted relative risks for consumers of no cups per day, one to two, three to four, and at least five were 1.0, 2.1, 2.8, and 3.2, respectively.

Table 3. Distribution of Cases and Controls According to Tea-Drinking Habits and Estimates of Risk Ratios.

SEX	CATEGORY	TEA DRINKING (CUPS PER DAY)			TOTAL
		0	1-2	>3	
Men	Cases (no.)	61	134	21	216
	Controls (no.)	72	205	29	306
	Adjusted relative risk *	1.0	0.7	0.8	0.7
	95% confidence interval	—	0.5-1.1	0.4-1.5	0.5-1.1
Women	Cases (no.)	40	85	25	150
	Controls (no.)	75	191	70	336
	Adjusted relative risk *	1.0	0.7	0.6	0.7
	95% confidence interval	—	0.5-1.2	0.3-1.2	0.5-1.2

*Chi-square (Mantel extension) with equally spaced scores, adjusted over age in decades: 1.4 for men, 1.9 for women. All data are analyzed as in Table 1.

Interaction

Since no association was observed with use of alcoholic drinks, tea, pipe tobacco, or cigars, the principal interaction of interest was that between cigarette use and coffee use. This relation was explored in the analysis presented in Table 5. The data showed a consistent association of pancreatic cancer with coffee drinking within each category of smoking, and the data for all smokers and nonsmokers showed a consistent trend with coffee drinking after adjustment for smoking. With the Mantel extension, the chi-square value for the trend with coffee consumption (after adjustment for smoking as well as age and sex) was 10.6 (P ~ 0.001). The association with smoking within categories of coffee consumption was less clear, and the relative risks for ex-smokers and current smokers, adjusted for coffee consumption, did not differ significantly from unity.

DISCUSSION

Our findings with regard to association of cancer of the pancreas with cigarette use and alcohol consumption are consistent with those of previous investigators. The association with cigarette use has been most extensively explored. Weakly positive associations were found in two other case-control studies^{4,5} and in the large cohort studies in British physicians,⁶ American veterans,⁷ and the American Cancer Society population.⁸ The relative risks for cigarette smokers as compared with nonsmokers were 2.3 in the larger case-control study and 1.6, 1.8, and an average of 2.2 in the three cohort studies. These values are comparable to the figure of 1.4 in our study. In one small case-control study, a weak and nonsignificant association was found only in women; among men, there was no difference in cigarette-smoking habits between cases and controls.⁵ However, the inclusion of patients with smoking-related diseases among the hospitalized controls in that study would have served to conceal a weak relation. Adjustment for coffee consumption did not entirely remove the association with cigarette smoking in our own data, although the association was not significant after such adjustment. The possible confounding influence of coffee consumption was not evaluated in the other studies.

The relation between alcohol use and pancreatic cancer has been less extensively studied, but a lack of association has been found in one case-control study⁴ and in a proportional mortality analysis of a large series of deaths of alcoholics.⁹ An association with wine drinking was reported in one study, but the numbers were relatively small, the difference was not conventionally significant, and potential confounding factors were not evaluated.⁵ Overall, it seems unlikely that alcohol consumption has any role in the origin of cancer of the pancreas — an observation that is of some interest in the light of the obvious role of this substance in chronic pancreatitis.

In a recently reported case-control study involving

Table 4. Distribution of Cases and Controls by Coffee-Drinking Habits and Estimates of Risk Ratios.

SEX	CATEGORY	COFFEE DRINKING (CUPS PER DAY)				TOTAL
		0	1-2	3-4	>5	
Men	Cases (no.)	9	94	53	60	216
	Controls (no.)	32	119	74	82	307
	Adjusted relative risk *	1.0	2.6	2.3	2.6	2.6
	95% confidence interval	—	1.2-5.5	1.0-5.3	1.2-5.8	1.2-5.4
Women	Cases (no.)	11	59	53	28	151
	Controls (no.)	56	152	80	48	336
	Adjusted relative risk *	1.0	1.6	3.3	3.1	2.3
	95% confidence interval	—	0.8-3.4	1.6-7.0	1.4-7.0	1.2-4.6

*Chi-square (Mantel extension) with equally spaced scores, adjusted over age in decades: 1.5 for men, 13.7 for women. All data are analyzed as in Table 1.

Table 5. Estimates of Relative Risk of Cancer of the Pancreas Associated with Use of Coffee and Cigarettes.*

CIGARETTE SMOKING	COFFEE DRINKING (CUPS PER DAY)			TOTAL †
	0	1-2	>3	
Never	1.0	2.1	3.1	1.0
Ex-smokers	1.3	4.0	3.0	1.3 (0.9-1.8)
Current smokers	1.2	2.2	4.6	1.2 (0.9-1.8)
Total †	1.0	1.8 (1.0-3.0)	2.7 (1.6-4.7)	

*The referent category is the group that uses neither cigarettes nor coffee. Estimates are adjusted for sex and for age in decades.

†Values are adjusted for the other variable, in addition to age and sex, and are expressed in relation to the lowest category of each variable. Values in parentheses are 95 per cent confidence intervals of the adjusted estimates.

94 patients with pancreatic adenocarcinoma and a similar number of hospital controls, Lin and Kessler noted that the cases tended to drink more decaffeinated coffee than did the controls.⁵ In view of the relatively recent use of decaffeinated coffee on a large scale, it seems unlikely that this particular type of beverage has a causal relation to cases of pancreatic cancer appearing at present. It seems more likely that the high consumption of decaffeinated coffee noted by Lin and Kessler is a reflection of generally high coffee consumption by these patients in the past. These authors gave no data on the use of regular coffee by their subjects.

Although the positive association with coffee consumption that we observed must be evaluated with other data before serious consideration is given to the possibility of a causal relation, it is worth noting that some of the descriptive features of the epidemiology of cancer of the pancreas seem to be consistent with such a relation. The apparent increase in frequency of cancer of the pancreas in recent decades¹⁰ and the low rates observed in Mormons^{11,12} and Seventh-Day Adventists¹³ would be compatible with a causative role for either coffee consumption or cigarette smoking. However, the relatively small excess of men with the disease in proportion to women would seem to be more suggestive of a role for coffee rather than for cigarettes. Some 10 years ago, correlating trade statistics in 20 countries with rates of death from cancer, Stocks reported a positive correlation between coffee consumption and rates of pancreatic cancer; the association was present in both sexes, although it was significant only in men.¹⁴ We note also the recent report of the simultaneous occurrence of cancer of the pancreas in a husband and wife who both added "coffee syrup" to ground coffee before percolating it.¹⁵

Our use of a control group composed of hospitalized patients must be discussed. It is possible that these patients reduced their coffee consumption because of illness and that their replies were affected,

Pancreas
les.*
TOTAL †
1.0
1.3
(0.9-1.8)
1.2
(0.9-1.8)

Estimates are
and sex, and are
parentheses are 95

ma and a
and Kessler
decaffein-
ew of the
on a large
r type of
pancreatic
likely that
noted by
igh coffee
These au-
ce by their

offee con-
ated with
ven to the
oting that
niology of
with such
cy of can-
d the low
-Day Ad-
ative role
smoking.
a with the
em to be
an for cig-
ide statis-
m cancer,
een coffee
nner; the
although
so the re-
ce of can-
who both
ore perco-

hospital-
sible that
ption be-
e affected,

even though the question was related to the time before the onset of their illness. Indeed, Rosenberg et al. reported a lower proportion of coffee consumers among hospitalized women with chronic disease than among women admitted for emergencies.¹⁶ However, the differences noted by Rosenberg et al. between patients with acute and chronic illness were much smaller than those between the cases and controls in our study. Although the majority of control patients in our series had chronic disease, pancreatic cancer itself is a chronic disease, and in theory it would seem as likely as any other disorder to induce a change in coffee consumption. It is a matter for speculation whether such a bias is likely to be greater in our case series or in patients with the diagnoses represented in our control series. It is inconceivable that this bias would account for the total difference between cases and controls, but it is possible that risk may be either overestimated or underestimated on this account. We note, however, that the relative risks shown in Table 4 were similar whether the patients with other cancers or the patients with nonmalignant disorders were used as the control group.

If the association between coffee consumption and pancreatic cancer is confirmed and found to be causal, the relation will have some importance in quantitative terms. Cancer of the pancreas is now the fourth most common fatal malignant disease in the United States. If the distribution of coffee consumption in our control group reflects that in the general population, with relative risks of 1.8 associated with the use of one to two cups daily and 2.7 associated with three or more cups daily, we estimate the proportion of pancreatic cancer that is potentially attributable to coffee consumption to be slightly more than 50 per cent.¹⁷ This estimate emphasizes the need to determine whether the association exists in other data and to evaluate its causal or noncausal nature.

We are indebted to the administrative, nursing, and record-room staffs of Beth Israel Hospital, Carney Hospital, Lahey Clinic, Massachusetts General Hospital, New England Baptist Hospital, New England Deaconess Hospital, Peter Bent Brigham Hospital, Rhode Island Hospital, Tufts-New England Medical Center, University Hospital, and the Veterans Administration Hospital of Jamaica Plain, Mass.; to the physicians on their staffs who gave us permission to interview patients; and to Mrs. Kim Neave and Miss Mary Curran for conducting the interviews.

REFERENCES

1. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959; 22:719-48.
2. Mantel N. Chi-square tests with one degree of freedom: extensions of the Mantel-Haenszel procedure. *J Am Stat Assoc.* 1963; 58:690-700.
3. Roitman KJ, Boice JD. Epidemiologic analysis with a programmable calculator. Washington, D.C.: Government Printing Office, 1979. (DHEW publication no. (NIH)79-1649).
4. Wynder EL, Mabuchi K, Maruchi N, Fortner JG. Epidemiology of cancer of the pancreas. *J Natl Cancer Inst.* 1973; 50:645-67.
5. Lin RS, Kessler II. A multifactorial model for pancreatic cancer in man: epidemiologic evidence. *JAMA.* 1981; 245:147-52.
6. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations of male British doctors. *Br Med J.* 1976; 2:1525-36.
7. Kahn HA. The Dorn study of smoking and mortality among U.S. veterans: report on eight and one-half years of observation. *Natl Cancer Inst Monogr.* 1966; 19:1-125.
8. Hammond EC. Smoking in relation to the death rates of one million men and women. *Natl Cancer Inst Monogr.* 1966; 19:127-204.
9. Monson RR, Lyon JL. Proportional mortality among alcoholics. *Cancer.* 1975; 36:1077-9.
10. Devesa SS, Silverman DT. Cancer incidence and mortality trends in the United States: 1935-74. *J Natl Cancer Inst.* 1978; 60:545-71.
11. Lyon JL, Gardner JW, West DW. Cancer incidence in Mormons and non-Mormons in Utah during 1967-75. *J Natl Cancer Inst.* 1980; 65:1055-62.
12. Enstrom JE. Cancer mortality among Mormons in California during 1968-75. *J Natl Cancer Inst.* 1980; 65:1073-82.
13. Phillips RL, Garfinkel L, Kuzma JW, Beeson WL, Lotz T, Brin B. Mortality among California Seventh-Day Adventists for selected cancer sites. *J Natl Cancer Inst.* 1980; 65:1097-108.
14. Stocks P. Cancer mortality in relation to national consumption of cigarettes, solid fuel, tea and coffee. *Br J Cancer.* 1970; 24:215-25.
15. Ferguson LJ, Watts JM. Simultaneous cancer of the pancreas occurring in husband and wife. *Gut.* 1980; 21:537-40.
16. Rosenberg L, Slone D, Shapiro S, Kaufman DW, Stolley PD, Miettinen OS. Coffee drinking and myocardial infarction in young women. *Am J Epidemiol.* 1980; 111:675-81.
17. Cole P, MacMahon B. Attributable risk percent in case-control studies. *Br J Prev Soc Med.* 1971; 25:242-4.

Welcome to TimesPeople
Get Started

 TimesPeople Lets You Share and Discover the Best of NYTimes.com

0:15 AM [Recommend](#)

[HOME PAGE](#) [TODAY'S PAPER](#) [VIDEO](#) [MOST POPULAR](#) [TIMES TOPICS](#) [Log In](#) | [Register Now](#)

The New York Times
Sunday, August 16, 2009

Science

Search All NYTimes.com

[WORLD](#) [U.S.](#) [N.Y. / REGION](#) [BUSINESS](#) [TECHNOLOGY](#) [SCIENCE](#) [HEALTH](#) [SPORTS](#) [OPINION](#) [ARTS](#) [STYLE](#) [TRAVEL](#) [JOBS](#) [REAL ESTATE](#) [AUTOS](#)

 With the TELUS Mobile Internet Key. [Learn more](#)

CRITICS SAY COFFEE STUDY WAS FLAWED

By HAROLD M. SCHMECK JR.
Published: June 30, 1981

THERE were flaws in a study showing links between coffee drinking and a common form of cancer, several medical scientists and physicians said in letters published in the latest issue of The New England Journal of Medicine.

In March, the journal carried a report showing statistical links between coffee drinking and cancer of the pancreas, the fourth most common cause of cancer deaths among Americans.

"This otherwise excellent paper may be flawed in one critical way," said a letter from Dr. Steven Shedlofsky of the Veterans Administration Hospital in White River Junction, Vt. He questioned the comparison of pancreatic cancer patients with persons hospitalized for noncancerous diseases of the digestive system.

Such patients, he noted, might be expected to give up coffee drinking because of their illness. This, he argued, would tilt the proportion of coffee drinkers away from the "control" group who were being compared with the cancer patients. Amplifying the letter in an interview, Dr. Shedlofsky said many patients with digestive diseases give up coffee because they believe it aggravates their discomfort, and others do so because their doctors have advised them to.

Dr. Thomas C. Chalmers, president of the Mount Sinai Medical Center and dean of its medical school, commented that the investigators who questioned patients on their prehospitalization coffee habits knew in advance which ones had cancer. This could have introduced unintentional bias in the results, Dr. Chalmers asserted.

Among the comments from other physicians were these: the question of whether noncancerous illness might have kept the control patients from drinking coffee was raised; a correspondent pointed out the problem inherent in trying to judge coffee consumption simply by asking about typical daily consumption before hospitalization; and another noted the possible role of other health habits that are closely related to coffee drinking. These habits included cigarette smoking and the use of sugar, milk, cream or nondairy "creamers" with the coffee.

The authors of the original report, led by Dr. Brian MacMahon of the Harvard School of Public Health, defended their study against all of the comments. They agreed that concern was "reasonable" over the large number of patients in their control group who had gastrointestinal disorders. But they said the association between coffee drinking and cancer of the pancreas was present in all the control groups.

The introduction of unintentional bias was unlikely, they said, because the study team had no hypotheses about coffee when it began the study. Coffee drinking only emerged as statistically important when most of the data had already been gathered, they said.

Differences Between Sexes

SIGN IN TO
RECOMMEND

E-MAIL

SEND TO PHONE

PRINT

REPRINTS

SHARE

BEST BUY

WE'LL
HOOK YOU UP
for school

HP DAYS

[CLICK FOR DETAILS](#)

MOST POPULAR

[E-MAILED](#) [BLOGGED](#)

1. [Patient Money: The Expense of Eating With Celiac Disease](#)
2. [Noticed: It's Hip to Be Round](#)
3. [Bob Herbert: Hard to Believe!](#)
4. [Gail Collins: To Be Old and in Woodstock](#)
5. [Getting Your Wireless Network Up to Speed](#)
6. [Multicultural Stages in a Small Oregon Town](#)
7. [Well: Fatty Foods Affect Memory and Exercise](#)
8. [Believers Invest in the Gospel of Getting Rich](#)
9. [Shortcuts: New Worries About Children With Cellphones](#)
10. [Paul Krugman: Republican Death Trip](#)

[Go to Complete List](#) »



The authority of informal power

ALSO IN JOBS »

Looking for a new job?
Are three martinis three too many?

nytimes.com

JOBS

RELATED ADS

[what are related ads?](#)

- » [Coffee Car](#)
- » [MR Coffee Pots](#)
- » [Study Skills](#)

The study showed no difference in risk between men who said they drank only about two cups of coffee a day and those who drank much more. Among women, however, the risk seemed to be related to the amount consumed. Some of the physicians who commented on the study considered the lack of a dose effect in men puzzling and a cause of doubt concerning the overall implications of the study.

In their original report, Dr. MacMahon and his colleagues treated their evidence cautiously, asserting that further studies were needed to determine whether coffee drinking was actually a factor in causing the cancers. If it is a matter of cause and effect, they said, and if the findings apply to the nation as a whole, coffee drinking might be a factor in slightly more than half of the roughly 20,000 cases a year of that form of cancer in the United States.

Coffee industry spokesmen, who were critical of the report when it was published in March, estimate that more than half of Americans over the age of 10 drink coffee.

- » [Sleep Study](#)
- » [Coffee Allergy](#)



Ads by Google what's this?

[Bell Traveller Q3](#)

Bell Traveller Q3 Bell Traveller Q3
affairesmobiles.bell.ca/fr/C

INSIDE NYTIMES.COM

[Sign in to Recommend](#)

[More Articles in Science >](#)

HEALTH »



Roving Runner: Baseball Nostalgia in the Bronx

WORLD »



Kiev Residents Wonder if Mayor Is Fit for Office

OPINION »

But They Were Next in Line for Takeoff

Airplane passengers should demand approval of the merciful Airline Passengers Bill of Rights.

FASHION & STYLE »



The Spirit of '69, Circa 1972

ARTS »



In Dresden, High Culture and Ugly Reality Clash

OPINION »



Weekend Opinionator: Cheney v. Bush

EPIDEMIOLOGY Fourth Edition

LEON GORDIS, MD, MPH, DrPH

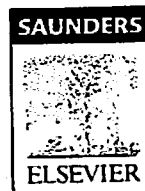
Professor of Epidemiology

Johns Hopkins Bloomberg School of Public Health

Professor of Pediatrics

Johns Hopkins School of Medicine

Baltimore, Maryland



although it is attractive to choose as hospitalized controls a disease group that is obviously unrelated to the putative causative factor under investigation, such controls are unlikely to be representative of the general reference population. As a result, it will not be clear whether it is the cases or the controls who differ from the general population.

The issue of which diagnostic groups would be eligible for use as controls and which would be ineligible (and therefore excluded) is very important. Let us say we are conducting a case-control study of lung cancer and smoking: we select as cases patients who have been hospitalized with lung cancer, and as controls we select patients who have been hospitalized with emphysema. What problem would this present? Because we know that there is a strong relationship between smoking and emphysema, our controls, the emphysema patients, would include a high number of smokers. Consequently, any relationship of smoking to lung cancer would not be detectable in this study, because we would have selected as controls a group of persons in which there is a greater-than-expected prevalence of smoking. We might therefore want to exclude from our control group those persons who have other smoking-related diagnoses, such as coronary heart disease, bladder cancer, pancreatic cancer, and emphysema. Such exclusions might yield a control group with a lower-than-expected prevalence of smoking and the exclusion process becomes complex. One alternative is to not exclude any groups from selection as controls in the design of the study, but to analyze the study data separately for different diagnostic subgroups that constitute the control group.

PROBLEMS IN CONTROL SELECTION

The following example demonstrates the problem of exclusions in the process of control selection:

In 1981, MacMahon and coworkers⁶ reported a case-control study of cancer of the pancreas. The cases were patients with a histologically confirmed diagnosis of pancreatic cancer in 11 Boston and Rhode Island hospitals from 1974 to 1979. Controls were selected from all patients who were hospitalized at the same time as the cases; and they were selected from other inpatients hospitalized by the attending physicians who had hospitalized the cases. One finding in this study was an apparent dose-response relationship between coffee consumption and cancer of the pancreas, particularly in women (Table 10-6).

When such a relationship is observed, it is difficult to know whether the disease is *caused* by the coffee consumption or by some factor closely related to the coffee consumption. Because smoking is a known risk factor for cancer of the pancreas, and because coffee consumption is closely related to cigarette smoking (it is rare to find a smoker who does not drink coffee), did MacMahon and others observe an association of coffee consumption with pancreatic cancer because the coffee caused the pancreatic cancer, or because coffee consumption is related to cigarette smoking, and cigarette smoking is known to be a risk factor for cancer of the pancreas? Recognizing this problem, the authors analyzed the data after stratifying for smoking history. The relationship with coffee consumption held both for current smokers and for those who had never smoked (Table 10-7).

This report aroused great interest in both the scientific and lay communities, particularly among coffee manufacturers. Given the widespread exposure of human beings to coffee, if the reported relationship were true, it would have major public health implications.

Let us examine the design of this study. The cases were white patients with cancer of the pancreas at 11 Boston and Rhode Island hospitals. The controls are of particular interest: They were patients with other diseases who were hospitalized by the same physicians who had hospitalized the cases. That is, when a case had been identified, the attending physician was asked if another of his or her patients who was hospitalized at the same time for another condition could be interviewed as a control. This unusual method of control selection had a practical advantage: One of the major obstacles in obtaining participation of hospital controls in case-control studies is that permission to contact the patient is requested of the attending physician. The physicians are often not motivated to have their patients serve as controls, because the patients do not have the disease that is the focus of the study. By asking physicians who had already given permission for patients with pancreatic cancer to participate, the likelihood was increased that permission would be granted for patients with other diseases to participate as controls.

Did that practical decision introduce any problems? The underlying question that the investigators wanted to answer was whether patients with cancer of the pancreas drank more coffee than did people without cancer of the pancreas in the same population (Fig. 10-3). What MacMahon and coworkers

TABLE 10-6. Distribution of Cases and Controls by Coffee-Drinking Habits and Estimates of Risk Ratios

Sex	Category	Coffee Consumption (Cups/Day)				Total
		0	1-2	3-4	≥5	
M	Number of cases	9	94	53	60	216
	Number of controls	32	119	74	82	307
	Adjusted relative risk*	1.0	2.6	2.3	2.6	2.6
	95% Confidence interval	-	1.2-5.5	1.0-5.3	1.2-5.8	1.2-5.4
F	Number of cases	11	59	53	28	151
	Number of controls	56	152	80	48	336
	Adjusted relative risk*	1.0	1.6	3.3	3.1	2.3
	95% Confidence interval	-	0.8-3.4	1.6-7.0	1.4-7.0	1.2-4.6

*Chi-square (Mantel extension) with equally spaced scores, adjusted over age in decades: 1.5 for men, 13.7 for women. Mantel-Haenszel estimates of risk ratios, adjusted over categories of age in decades. In all comparisons, the referent category was subjects who never drank coffee. From MacMahon B, Yen S, Trichopoulos D, et al: Coffee and cancer of the pancreas. *N Engl J Med* 304(11):630-633, 1981.

found was that the level of coffee consumption in cases was greater than the level of coffee consumption in controls.

The investigators would like to be able to establish that the level of coffee consumption observed in the controls is what would be expected in the general population without pancreatic cancer and that cases therefore demonstrate excessive coffee consumption (Fig. 10-4). But the problem is this: Which physicians are most likely to admit patients with cancer of the pancreas to the hospital? Gastroenterologists are

often the admitting physicians. Many of their other hospitalized patients (who served as controls) also have gastrointestinal problems, such as esophagitis and peptic ulcer. So in this study, the persons who served as controls may very well have reduced their intake of coffee, either because of a physician's instructions or because of their own realization that reducing their coffee intake could relieve their symptoms. We cannot assume that the controls' levels of coffee consumption are representative of the level of coffee consumption expected in the general popula-

TABLE 10-7. Estimates of Relative Risk* of Cancer of the Pancreas Associated with Use of Coffee and Cigarettes

Cigarette Smoking Status	Coffee Drinking (Cups/Day)			Total†
	0	1-2	≥3	
Never smoked	1.0	2.1	3.1	1.0
Ex-smokers	1.3	4.0	3.0	1.3
Current smokers	1.2	2.2	4.6	1.2 (0.9-1.8)
Total*	1.0	1.8 (1.0-3.0)	2.7 (1.6-4.7)	

*The referent category is the group that uses neither cigarettes nor coffee. Estimates are adjusted for sex and age in decades.
†Values are adjusted for the other variables, in addition to age and sex, and are expressed in relation to the lowest category of each variable. Values in parentheses are 95% confidence intervals of the adjusted estimates.
From MacMahon B, Yen S, Trichopoulos D, et al: Coffee and cancer of the pancreas. *N Engl J Med* 304(11):630-633, 1981.

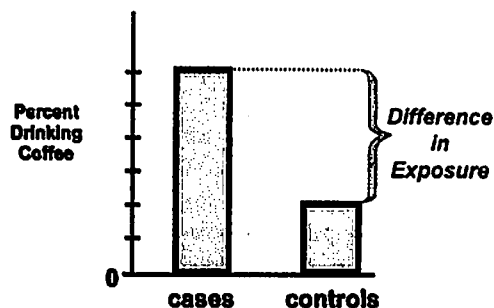


Figure 10-3. Interpreting the results of a case-control study of coffee drinking and pancreatic cancer.

tion; their rate of coffee consumption may be abnormally low. Thus, the observed difference in coffee consumption between pancreatic cancer cases and controls may not necessarily have been the result of cases drinking more coffee than expected, but rather of the controls drinking less coffee than expected (Fig. 10-5).

MacMahon and his colleagues subsequently repeated their analysis but separated controls with gastrointestinal illness from controls with other conditions. They found that the risk associated with coffee consumption was indeed higher when the comparison was with controls with gastrointestinal illness but that the relationship between coffee consumption and pancreatic cancer persisted, albeit at a lower level, even when the comparison was with controls with other illnesses. Several years later, Hsieh and coworkers reported a new study that attempted to replicate these results; it did not support the original findings.⁷

In summary, when a difference in exposure is observed between cases and controls, we must ask

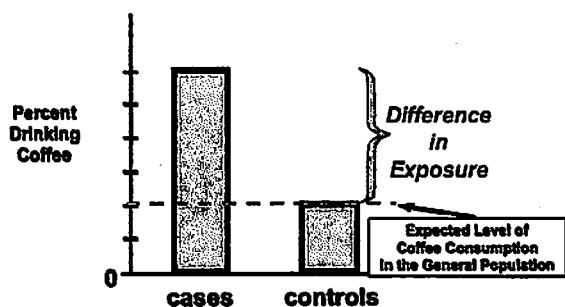


Figure 10-4. Interpreting the results of case-control studies: Is the lower level the expected level of exposure?

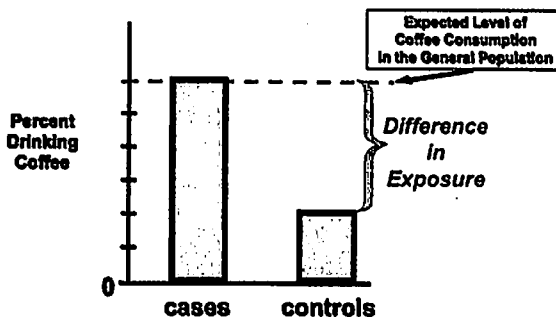


Figure 10-5. Interpreting the results of case-control studies: Is the higher level the expected level of exposure?

whether the level of exposure observed in the controls is really the level expected in the population in which the study was carried out or whether—perhaps given the manner of selection—the controls may have a particularly high or low level of exposure that might not be representative of the level in the population in which the study was carried out.

MATCHING

A major concern in conducting a case-control study is that cases and controls may differ in characteristics or exposures other than the one that has been targeted for study. If more cases than controls are found to have been exposed, we may be left with the question of whether the observed association could be due to differences between the cases and controls in factors other than the exposure being studied. For example, if more cases than controls are found to have been exposed, and if most of the cases are poor and most of the controls are affluent, we would not know whether the factor determining development of disease is exposure to the factor being studied or another characteristic associated with being poor. To avoid such a situation, we would like to ensure that the distribution of the cases and controls by socioeconomic status is similar, so that a difference in exposure will likely constitute the critical difference, and the presence or absence of disease is not likely to be attributable to a difference in socioeconomic status.

One approach to dealing with this problem in the design and conduct of the study is to match the cases and controls for factors about which we may be con-

Coffee and Pancreatic Cancer

The Problems of Etiologic Science and Epidemiologic Case-Control Research

THE RECENT report that coffee may cause pancreatic cancer¹ was presented in a pattern that has become distressingly familiar. The alleged carcinogen is a commonly used product. The report was given widespread publicity before the supporting evidence was available for appraisal by the scientific community, and the public received renewed fear and uncertainty about the cancerous hazards lurking in everyday life.

The research on coffee and pancreatic cancer was done with the case-control technique that has regularly been used in epidemiologic circumstances where the more scientifically desirable forms² of clinical investigation—a randomized controlled trial or a suitably performed observational cohort study—are either impossible or unfeasible. In case-control studies, the investigators begin at the end, rather than at the beginning, of the cause-effect pathway. The cases are selected from persons in whom the target disease has already developed. The controls are selected from persons in whom that disease has not been noted. The cases and controls are then investigated in a backward temporal direction, with inquiries intended to determine antecedent exposure to agents that may have caused the disease. If the ratio of antecedent exposure to a particular agent is higher in the cases than in the controls, and if the associated mathematical calculations are “statistically significant,” the agent is suspected of having caused the disease.

In the recently reported study¹ of coffee and pancreatic cancer, the investigators began by assembling records for 578 cases of patients with “histologic diagnoses of cancer of the exocrine pancreas.” The investigators next created two “control” groups, having other diagnoses. The cases and controls were then interviewed regarding antecedent exposure to tobacco, alcohol, tea, and coffee. When the data were analyzed for groups demarcated according to gender and quantity of coffee consumption, the calculated relative-risk ratios for pancreatic cancer were the values shown in Table 1.

From these and other features of the statistical analysis, the investigators concluded that “a strong association between coffee consumption and pancreatic cancer was evident in both sexes.” The conclusions were presented

with the customary caveats about the need for more research and with the customary restraints shown in such expressions as “coffee use *might* [our italics] account for a substantial proportion” of pancreatic cancers. Nevertheless, the impression was strongly conveyed that coffee had been indicted as a carcinogen.

Although the major public attention has been given to the “Results” and “Discussion” sections of the published report, readers concerned with scientific standards of evidence will want to focus on the “Methods.” The rest of this commentary contains a review of pertinent principles of case-control methodology, together with a critique of the way these principles were applied in the coffee-pancreatic cancer study to formulate a hypothesis, assemble the case and control groups, collect the individual data, and interpret the results.

Scientific Hypotheses and ‘Fishing Expeditions’

Most case-control studies are done to check the hypothesis that the target disease has been caused by a specified suspected agent, but after the cases and controls are assembled the investigators can also collect data about many other possible etiologic agents. The process of getting and analyzing data for these other agents is sometimes called a “fishing expedition,” but the process seems entirely reasonable. If we do not know what causes a disease, we might as well check many different possibilities. On the other hand, when an unsuspected agent yields a positive result, so that the causal hypothesis is generated by the data rather than by the investigator, the results of the fishing expedition require cautious interpretation. Many scientists would not even call the positive association a “hypothesis” until the work has been reproduced in another investigation.

The investigators who found a positive association between coffee consumption and pancreatic cancer have been commendably forthright in acknowledging that they were looking for something else. When the original analyses showed nothing substantial to incriminate the two principal suspects—tobacco and alcohol—the exploration of alternative agents began. The investigators do not state how many additional agents were examined besides

From the Robert Wood Johnson Clinical Scholars Program, Yale University School of Medicine, New Haven, Conn (Drs Feinstein and Horwitz), and the Cooperative Studies Program Support Center, Veterans Administration Hospital, West Haven, Conn (Dr Feinstein), and the McGill Cancer Center, McGill University (Dr Spitzer), and the Kellogg Center for Advanced Studies in Primary Care, Montreal General Hospital (Drs Spitzer and Battista), Montreal.

Reprint requests to Robert Wood Johnson Clinical Scholar Program, Yale University School of Medicine, 333 Cedar St, Box 3333, New Haven, CT 06510 (Dr Feinstein).

Table 1.—Relative-Risk Ratios According to Gender and Quantity of Coffee Consumption

	Coffee Consumption, Cups per Day			
	0	1-2	3-4	≥5
Men	1.0	2.6	2.3	2.6
Women	1.0	1.6	3.3	3.1

tea and coffee, but tea was exonerated in the subsequent analyses, while coffee yielded a positive result.

The investigators suggest that this result is consistent with coffee-as-carcinogen evidence that had appeared in a previous case-control study³ of pancreatic cancer. In fact, however, coffee was not indicted in that previous study. The previous investigators found an elevated risk ratio for only decaffeinated coffee, and they drew no conclusion about it, having found elevated risks for several other phenomena that led to the decision that pancreatic cancer had a nonspecific multifactorial etiology. Thus, the new hypothesis that coffee may cause pancreatic cancer not only arises from a "fishing expedition," but also contradicts the results found in previous research.

Selection and Retention of Cases and Controls

Because the investigators begin at the end of the causal pathway and must explore it with a reversal of customary scientific logic, the selection of cases and controls is a crucial feature of case-control studies. Both groups are chosen according to judgments made by the investigators. The decisions about the cases are relatively easy. They are commonly picked from a registry or some other listing that will provide the names of persons with the target disease. For the controls, who do not have the target disease, no standard method of selection is available, and they have come from an extraordinarily diverse array of sources. The sources include death certificates, tumor registries, hospitalized patients, patients with specific categories of disease, patients hospitalized on specific clinical services, other patients of the same physicians, random samples of geographically defined communities, people living in "retirement" communities, neighbors of the cases, or personal friends of the cases.

One useful way of making these decisions less arbitrary is to choose cases and controls according to the same principles of eligibility and observation that would be used in a randomized controlled trial of the effects of the alleged etiologic agent. In such a trial, a set of admission criteria would be used for demarcating persons to be included (or excluded) in the group who are randomly assigned to be exposed or non-exposed to the agent. Special methods would then be used to follow the members of the exposed and non-exposed groups thereafter, and to examine them for occurrence of the target disease. Those in whom this disease develops would become the cases, and all other people would be the controls.

When cases and controls are chosen for a case-control study, the selection can be made from persons who would have been accepted for admission to such a randomized trial and who have been examined with reasonably similar methods of observation. As a scientific set of guidelines for choosing eligible patients, the randomized-trial principles could also help avoid or reduce many of the different forms of bias that beset case-control studies. Among these difficulties are several biases to be discussed later, as well as other problems such as clinical susceptibility bias, surveillance bias, detection bias, and "early death" bias, which are beyond the scope of this discussion and have been described elsewhere.^{4,9}

The randomized-trial principles can also help illuminate the problems created and encountered by the investigators in the study of coffee and pancreatic cancer. In a randomized trial, people without pancreatic cancer would

be assigned either to drink or not to drink coffee. Anyone with clinical contraindications against coffee drinking or indications for it (whatever they might be) would be regarded as ineligible and not admitted. Everyone who did enter the trial, however, would thereafter be included in the results as the equivalent of either a case, if later found to have pancreatic cancer, or a control. The cases would be "incidence cases," with newly detected pancreatic cancer, whose diagnoses would be verified by a separate panel of histological reviewers. All of the other admitted persons would eventually be classified as unaffected "controls," no matter what ailments they acquired, as long as they did not have pancreatic cancer. If large proportions of the potential cases and controls were lost to follow-up, the investigators would perform detailed analyses to show that the remaining patients resembled those who were lost, thus providing reasonable assurance that the results were free from migration bias.²

In the coffee-pancreatic cancer study, the source of the cases was a list of 578 patients with "histologic diagnoses of cancer of the exocrine pancreas." The histologic material was apparently not obtained and reviewed; and the authors do not indicate whether the patients were newly diagnosed "incidence cases," or "prevalence cases" who had been diagnosed at previous admissions. Regardless of the incidence-prevalence distinction, however, the published data are based on only 369 (64%) of the 578 patients who were identified as potential cases. Most of the "lost" patients were not interviewed, with 98 potential cases being too sick or already dead when the interviewer arrived. The investigators report no data to indicate whether the "lost" cases were otherwise similar to those who were retained.

In choosing the control group, the investigators made several arbitrary decisions about whom to admit or exclude. The source of the controls was "all other patients who were under the care of the same physician in the same hospital at the time of an interview with a patient with pancreatic cancer." From this group, the investigators then excluded anyone with any of the following diagnoses: diseases of the pancreas; diseases of the hepatobiliary tract; cardiovascular disease; diabetes mellitus; respiratory cancer; bladder cancer; or peptic ulcer. Since none of these patients would have been excluded as nonpancreatic-cancer controls if they acquired these diseases after entry into a randomized trial of coffee drinking, their rejection in this case-control study is puzzling. The investigators give no reasons for excluding patients with "diseases of the pancreas or hepatobiliary tract." The reason offered for the other rejections is that the patients had "diseases known to be associated with smoking or alcohol consumption." The pertinence of this stipulation for a study of coffee is not readily apparent.

Since the investigators do not state how many potential controls were eliminated, the proportionate impact of the exclusions cannot be estimated. The remaining list of eligible control patients, however, contained 1,118 people, of whom only a little more than half—644 patients—became the actual control group used for analyses. Most of the "lost" controls were not interviewed because of death, early discharge, severity of illness, refusal to participate, and language problems. Of the 700 interviewed controls, 56 were subsequently excluded because they were non-white, foreign, older than 79 years, or "unreliable." No

data are offered to demonstrate that the 644 actual controls were similar to the 474 "eligible" controls who were not included.

The many missing controls and missing interviews could have led to exclusion biases^{10,11} whose effects cannot be evaluated in this study. The investigators have also given no attention to the impact of selective hospitalization bias, perceived by Berkson⁴ and empirically demonstrated by Roberts et al,⁶ that can sometimes falsely elevate relative-risk ratios in a hospital population to as high as 17 times their true value in the general population. For example, in a hospitalized population, Roberts et al⁶ found a value of 5.0 for the relative-risk ratio of arthritic and rheumatic complaints in relation to laxative usage; but in the general population that contained the hospitalized patients, the true value was 1.5. Whatever may have been the effects of selective hospitalization in the current study (including the possibility of having masked real effects of tobacco and alcohol), the way that the cases and controls were chosen made the study particularly vulnerable to the type of bias described in the next section.

Protopathic Bias in Cases and Controls

"Protopathic" refers to early disease. A protopathic problem occurs if a person's exposure to a suspected etiologic agent is altered because of the early manifestations of a disease, and if the altered (rather than the original) level of exposure is later associated with that disease. By producing changes in a person's life-style or medication, the early manifestations of a disease can create a bias unique to case-control studies.¹² In a randomized trial or observational cohort study, the investigator begins with each person's baseline state and follows it to the subsequent outcome. If exposure to a suspected etiologic agent is started, stopped, or altered during this pathway, the investigator can readily determine whether the change in exposure took place before or after occurrence of the outcome. In a case-control study, however, the investigator beginning with an outcome cannot be sure whether it preceded or followed changes in exposure to the suspected agent. If the exposure was altered because the outcome had already occurred and if the timing of this change is not recognized by the investigator, the later level of exposure (or non-exposure) may be erroneously linked to the outcome event.

For example, in circumstances of ordinary medical care, women found to have benign breast disease might be told by their physicians to avoid or stop any form of estrogen therapy. If such women are later included as cases in a case-control study of etiologic factors in benign breast disease, the antecedent exposure to estrogens will have been artifactually reduced in the case group. Oral contraceptives or other forms of estrogen therapy may then be found to exert a fallacious "protection" against the development of benign breast disease.

The problem of protopathic bias will occur in a case-control study if the amount of previous exposure to the suspected etiologic agent was preferentially altered—either upward or downward—because of clinical manifestations that represented early effects of the same disease that later led to the patient's selection as either a case or control. The bias is particularly likely to arise if the preferential decisions about exposure were made in opposite directions in the cases and controls. The coffee-

pancreatic cancer study was particularly susceptible to this type of bi-directional bias. The customary intake of coffee may have been increased by members of the pancreatic-cancer case group who were anxious about vague abdominal symptoms that had not yet become diagnosed or even regarded as "illness." Conversely, control patients with such gastrointestinal ailments as regional enteritis or dyspepsia may have been medically advised to stop or reduce their drinking of coffee. With a strict set of admission criteria, none of these patients would be chosen as cases or controls, because the use of the alleged etiologic agent would have been previously altered by the same ailment that led to the patient's selection for the study.

This problem of protopathic bias is a compelling concern in the investigation under review here. Because so many potential control patients were excluded, the remaining control group contained many people with gastrointestinal diseases for which coffee drinking may have been previously reduced or eliminated. Of the 644 controls, 249 (39%) had one of the following diagnoses: cancer of the stomach, bowel, or rectum; colitis, enteritis, or diverticulitis; bowel obstruction, adhesions, or fistula; gastritis; or "other gastroenterologic conditions." If coffee drinking is really unrelated to pancreatic cancer, but if many of these 249 patients had premonitory symptoms that led to a cessation or reduction in coffee drinking "before the current illness was evident," the subsequent distortions could easily produce a false-positive association.

The existence of this type of bias could have been revealed or prevented if the investigators had obtained suitable data. All that was needed during the interview with each case or control patient was to ask about duration of coffee drinking, changes in customary pattern of consumption, and reasons for any changes. Unfortunately, since coffee was not a major etiologic suspect in the research, this additional information was not solicited. After the available data were analyzed, when the investigators became aware of a possible problem, they tried to minimize its potential importance by asserting that "although the majority of control patients in our series had chronic disease, pancreatic cancer is itself a chronic disease, and in theory it would seem as likely as any other disorder to induce a change in coffee [consumption]." This assertion does not address the point at issue. The bias under discussion arises from changes in exposure status because of the early clinical manifestations of a disease, not from the chronic (or acute) characteristics of the conditions under comparison.

The investigators also claimed that "it is inconceivable that this bias would account for the total difference between cases and controls." The conception is actually quite easy. To make the demonstration clear, let us eliminate gender distinctions and coffee quantification in the investigators' Table 4, which can then be converted into a simple fourfold table (Table 2). In this table, the odds ratio, which estimates the relative-risk ratio, is $(347/20)/(555/88)=2.75$, which is the same magnitude as the relative risks cited by the investigators.

Let us now assume that 5% of the coffee-drinker cases were formerly non-coffee-drinkers. If so, 17 people in the case group would be transferred downward from the coffee drinkers to the nondrinkers. Although 249 members

Table 2.—Status of Study Subjects According to Coffee Consumption		
	Cases	Controls
Coffee-drinkers	347	555
Non-coffee-drinkers	20	88
Total	367	643

Table 3.—Hypothetical* Status of Study Subjects Shown in Table 2		
	Cases	Controls
Coffee-drinkers	330	573
Non-coffee-drinkers	37	70
Total	367	643

*Based on estimate that 5% of coffee-drinkers in case group were previously non-coffee-drinkers and that 20% of non-coffee-drinkers in control group ceased coffee consumption because of symptoms.

of the control group had gastrointestinal conditions that might have led to a cessation of coffee consumption, let us conservatively estimate that only 20% of the 88 controls listed in the non-coffee-drinkers category were previous coffee-drinkers who had stopped because of symptoms. If so, 18 of the non-coffee-drinking controls should move upward into the coffee-drinking group. With these reclassifications, the adjusted fourfold table would be as presented in Table 3. For this new table, the odds ratio is $(330/37)/(573/70)=1.09$, and the entire positive association vanishes.

Acquisition of Basic Data

All of the difficulties just described arise as consequences of basic decisions made in choosing cases and controls. After these decisions are completed, the case-control investigator acquires information about each person's antecedent exposure. This information becomes the basic research data, analogous to the description of each patient's outcome in a randomized controlled trial. The information about exposure should therefore be collected with thorough scientific care, using impeccable criteria to achieve accuracy, and, when necessary, using objective (or "blinded") methods to prevent biased observations.

These scientific requirements are seldom fulfilled in epidemiologic research. The primary data about exposure are verified so infrequently in case-control studies that prominent epidemiologists¹³ have begun to make public pleas for improved scientific standards and methods. In the few instances where efforts have been made to confirm recorded data,^{14,15} to repeat interviews at a later date¹⁶ or to check the agreement of data obtained from different sources,¹¹ the investigators have encountered discrepancies of major magnitude. In one of these studies,¹⁷ when the agent of exposure (occupation as a fisherman) was confirmed, the original numbers of exposed people were reduced by 17%. Had these numbers not been corrected, the study would have produced misleading conclusions.

Although errors of similar magnitude could easily have occurred in the coffee-pancreatic cancer investigation, the investigators did not publish even a brief text of the actual questions used for the interviews, and no efforts are mentioned to check the quality of the data that were obtained in the single interview with each patient. Family

members or friends were not asked to confirm the patients' answers; the information was not checked against previous records; and no patients were reinterviewed after the original interrogation to see whether subsequent responses agreed with what was said previously. Although a verification of each interview is difficult to achieve in a large study, the scientific quality of the data could have been checked in a selected sample.

Because of the high likelihood of the protopathic bias noted earlier, the quality of the coffee-drinking data is a major problem in the study under review. The investigators state that "the questions on tea and coffee were limited to the number of cups consumed in a typical day before the current illness was evident." This approach would not produce reliable data, since it does not indicate what and when was a "typical day," who decided what was the "time before the current illness was evident," or who determined which of the patient's symptoms were the first manifestation of "illness" either for pancreatic cancer or for the diverse diseases contained in the control group.

Although the investigators acknowledge the possibility that "patients reduced their coffee consumption because of illness," nothing was done to check this possibility or to check the alternative possibility that other patients may have increased their customary amounts of coffee drinking. In addition to no questions about changes in coffee consumption, the patients were also asked nothing about duration. Thus, a patient who had started drinking four cups a day in the past year would have been classified as having exactly the same exposure as a patient who had been drinking four cups a day for 30 years.

The Problem of Multiple Contrasts

When multiple features of two groups are tested for "statistically significant" differences, one or more of those features may seem "significant" purely by chance. This multiple-contrast problem is particularly likely to arise during a "fishing expedition." In the customary test of statistical significance, the investigator contrasts the results for a single feature in two groups. The result of this single-feature two-group contrast is declared significant if the *P* value falls below a selected boundary, which is called the α level. Because α is commonly set at .05, medical literature has become replete with statements that say "the results are statistically significant at $P<.05$." For a single two-group contrast at an α level of .05, the investigator has one chance in 20 (which can also be expressed as contrary odds of 19 to 1) of finding a false-positive result if the contrasted groups are really similar.

For the large series of features that receive two-group contrasts during a "fishing expedition," however, statistical significance cannot be decided according to the same α level used for a single contrast. For example, in the coffee-pancreatic cancer study, the cases and controls were divided for two-group contrasts of such individual exposures (or non-exposures) as cigars, pipes, cigarettes, alcohol, tea, and coffee. (If other agents were also checked, the results are not mentioned.) With at least six such two-group contrasts, the random chance of finding a single false-positive association where none really exists is no longer .05. If the characteristics are mutually independent, the chance is at least $.26[=1-(.95)^6]$. Consequently, when six different agents are checked in the same study,

the odds against finding a spurious positive result are reduced from 19 to 1 and become less than 3 to 1 [= .74/.26].

To guard against such spurious conclusions during multiple contrasts, the customary statistical strategy is to make stringent demands on the size of the P value required for "significance." Instead of being set at the customary value of .05, the α level is substantially lowered. Statisticians do not agree on the most desirable formula for determining this lowered boundary, but a frequent procedure is to divide the customary α level by k , where k is the number of comparisons.¹⁸ Thus, in the current study, containing at least six comparisons, the decisive level of α would be set at no higher than .05/6=.008.

In the published report, the investigators make no comment about this multiple-contrast problem and they do not seem to have considered it in their analyses. In one of the results, a P value is cited as "<.001," but most of the cogent data for relative risks are expressed in "95% confidence intervals," which were calculated with α =.05. Many of those intervals would become expanded to include the value of 1, thereby losing "statistical significance," if α were re-set at the appropriate level of .008 or lower.

Comment

The foregoing discussion has been confined to the main reasons for doubting the reported association between coffee and pancreatic cancer. Readers who are interested in evaluating other features of the study can check its constituent methods by referring to the criteria listed in several published proposals⁸⁻¹⁰ of scientific standards for case-control research.

A separate problem, to be mentioned only in passing, is the appropriateness of forming conclusions and extensively diffusing results from a study in which the hypothesis develops as an analytic surprise in the data. Scientists and practitioners in the field of human health face difficult dilemmas about the risks and benefits of their activities. The old principle of avoiding harm whenever possible holds true whether a person or a population is at risk. Whether to shout "Fire!" in a crowded theater is a difficult decision, even if a fire is clearly evident. The risk of harm seems especially likely if such shouts are raised when the evidence of a blaze is inconclusive or meager. Aside from puzzled medical practitioners and a confused lay public, another possible victim is the developing science of chronic disease epidemiology. Its credibility can withstand only a limited number of false alarms.

Because the epidemiologic case-control study is a necessary, currently irreplaceable research mechanism in etiologic science, its procedures and operating paradigms need major improvements in scientific quality. In the evaluation of cause-effect relationships for therapeutic agents, the experimental scientific principles of a randomized trial have sometimes required huge sample sizes and massive efforts that have made the trials become an "indispensable ordeal."¹⁹ In the evaluation of cause-effect relationships for etiologic agents, the case-control technique has eliminated the "ordeal" of a randomized controlled trial by allowing smaller sample sizes, the analysis of natural events and data, and a reversed observational direction. Since the use of scientific principles remains "indispensable," however, the development and application of suitable scientific standards in case-

control research is a prime challenge in chronic disease epidemiology today.

The current methodologic difficulties arise because case-control investigators, having recognized that etiologic agents cannot be assigned with experimental designs, and having necessarily abandoned the randomization principle in order to work with naturally occurring events and data, have also abandoned many other scientific principles that are part of the experimental method and that could be employed in observational research. The verification and suitably unbiased acquisition of basic raw data regarding diagnoses and exposures do not require randomized trials; and the patients admitted to an observational study can be selected in accordance with the same eligibility criteria and the same subsequent diagnostic procedures that would have been used in a randomized trial.²⁰ These scientific experimental principles, however, are still frequently disregarded in case-control research, despite the celebrated warning of the distinguished British statistician, Sir Austin Bradford Hill.²¹ In discussing the use of observational substitutes for experimental trials, he said that the investigator "must have the experimental approach firmly in mind" and must work "in such a way as to fulfill, as far as possible, experimental requirements."

ALVAN R. FEINSTEIN, MD
RALPH I. HORWITZ, MD
WALTER O. SPITZER, MD
RENALDO N. BATTISTA, MD

1. MacMahon B, Yen S, Trichopoulos D, et al: Coffee and cancer of the pancreas. *N Engl J Med* 1981;304:630-633.
2. Feinstein AR: Clinical biostatistics: XLVIII. Efficacy of different research structures in preventing bias in the analysis of causation. *Clin Pharmacol Ther* 1979;26:129-141.
3. Lin RS, Kessler II: A multifactorial model for pancreatic cancer in man. *JAMA* 1981;245:147-152.
4. Berkson J: Limitations of the application of four-fold tables to hospital data. *Biometrics Bull* 1946;2:47-53.
5. Neyman J: Statistics: Servant of all sciences. *Science* 1955;122:401.
6. Roberts RS, Spitzer WO, Delmore T, et al: An empirical demonstration of Berkson's bias. *J Chronic Dis* 1978;31:119-128.
7. Horwitz RI, Feinstein AR: Methodologic standards and contradictory results in case-control research. *Am J Med* 1979;66:556-564.
8. Feinstein AR: Methodologic problems and standards in case-control research. *J Chronic Dis* 1979;32:35-41.
9. Sackett DL: Bias in analytic research. *J Chronic Dis* 1979;32:51-63.
10. Horwitz RI, Feinstein AR, Stewart KR: Exclusion bias and the false relationship of reserpine/breast cancer, abstracted. *Clin Res* 1981;29:563.
11. Horwitz RI, Feinstein AR, Stremlau JR: Alternative data sources and discrepant results in case-control studies of estrogens and endometrial cancer. *Am J Epidemiol* 1980;111:389-394.
12. Horwitz RI, Feinstein AR: The problem of 'protopathic bias' in case-control studies. *Am J Med* 1980;68:255-258.
13. Gordis L: Assuring the quality of questionnaire data in epidemiologic research. *Am J Epidemiol* 1979;109:21-24.
14. Chambers LW, Spitzer WO, Hill GB, et al: Underreporting of cancer in medical surveys: A source of systematic error in cancer research. *Am J Epidemiol* 1976;104:141-145.
15. Chambers LW, Spitzer WO: A method of estimating risk for occupational factors using multiple data sources: The Newfoundland lip cancer study. *Am J Public Health* 1977;67:176-179.
16. Klemetti A, Saxen L: Prospective versus retrospective approach in the search for environmental causes of malformations. *Am J Public Health* 1967;57:2071-2075.
17. Spitzer WO, Hill GB, Chambers LW, et al: The occupation of fishing as a risk factor in cancer of the lip. *N Engl J Med* 1975;293:419-424.
18. Brown BW Jr, Hollander M: *Statistics: A Biomedical Introduction*. New York, John Wiley & Sons Inc, 1977, pp 231-234.
19. Fredrickson DS: The field trial: Some thoughts on the indispensable ordeal. *Bull NY Acad Med* 1968;44:985-993.
20. Horwitz RI, Feinstein AR: A new research method, suggesting that anticoagulants reduce mortality in patients with myocardial infarction. *Clin Pharmacol Ther* 1980;27:258.
21. Hill AB: Observation and experiment. *N Engl J Med* 1953;248:995-1001.