PREVENTIVE MEDICINE 16, 165–182 (1987)

Confounding[1]

STEVEN D. STELLMAN, PH.D.

American Cancer Society, 4 West 35th Street, New York, New York 10001

The magnitude of confounding is examined in nine case studies of two "weak" relationships: between artificial sweeteners and bladder cancer, and between oral contraceptives and cervical dysplasia. Confounding had little or no influence on the results of any published study. The responsible epidemiologist must always consider the possibility of confounding, no less when associations are weak than when they are strong. Identification of potentially confounding variables is an integral part of good epidemiologic practice. Rarely, however, does confounding itself, especially from unidentified sources, live up to its reputation for introducing seriously spurious associations. An investigator is more likely to be led astray by undetected biases than by pure confounding. © 1987 Academic Press, Inc.

INTRODUCTION

Confounding is the phenomenon whereby an association between two study variables, usually the exposure and the disease outcome it is thought to cause, is over- or underestimated through the agency of a third variable, the confounder. Confounding is the cause of great angst among epidemiologists, who go to great lengths searching out confounders and exorcising them through elaborate schemes.

While there is no escaping the central importance of confounding, many unnecessary and possibly damaging activities may take place in its name. Controversial studies are often criticized by appeal to hypothetical confounding factors, as often as not unnamed or unidentified, which, according to the critics, might have led the unwary investigators to an opposite conclusion had the factors been properly recognized. Investigators, anticipating such criticism, sometimes feel compelled to invoke excessively elaborate (and possibly inappropriate) models, or the latest in sophisticated software, to control large numbers of variables, sacrificing interpretability of results in the process.

In this article I tackle the question of just how bad confounding is, using case studies drawn from two major problem areas: the relation of artificial sweetener use to bladder cancer, and the relation of cervical dysplasia to oral contraceptive use. Both are examples of weak associations. There are over a dozen studies in each area from which to draw illustrative material, and there is reasonable consensus among epidemiologists about what the "right" answer is. Furthermore, recent reviews have been published for both (32, 40).

The two problem areas complement each other, since the subjects are drawn from greatly different age groups (bladder cancer cases are far older) and social strata (cervical dysplasia is more prevalent among those of lower socioeconomic

[1] From the Workshop on Guidelines to the Epidemiology of Weak Associations, American Health Foundation, New York, December 4–5, 1985.

0091-7435/87 $3.00

status). Spurious associations, if any, which could masquerade as confounding are related to different factors in the two examples.

STUDIES OF ARTIFICIAL SWEETENER USE AND BLADDER CANCER

In 1985 a National Academy of Sciences committee evaluated the carcinogenicity of cyclamate (32), which, together with saccharin, was used as an artificial sweetener (AS) in sugar substitutes, diet beverages, and diet foods. Table 1 shows the relative risk (RR) estimates, by sex, with confidence intervals. In 8 of the 10 studies, the RRs are very close to 1.0. Two studies (13, 27) found RRs significantly higher than 1.0, but for different sexes. While all were case–control studies, they were done in three different countries, involved exposures over different time periods, used different methods to assess exposure, differed greatly in selection of controls, and covered populations with vastly different distributions of confounding variables, such as smoking, occupation, and residence.

Study 1: Howe et al., 1977 (13); Miller and Howe, 1977 (24); Howe et al., 1980 (14)

Howe and co-workers conducted a population-based case–control study in three provinces of Canada. Cases consisted of 480 men and 152 women with bladder cancer. Controls were individually matched to cases on sex, age, and

TABLE 1
RELATIVE RISK FOR BLADDER CANCER AMONG ARTIFICIAL SWEETENER EVER-USERS

Study	Sex	No. exposed/total no. Cases	No. exposed/total no. Controls	Relative risk	Significance or 95% confidence interval
Morgan and Jain,	M	30/158	30/158	1.0	NS
1974 (28)	F	13/74	28/74	0.4	$P < 0.01$
Wynder and Goldsmith,	M	13/132	16/124	0.7	NG
1977 (46)	F	4/31	5/29	0.7	NG
Howe *et al.*,	M	73/480	47/480	1.6	$P = 0.009$
1977 (13)	F	18/152	30/152	0.6	NS
Kessler and Clark,	M	29/365	126/365	1.1	NS
1978 (15)	F	79/154	76/154	0.8	NS
Morrison and Buring,	M	NG	NG	0.8	0.6–1.1
1980 (30)	F	NG	NG	1.6	0.9–2.7
Hoover and Strasser,	M	909/2258	1723/3977	1.0	0.9–1.1
1980 (12)	F	384/742	732/1499	1.1	0.9–1.3
Wynder and Stellman,	M	76/302	77/299	0.9	0.7–1.3
1980 (47)	F	14/65	19/65	0.6	0.3–1.4
Najem *et al.*,					
1982 (31)	Both	12/74	19/142	1.3	0.6–2.8
Moller-Jensen *et al.*,	M	55/284	150/583	0.7	0.5–1.0
1983 (25)	F	26/96	50/193	1.1	0.6–1.9
Mommsen *et al.*,	M	NG	NG	NS	NG
1982, 1983 (26, 27)	F	6/47	2/97	6.7	1.5–30.2

Note. NS = not significant; NG = not given.
Source. Ref. (32). © 1985, by the National Academy of Sciences. Reprinted with permission.

neighborhood of residence. The exposure of interest was current or past use of sugar substitutes.

Potential confounding variables included education, occupational exposures, history of bladder or kidney diseases, kidney infection, smoking, instant and regular coffee use, and use of nonpublic water supply. Most of these variables were controlled at two levels. RRs were estimated by the methods of Miettinen (20) and of Pike *et al.* (35) for matched pairs; for those analyses in which the matching had to be broken because of small numbers of subjects, the Mantel–Haenszel procedure was used (17).

The finding of an RR of 1.6 in men, when first published in 1977, was at odds with all other human studies published to date, though not with three high-dose studies on rats which triggered a proposal to ban saccharin by the Food and Drug Administration (8). The RR cited was based on the standard matched-group method, using the ratio of discordant pairs. Through study design, the variables sex, age, race, and neighborhood of residence were matched, and therefore could not confound. Even when the matching was broken for subsequent analyses, the underlying distributions of those four variables were practically identical in cases and controls.

There are, of course, many other "potentially confounding" variables, some socioeconomic and others having to do with personal health history and lifestyle habits. Education, occupation, history of bladder and kidney infections, use of nonpublic water supply, smoking, and instant coffee use were controlled based on the observation that they were distributed differently among cases and controls. These adjustments had an inconsequential influence on the basic RR of 1.6.

The Howe study was published in the September 17, 1977, issue of *Lancet*. The same issue carried an editorial highly critical of the findings, citing, in particular, "potentially important confounding" which might have been introduced by failure to control more closely for smoking and instant coffee use. Replying to this criticism, Miller and Howe (24) reported unpaired RRs in individual strata of smoking and instant coffee consumption, shown here in Table 2.

The authors cited these RRs as evidence that neither exposure confounds the association with AS use. For coffee, that conclusion is certainly true, since this control results in practically no RR change. The situation is less clear with smoking, which shows that the elevated RR comes mainly from strata of heavy current or former smokers. This is difficult to explain. I favor the view that the result in ex-smokers is a manifestation of body weight: ex-smokers tend to gain weight (or to believe they do), and consume AS because of it, as other studies show AS use to be positively correlated with weight (37).

Cases and controls were similar in a very large number of other variables which were not matched by design: marital status; religion; years of residence in the province; consumption of milk, tea, fiddlehead greens, meats containing nitrates, nitrites; history of asthma, tuberculosis, nonbladder operations; illnesses other than urinary infections; use of analgesics; and cooperation and interest in the interview. Therefore, none of these variables, or numerous other unmeasured variables related to them, are likely to have been confounders.

Several years after their two initial publications, Howe and colleagues published a more detailed analysis of many bladder cancer risk factors (14), including

TABLE 2

RELATIVE RISK FOR BLADDER CANCER IN ARTIFICIAL SWEETENER USERS WITHIN STRATA OF INSTANT COFFEE AND CIGARETTE CONSUMPTION

	RR
Current instant coffee use	
None	1.6
Less than 1.5 cups per day	1.7
At least 1.5 cups per day	1.7
Summary RR	1.6 (P = 0.006)
Cigarette smoking	
Never smoked	0.7
Ex-smoker, under 5,000 packs	1.5
Ex-smoker, at least 5,000 packs	2.1
Current smoker, under 15 cigarettes/day	1.0
Current smoker, at least 15 cigarettes/day	1.7
Summary	1.7 (P = 0.01)

Source. Ref. (24). Reprinted with permission.

AS. Whereas their original analyses used simple Mantel–Haenszel style methods, this later paper contained elaborate sets of linear logistic models, allowing control for many extraneous factors, including those listed above that were deemed unlikely to confound anyway. Table 3 presents the models used and the RRs for AS consumption, first as a continuous variable (frequency per day) and then in four categories of dosage.

The RRs given in the final column of Table 3, for Model 4, reflect, though weakly, the dose–response relationship reported in the original article. In this analysis, 10 variables were controlled, and the RR at the highest dosage was 2.8 (P = 0.08). It would be interesting to examine what influence control for those 10 potential confounders had on these RR estimates; fortunately, these values were reported by Howe *et al.*: 1.2, 1.8, 1.4, and 2.8, respectively. At the highest dosage, P = 0.02. The authors reasonably concluded that, whatever other biases may have been present, their reported RRs, whether obtained by stratified analysis or obtained by logistic regression, were not tainted by unidentified confounding, as the editorial in *Lancet* had alleged.

No explanation has been posed for why this study differs from the majority of others in its finding of a significant elevation of risk among men. Chance is one possibility, and some unidentified bias is another. However, confounding is *not* a likely explanation. The results of the further analyses by Howe and colleagues demonstrate the futility of blaming an unusual or undesired result on unspecified confounding.

Study 2: Hoover and Strasser, 1980 (12)

This is the largest of all the AS studies reviewed. In fact, it is the largest epidemiologic bladder cancer study ever undertaken. It was designed as a case–control, population-based study, with interviews done in the patients' and controls' homes. The population base was the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program, along with the New

TABLE 3
LOGISTIC REGRESSION-DERIVED RELATIVE RISKS FOR BLADDER CANCER AMONG MALES IN RELATION TO ARTIFICIAL SWEETENER USAGE (× = VARIABLE IS CONTROLLED)

	Model 1	Model 2	Model 3	Model 4
Potential confounders				
Ex-smoking	×	×	×	×
Pipe smoking	×	×	×	×
Heavy inhalation of pipe smoke	×	×	×	×
Employment in high-risk industry	×	×	×	×
Exposure to dust or fumes in high-risk industry	×	×	×	×
Use of nonpublic water supply	×	×	×	×
History of bladder infection	×	×	×	×
History of diabetes	×		×	×
No. grades of school	×	×	×	×
Aspirin: lifetime use—tablets	×	×		×
Aspirin: 3 dosage categories			×	
Coffee: cups per day			×	
Coffee: 3 dosage categories	×			
Instant coffee: 3 dosage categories		×		
Regular coffee: 3 dosage categories		×		
Artificial sweetener				
Avg frequency per day	1.09	1.06	1.09	
0.1–4 per day				0.9
5–6 per day				1.6
7–8 per day				1.1
9–10 per day				2.8

Source: Ref. (14).

Jersey Cancer Registry, which together covered about one-tenth of the population of the United States. There were 2,258 male and 752 female bladder cancer cases.

The comparison group consisted of 4,337 males and 1,446 females drawn from an age- and sex-stratified random sample of the general population in the SEER areas, frequency-matched 2:1 to cases. To evaluate exposure, a detailed history of AS use in tabletop sweeteners, diet beverages, and diet foods was obtained from each subject during the interview.

Among the dozen or more confounding variables considered were age, sex, race, geographical region, smoking, coffee consumption, occupational exposure to chemicals (dye, leather, rubber, ink, paint), history of diabetes, education, obesity, use of hair dyes, and history of urinary infections. The principal statistical methods used were the stratified analyses of Gart (10) and Mantel–Haenszel, and the Mantel extension test for dose–response (16). The study presented results with Mantel–Haenszel adjustments for age, race, cigarette smoking, coffee drinking, and occupational exposures, or subsets of these variables.

The main findings of this study were absence of elevated RR in the group as a whole, but positive associations with AS use in certain otherwise low-risk subgroups. Many RR tables were given for groups satisfying various restrictions, such as for male heavy smokers and female nonsmokers who were also not occupationally exposed to known bladder carcinogens. While comparable RRs without adjustment for confounding were not presented, Hoover and Strasser stated explicitly that the adjusted RRs were not different from those crude estimates.

Study 2A: Walker et al., 1982 (43). A private organization, Epidemiology Resources, Inc., undertook a reanalysis of the Hoover–Strasser data, which were obtained under the Freedom of Information Act. The novel feature of this analysis is the concept of the *risk score*, a multivariate function which combines the effects of several risk factors through a logistic regression method. This technique, developed by Prentice and Pyke (36), is distantly related to the "confounder score" discriminant method proposed by Miettinen (23), and shares the conceptual feature of condensing information from many extraneous factors into a single variable.

The variables upon which the risk score was based were cigarette smoking (five levels), coffee consumption (four levels), history of bladder infection (yes/no), occupational exposure to known bladder carcinogens (yes/no), and education (two levels).

Walker's Table 2 (not reproduced here) enumerates the RRs for three different dosage measures of nonnutritive sweeteners: duration (five levels), amount (five levels), and cumulative dosage (six levels). At each of these 16 dosages the RR estimate was given three ways: crude, adjusted for the four study design variables (region, age, sex, race), and adjusted for those four variables plus five others: cigarette smoking, education, history of bladder infection, coffee consumption, and job exposure. At 13 of the 16 dosages the RRs were identical no matter how much adjustment was done; in the other three, the RR changed by 0.1 unit.

Walker *et al.* also presented the data from Hoover's analysis of low-risk white females, which was adjusted for age and restricted to nonsmokers not occupationally exposed to known bladder carcinogens. They further adjusted these data for geographical region, with results as shown in Table 4.

It can be seen from Table 4 that the shifts in RR are minor and fall within the published confidence intervals. Disagreement between the two teams of investigators centered around legitimacy of the risk score method itself [see Ref. (11)] and use of subgroup analysis. Confounding or potential confounding was scarcely mentioned, nor was there any reason for it to be; confounding was a negligible factor in this study.

Study 3: Morrison and Buring, 1980 (30)

The third artificial sweetener study which contains useful information on confounding is that of Morrison and Buring. It was also a case–control, population-based study, comprising cases of lower urinary tract cancer (mainly bladder cancer, but including cancers of the renal pelvis, ureter, and urethra). The sex distribution of the 592 cases was not given.

The comparison group was drawn from a random sample taken from resident lists of the Boston standard metropolitan statistical area (SMSA) towns from

TABLE 4
RELATIVE RISK FOR BLADDER CANCER WITH VARIOUS ADJUSTMENTS USING THE RISK SCORE METHOD

	Relative risk, adjusted for			
Category of artificial sweetener use	Age	Age and region	Age	Age and region
Daily consumption of:	Tabletop sweetener		Diet drink use	
Never used	1.0		1.0	
Ever used	1.2		1.1	
<1 use	0.9	1.1	0.9	1.0
1–1.9	1.2	1.4	1.2	1.3
2–2.9	1.8	2.0	1.6	1.9
3 or more	1.8	1.5	1.6	1.5
2 or more, <5 years	1.3	1.3	0.5	0.6
5–9 years	1.8	1.9	1.4	1.3
10 or more years	2.7	2.8	3.0	2.7

Source. Ref. (43). Reprinted with permission of American Journal of Public Health and Epidemiology Resources, Inc.

which the cases came, matched for age and sex, though not town. The exposure questions included a detailed history of consumption of specific sugar substitutes and diet beverages.

The seven potential confounding variables considered were age (under 65, 65–74, 75 and over), sex, smoking history, high-risk occupation or occupational exposure to known bladder carcinogens, age at leaving school, marital status, and religion. The principal statistical methods used were the RR calculations of Gart (10) and Miettinen (21), adjusting for age, sex, and smoking. Further stratification was done for occupation; and a subsidiary logistic regression analysis was done to control the remaining variables.

The RR estimates according to the level of control for combinations of confounders are presented in Table 5. It is evident that confounding played no discernible role in influencing the results of this study.

Study 4: Wynder and Stellman, 1980 (47)

As our final AS example, we consider the two studies of Wynder and Stellman. The published portion is a hospital-based, case–control study, with 302 male and 65 female bladder cancer cases. The comparison group consisted of other hospitalized patients without a diagnosis of a tobacco-related disease, individually matched from the control pool, on age, sex, hospital, and hospital room status (private, semiprivate, ward). Exposure was consumption of sugar substitutes and diet beverages. The 10 confounding variables considered included the four matching variables mentioned above, history of diabetes, occupation, obesity, religion, consumption of coffee and tea, and smoking history. RR estimates were made via the Miettinen matched-pair method (20), and t tests were used to compare means of AS consumption variables. The results were completely consistent with at least eight other studies [shown in Table 1 and summarized by the NAS/

TABLE 5
RELATIVE RISK (AND 95% CONFIDENCE INTERVALS) FOR BLADDER CANCER IN ARTIFICIAL SWEETENER USERS, ADJUSTED FOR DIFFERENT VARIABLES

		Men		Women	
"Exposure"	Adjusted for	RR	95% C.I.	RR	95% C.I.
Diet drinks	Age	0.8	0.6–1.1	1.6	0.9–2.7
	Age, smoking	0.9	0.7–1.2	1.8	1.0–3.3
	All variables (L.R.)	0.7	Not given	Not given	
Sugar substitute	Age	0.8	0.5–1.1	1.5	0.9–2.6
	Age, smoking	0.9	0.6–1.3	1.9	1.0–3.6
	All variables (L.R.)	0.8	Not given	Not given	

Source. Ref. (30). Reprinted with permission of The New England Journal of Medicine. L.R., logistic regression.

NRC Committee on Carcinogenicity of Cyclamate (32)]. Adjustment for the many variables noted above had no appreciable effect on risk estimates in this population.

A far more interesting and illustrative aspect of this project came out of a preliminary phase which has never been formally published. Case–control studies of tobacco-related cancers, including bladder cancer, have been conducted by the American Health Foundation since its founding in 1968. Every few years, data bases are consolidated and analyzed together. The study reported above was based on data collected from 1977 to 1979. However, a limited number of questions concerning AS usage had been introduced into an earlier version of the questionnaire, in 1974. Interim analyses yielded provocative results which the authors were reluctant to publish without further confirmation. Some of these results, however, were included in a progress report to the funding agency, the National Cancer Institute, and were made public. The basic findings are shown in Table 6.

TABLE 6
RELATIVE RISK FOR BLADDER CANCER IN RELATION TO CONSUMPTION OF ARTIFICIAL SWEETENERS (ANY FORM)

Sex	Variables adjusted (Mantel–Haenszel)	Relative risk	95% C.I.
Males (N = 402)	None	1.85	1.45–2.36
	Age, hospital, hospital room status, year of interview	1.43	1.10–1.86
	Above plus education	1.13	0.60–2.09
Females (N = 122)	None	0.99	0.61–1.59
	Age, hospital, hospital room status, year of interview	0.89	0.48–1.64
	Above plus education	0.80	0.20–2.98

Source. Ref. (47). Copyright 1980 by the AAAS. Reprinted with permission.

The adjustment cited in Table 6 brings the final RR estimates closer into line with those of the article published on the subsequent series. It is the only illustration available from AS studies of a bona fide case of identifiable confounding, for which appropriate adjustment through stratification removed the spurious effect. Furthermore, the authors recognized the artifactual nature of the elevated RR in men, and chose not to publish the results. Bross (5) proposed alternatively that the unpublished odds ratios represented the true "effect" of AS, but the authors have maintained their original position.

There is little doubt that confounding is at the core of these results. Where did it come from? Probably, it arose from "oversampling" from hospitals with patients of lower socioeconomic status (SES). Quotation marks are used because, in fact, sampling was done properly from the viewpoint of lung cancer, numerically the largest rubric in a study that included cancers of the lung, larynx, mouth, esophagus, and bladder. Bladder cancer in men occurs in higher socioeconomic strata than does lung cancer, so that if controls are drawn from the same hospitals as the cases cited above, they will tend to have the same (lower) SES to match the lung cancer cases. As SES was also a strong correlate of AS use, the controls in the first study thereby had lower levels of AS consumption, artificially raising the RR.

To correct this problem, the second study was begun afresh with rigorous one-to-one matching (rather than pool or frequency matching), on four variables, one of which was a socioeconomic indicator: hospital room status. When the time came for analysis, all bladder cancer cases, along with their specifically matched controls, were sequestered in a separate computer file, so that the SES distribution was automatically identical for cases and controls. The result was the same as what could be anticipated if a substantial number of high-SES controls had been included in the first study, and stratification was done on the basis of SES. In the second study, the number of controls was substantially smaller (and the relative study cost much lower).

In summary, the four AS studies considered in depth here are consistent with one another. Despite a wide variety of techniques for controlling confounding variables, and many differences in distributions of those variables in the underlying populations, control for confounding had very little effect on, and probably was not an important feature of, any published study. Confounding *was* important in a preliminary segment of one study which, in fact, was not published precisely because of this concern. The source of confounding was identified and eliminated by design, not analysis, in a subsequent study phase.

CERVICAL DYSPLASIA AND ORAL CONTRACEPTIVES

Numerous studies have been made of the possible influence of steroid contraceptive (birth control pill) use on neoplastic processes in the cervix uteri. I have used the vague term *neoplastic processes* deliberately, because the ambiguity of outcome is part of the analytical problem. Even though invasive cervical carcinoma is, theoretically, a possible result from pill use, its incidence is so low that few cases are available for study. The high rate of surveillance in the United States and the United Kingdom is at once a factor in the low rate of carcinoma as well as a serious biasing factor in assembling populations for study. Most investigators have elected to study, as a proxy outcome, various degrees of cervical dysplasia, based on the widely held belief that this condition is a necessary stage in the pathological development of cancer (38).

Table 7 presents a summary of the overall RRs for each of the studies reviewed by Swan and Petitti (40), some of which are considered in detail in this article. With regard to confounding, the most obvious feature of these studies is the much larger number of variables initially singled out as potentially troublesome and worthy of statistical control. This is an indication that, even in the earliest of these works, investigators were building on a knowledge base in which an understanding of the importance of the potential for confounding by these factors existed, in contrast to the AS studies, in which the relation of other variables to AS usage usually had to be determined anew at the start of each study (29).

Study 5: Thomas, 1972 (41)

In 1972, the statistically most elaborate study done up to that time was pub-

TABLE 7
STUDIES, NUMBER AND TYPE OF CASES, COMPARISON GROUPS, AND RELATIVE RISKS FOR CERVICAL DYSPLASIA IN ORAL CONTRACEPTIVE (OC) USERS

Study	No. cases	Groups compared	RR
Melamed *et al.*, 1969 (19)	207 CIS and invasive	Ever OC users vs diaphragm users	1.7
Thomas, 1972 (41)	324 Pap classes III, IV, V	Ever OC users vs all other methods	0.6
Worth and Boyes, 1972 (44)	310 CIS	Ever OC users vs all other methods	1.2
Boyce *et al.*, 1972, 1977 (1, 2)	196 CIS and invasive	Ever OC users vs all other methods	1.2
Meisels *et al.*, 1977 (18)	3,463 dysplasia	Ever OC users vs all other methods	1.0
Stern *et al.*, 1977 (38)	300 dysplasia	Progression of dysplasia to CIS in new users vs all other methods	—
Ory *et al.*, 1977 (33)	147 CIS, 854 dysplasia	Ever OC users vs IUD users	1.3–4.7
Peritz *et al.*, 1977 (34)	35 CIS, 31 dysplasia	Ever OC users vs other methods	1.2–4.7
Wright *et al.*, 1978 (45)	39 CIS and invasive, 26 dysplasia	Ever OC users vs IUD users vs diaphragm users	1.2–4.3
Swan and Brown, 1981 (39)	69 CIS and invasive	Ever OC users vs all other methods	0.6[a] 1.7[b]
Clarke *et al.*, 1985 (6)	250 dysplasia	Ever OC users vs never users	1.7

Note. CIS = Carcinoma *in situ*.
Source. Ref. (40). Reprinted with permission.
[a] Prevalent cases.
[b] Incident cases.

lished by Thomas. It was a case–control study based on the outcome of Pap smears from 324 women ages 15–50, which fell into classes III, IV, or V. The exposure consisted of the type of oral contraceptive (OC), number of months of use, and cumulative dose.

The comparison group consisted of 360 women randomly drawn from those in the same Maryland county who had at least one smear during 1965–1969. Thirteen variables were considered as potentially confounding. All were considered as binary or yes/no variables: age 30+, husband circumcised, barrier contraceptives, smoker, attend church weekly, trichomonads on smear, history of vaginal discharge, high school education, marital instability, any husband previously married, three or more live births, child conceived before marriage, first pregnancy before age 20.

The statistical method was as follows: first the 2 × 2 association between OC use and each of the 13 above variables was displayed in cases and controls separately, and significant differences were noted. OC users differed significantly from nonusers in both cases and controls only with respect to two variables: age and first pregnancy before age 20.

Unadjusted RRs were then computed using the method of Cornfield (7), followed by adjustment using an unspecified direct method for each of the 13 confounders, one at a time. Finally, a multiple regression technique attributed to Feldstein (9) was used to adjust for all 13 simultaneously.

Thomas's Table 6 (not reproduced here) gives the RRs for four categories of outcome [carcinoma *in situ* (CIS), dysplasia, no biopsy, and total cases] adjusted for each variable separately, and adjusted for none and for all simultaneously. The RR of 0.78 for CIS fell to 0.58 after adjustment for all 13 variables; the RR for all cases went from 1.09 to 0.91. Adjustment for individual variables pushed the RR up or down a little, but the multivariate adjustment produced the farthest excursion. For all four outcomes, adjustment lowered the RR, reinforcing the conclusion that OC use did not increase risk for cervical carcinoma or its pathologic precursors.

Study 6: Peritz et al., 1977 (34)

The Kaiser–Permanente Walnut Creek group reported a positive association between cervical carcinoma incidence and total duration of OC use. The results were based on 31 cases of dysplasia and 35 cases of CIS observed in a sample of 17,942 women, mostly ages 20–54 years, followed prospectively. The comparison population consisted of nonusers enrolled in the same health plan as the users.

Adjustment was made for eight confounding variables, measured at entry into the study (with the number of levels): age (4), education (3), marital status (2), number of Pap smears (2), religion (2), number of pregnancies (2), smoking (2), history of certain infections (2).

Instead of the traditional statistical methods, two new methods were proposed and used, each involving a rank-order z statistic; no reason was given for avoiding standard prospective methods. Both procedures involve classifying women into like strata based on similar values of the confounding variables, making this, in principle, a variant matched-group method. RRs were calculated at each duration level, but trend analysis was not performed. Distributions of the confounders in cases and noncases was presented [Table 2, Ref. (34)], so that stratum-specific RRs could be calculated.

The effect of adjustment for age and other variables is given in Table 8. (The RR here is assumed for CIS; the source table in Peritz *et al.* is not clearly labeled.)

Although comparison between tables is hindered by an unexplained shift in duration categories, the changes in RR brought about by adjustment for different numbers of variables is trivial. There is no practical difference between the 4+ years RRs shown in Table 8, and any standard analysis would have concluded—as these authors did—that an effect exists. The interpretation and generalization of results depends more on consideration of questions of bias than of confounding, especially since there seems to be some uncertainty as to the exclusion of women with hysterectomies from the population, which would certainly affect risk estimates.

Study 7: Wright et al., 1978 (45)

In a study in the United Kingdom, 17,032 white married women, ages 25–39, attending family planning clinics were followed prospectively from 1968 through 1977. The purpose of the study was to investigate the reduced risk among diaphragm users, rather than increased risk among OC users, as was the case with other studies considered here. A supplemental case–control substudy was done within the cohort to examine in greater detail (and with greater efficiency) a subgroup of subjects who agreed to give intensive interviews, which included information about sexual activity not available from routine records for the other subjects.

The comparison group for the cohort study consisted of women who used either the IUD or diaphragm for contraception. Five confounding variables were considered (with the number of levels): age (4), social class (3), smoking (3), age first married (4), age first pregnant (5). For the case–control study, age at first coitus (4) and number of partners (2) were added.

In the cohort study, 26 dysplasias, 33 CIS, and 6 cases of invasive carcinoma were observed. The statistical method consisted of comparison of incidence rates

TABLE 8

RELATIVE RISK FOR CARCINOMA *in Situ* AMONG ORAL CONTRACEPTIVE (OC) USERS ACCORDING TO YEARS OF USAGE

	Duration of OC use			
RR	0	To 1 year	1–4 years	4+ years
Crude	1.0	1.4	1.8	4.2
Age-adjusted	1.0	2.0	3.0	5.4
	0	To 2 years	2–4 years	4+ years
5 variables controlled	1.0	2.0	1.8	4.0
8 variables controlled	1.0	2.2	2.0	5.1

Source. Ref. (34). Reprinted with permission.

between levels of confounders one at a time, using a chi-square test, and an unnamed trend test [probably Mantel (16)]. Smoking and age first married were significantly associated with incidence.

Incidence rates were given at each level of each confounder, taken one at a time, with chi-square tests for trend. In addition, an indirect method (42) was used to adjust for the five other factors, with the results shown in Table 9. Despite large intercorrelations between variables, adjustment made very little difference in rates.

In the case–control study, 52 cases and 139 controls were matched by age and clinic. RRs were estimated for combinations of exposure and confounders using logistic regression of Breslow *et al.* (3). Values of the estimated RR for cervical neoplasia in diaphragm users, relative to users of other methods, are given for various combinations of confounders in Table 10. The crude RR of 0.13 increased to 0.23 when smoking, age at first intercourse, and number of partners were added to the model (Table 10). Significant values in the last column of the table identify two important risk factors in their own right: cigarette smoking and number of partners. Age at first intercourse is the most important confounder, since it confers the greatest increase on the crude RR, but once it is absorbed into the model, its influence on the association between diaphragm use and cervical neoplasia is not significant.

Study 8: Swan and Brown, 1981 (39)

This second offering from the Walnut Creek group is a case–control study of CIS or severe dysplasia in relation to OC use and sexual activity. The study group consisted of 26 prevalent plus 43 incident cases. The exposure variable was "years of OC use."

The comparison group consisted of three controls per case, matched on age, study entry date, number of Pap smears at entry, marital status, religion, education, parity, and year of menopause. Besides the eight matching variables, six other potential confounders considered were age first married, age first pregnant, age regular coitus, number of sexual partners, contraceptive method, and history of certain infections.

The statistical method consisted first of displaying mean values of confounding variables for cases and matched controls separately for prevalent and incident cases. An RR analysis was done with the Miettinen confounder score method (23). A sexuality index, comprising a linear combination of sexual activity variables, was plotted in cases vs controls. The RR was 1.7 for incident cases and 0.6

TABLE 9
INCIDENCE PER 1,000 WOMAN-YEARS OF OBSERVATION

	Crude	Adjusted
Oral contraceptives	0.95	0.91
Diaphragm	0.17	0.19
IUD	0.87	0.93

Source. Ref. (45). Reprinted with permission.

TABLE 10
RELATIVE RISK AND CHI-SQUARE FOR CERVICAL NEOPLASIA IN DIAPHRAGM USERS, RELATIVE TO NONUSERS, ADJUSTING FOR A NUMBER OF VARIABLES

Model	RR	Chi-square	Chi-square for effect of other specified variables
C.M. alone	0.13	12.96***	—
C.M. plus social class	0.13	12.63***	0.72 (2 *df*)
C.M. plus cigarette smoking	0.14	10.13**	10.34 (2 *df*)**
C.M. plus age first marriage	0.15	9.64**	3.55 (3 *df*)
C.M. plus age first pregnant	0.14	10.48**	7.55 (3 *df*)
C.M. plus age first intercourse	0.19	6.11*	7.01 (3 *df*)
C.M. plus number of partners	0.15	9.30**	11.69 (1 *df*)***
C.M. plus cigarette smoking, age first intercourse, number of partners	0.23	4.33*	23.81 (6 *df*)***

* $P < 0.05$.
** $P < 0.01$.
*** $P < 0.001$.
Note. C.M. = contraceptive method. *df* = degrees of freedom.
Source. Ref. (45). Reprinted with permission.

for prevalent cases. A major determinant of this difference was the greater probability of screening for OC users compared with that of nonusers, and indeed it is this demonstrated ascertainment bias that colors the entire analysis.

The RR estimates as determined by the confounder score method were presented in specific strata of the sexuality index, according to OC use and whether prevalent or incident, making 24 cells in all. Unfortunately, small numbers of subjects made comparisons difficult. The authors' conclusion was that "adjusting for sexual factors explains some, but not all, of the relative risk attributable to OC use." That is, consideration of those factors removed some, but not all, of the confounding.

This study is the most innovative, aggressive, and perhaps successful in attempting to account for the confounding influence of sexual variables. The method still lacks true multivariate control (as one would obtain with logistic regression), because the Miettinen method and use of the linear combination sexual index attempt to condense multidimensional information onto a single axis. The small sample size makes firmer conclusions difficult.

Study 9: Clarke et al., 1985 (6)

This Canadian case–control study was designed to investigate the relationship between cervical dysplasia and the risk factors OC use, sexual behavior, and cigarette smoking. There were 250 cases of biopsy-confirmed dysplasia (CIS excluded). Duration of OC use was the main exposure under study. Considered as confounding variables were (with the number of levels) education (2), number of sexual partners (3), age at first coitus (2), smoking (3), and number of cigarettes smoked per day (3).

The comparison group consisted of 500 neighborhood controls matched on age, presence of a uterus, neighborhood of residence, and whether they lived in a house or an apartment.

The statistical method consisted of computing an unmatched RR using the Miettinen odds ratio (22) for confounders considered one at a time. In addition, a matched logistic regression analysis was done (4). The overall RRs for cervical dysplasia with respect to OC use were 3.0 (Miettinen) or 2.6 (logistic regression).

Table 11 presents the crude RRs in relation to each of the variables, along with the results of adjustment using logistic regression. When confounders were added to the logistic regression model (which always included OC use), their associated RRs were nearly identical to the crude RRs without OC use, with the following important exceptions: the crude RR for 6 or more sexual partners (relative to one partner) rose from 5.1 to 6.9 after OC use (ever vs never) was added to the model; and the RR for OC use fell from 3.0 to 1.7 when number of sexual partners was taken into account.

The latter observation, that the OC RR lost its statistical significance after adjustment for sexual partners, was taken by the authors as an indication that OC use itself was probably not a true causal factor. This conclusion might be correct, but it is equally likely, based on these data, that the residual risk (1.7) is a true one. The small number of cases reduces the statistical power ($P = 0.14$), and the

TABLE 11
RELATIVE RISK FOR CERVICAL DYSPLASIA IN ORAL CONTRACEPTIVE USERS, COMPARED WITH NONUSERS, ADJUSTED FOR DIFFERENT VARIABLES

		Logistic regression	
Variable	Crude RR (matched data)	Adjusted for	RR
Number of sexual partners			
1	1.0		
2–5	3.4**	OC use	3.4**
6 or more	5.1**	OC use	6.9**
Age first intercourse			
Under 18	1.0		
18 or older	2.7**		
18 or older		OC use	2.8**
Cigarette smoking			
Never	1.0		
Ever	3.2**		
Current		OC use	3.1**
Oral contraceptives			
Never	1.0		
Ever	3.0**	—	2.6**
Ever		Smoking	2.7**
Ever		Age first intercourse	2.3**
Ever		Number sexual partners	1.7[a]

** $P < 0.01$.
Source. Ref. (6). Reprinted with permission.
[a] $P = 0.14$.

ultimate RR level is more in line with that found in the majority of cervical dysplasia studies considered in this article. Nevertheless, this is the most clear-cut case of true confounding observed in the entire spectrum of studies.

SUMMARY

Two examples of weak associations have been explored in which a large number of studies have been done: artificial sweeteners in relation to bladder cancer, and oral contraceptives in relation to cervical dysplasia. The consensus among epidemiologists that has developed in each example area is remarkable, because it is based on a wide variety of study designs, populations, analytic techniques, and choice of confounding variables. Even those studies that differ in individual respects from the others (6, 13) can be viewed as mostly consistent with the consensus with respect to magnitude of effects.

It is unlikely that the overall judgment of investigators, as summarized by extensive reviews (32, 40), is incorrect; if there are inadvertent inaccuracies due to misunderstanding of data, it is extremely unlikely that they are due to failure to take confounding into account. The unresolved issues in the great majority of these studies have to do with bias and selection factors.

The most serious potential mistakes that might be made based on erroneous interpretation of these and similar types of data would have to do with regulating the products involved, which here involve a prescription drug (OC) and a consumer product (AS). The decision-making process, at least for drug regulation, involves weighing of risks and benefits. To the extent that these studies have placed reasonable confidence limits about the risks, they have contributed to the knowledge base for making those decisions, even though the associations are "weak."

REFERENCES

1. Boyce, J. G., Lu, T., Nelson, J. H., and Joyce, D. Cervical carcinoma and oral contraception. *Obstet. Gynecol.* **40**, 139–146 (1972).
2. Boyce, J. G., Lu, T., Nelson, J. H., and Fruchter, R. G. Oral contraceptives and cervical carcinoma. *Amer. J. Obstet. Gynecol.* **128**, 761–766 (1977).
3. Breslow, N. E., Day, N. E., Halvorsen, K. T., Prentice, R. L., and Sabal, C. "Estimation of Multiple Relative Risk Functions in Matched Case–Control Studies." Technical Report No. 13, Department of Biostatistics, School of Public Health and Community Medicine, Seattle, WA, 1977.
4. Breslow, N. E., and Day, N. E. "Statistical Methods in Cancer Research. I. The Analysis of Case–Control Studies." International Agency for Research on Cancer, Lyon, 1980.
5. Bross, I. D. J. Bladder cancer and artificial sweeteners: A methodological issue. *Science* **210**, 447 (1980).
6. Clarke, E. A., Hatcher, J., McKeown-Eyssen, G. E., and Lickrish, G. M. Cervical dysplasia: Association with sexual behavior, smoking and oral contraceptive use? *Amer. J. Obstet. Gynecol.* **151**, 612–616 (1985).
7. Cornfield, J. A method of estimating comparative rates from clinical data: Application to cancer of the lung, breast, and cervix. *J. Natl. Cancer Inst.* **11**, 1269–1275 (1951).
8. Epstein, S. S. "The Politics of Cancer." Sierra Club, New York, 1978.
9. Feldstein, M. S. A binary variable multiple regression method of analyzing factors affecting perinatal mortality and other outcomes of pregnancy. *J. Roy. Soc. Statist. Soc. Ser. A* **129**, 61–73 (1966).

10. Gart, J. J. Point and interval estimation of the common odds ratio in the combination of 2 × 2 tables with fixed marginals. *Biometrika* **57**, 471–475 (1970).
11. Hoover, R. N., and Hartge, P. Non-nutritive sweeteners and bladder cancer *Amer. J. Public Health* **72**, 382–383 (1982) (editorial).
12. Hoover, R. N., and Strasser, P. H. Artificial sweeteners and human bladder cancer: Preliminary results. *Lancet* **1**, 837–840 (1980).
13. Howe, G. R., Burch, J. D., Miller, A. B., Morrison, B., Gordon, P., Weldon, L., Chambers, L. W., Fodor, G., and Winsor, G. M. Artificial sweeteners and human bladder cancer. *Lancet* **2**, 578–581 (1977).
14. Howe, G. R., Burch, J. D., Miller, A. B., Cook, G. M., Esteve, J., Morrison, B., Gordon, P., Chambers, L. W., Fodor, G., and Winsor, G. M. Tobacco use, occupation, coffee, various nutrients, and bladder cancer. *J. Natl. Cancer Inst.* **64**, 701–713 (1980).
15. Kessler, I. I., and Clark, J. P. Saccharin, cyclamate, and human bladder cancer. *JAMA* **240**, 349–355 (1978).
16. Mantel, N. Chi-square tests with one degree of freedom: Extension of the Mantel–Haenszel procedure. *J. Amer. Stat. Assoc.* **58**, 690–700 (1963).
17. Mantel, N., and Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of diseases. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
18. Meisels, A., Begin, R., and Schneider, V. Dysplasias of the uterine cervix. Epidemiological aspects: Role of age at first coitus and use of oral contraceptives. *Cancer* **40**, 3076–3081 (1977).
19. Melamed, M. R., Koss, L. G., Flehinger, B. J., Kelisky, R. P., and Dubrow, H. Prevalence rates of uterine cervical carcinoma *in situ* for women using the diaphragm or contraceptive oral steroids. *Brit. Med. J.* **3**, 195–200 (1969).
20. Miettinen, O. S. Estimation of relative risk from individually matched series. *Biometrics* **26**, 75–86 (1970).
21. Miettinen, O. S. Standardization of risk ratios. *Amer. J. Epidemiol.* **96**, 383–388 (1973).
22. Miettinen O. S. Estimability and estimation in case-referent studies. *Amer. J. Epidemiol.* **103**, 226–235 (1976).
23. Miettinen O. S. Stratification by a multivariate confounder score. *Amer. J. Epidemiol.* **104**, 609–620 (1976).
24. Miller, A. B., and Howe, G. R. Artificial sweeteners and bladder cancer. *Lancet* **2**, 1221–1222 (1977).
25. Moller-Jensen, O., Knudsen, J. B., Sorensen, B. L., and Clemmesen, J. Artificial sweeteners and absence of bladder cancer risk in Copenhagen. *Int. J. Cancer* **32**, 577–582 (1983).
26. Mommsen, S., Aagaard, J., and Sell, A. An epidemiological case–control study of bladder cancer in males from a predominantly rural district. *Eur. J. Cancer Clin. Oncol.* **18**, 1205–1210 (1982).
27. Mommsen, S., Aagaard, J., and Sell, A. A case–control study of female bladder cancer. *Eur. J. Cancer Clin. Oncol.* **19**, 725–729 (1983).
28. Morgan, R. W., and Jain, M. G. Bladder cancer: Smoking, beverages and artificial sweeteners. *Canad. Med. Assoc. J.* **111**, 1067–1070 (1974).
29. Morrison, A. S. Use of artificial sweeteners by cancer patients. *J. Natl. Cancer Inst.* **62**, 1397–1399 (1979).
30. Morrison, A. S., and Buring, J. E. Artificial sweeteners and cancer of the lower urinary tract. *New Engl. J. Med.* **302**, 537–541 (1980).
31. Najem, G. R., Louria, D. B., Seebode, J. J., Thind, I. A., Prusakowski, J. M., Ambrose, R. B., and Fernicola, A. R. Life time occupation, smoking, caffeine, saccharine, hair dyes and bladder carcinogenesis. *Int. J. Epidemiol.* **11**, 212–217 (1982).
32. National Research Council. Committee on the Evaluation of Cyclamate for Carcinogenicity, Commission of Life Sciences. "Evaluation of Cyclamate for Carcinogenicity." National Academy Press, Washington, DC, 1985.
33. Ory, H. W., Conger, S. B., Naib, Z., Tyler, C. W., and Hatcher, R. A. Preliminary analysis of oral contraceptive use and risk of developing premalignant lesions of the uterine cervix, *in* "Pharmacology of Steroid Contraceptive Drugs" (S. Garattini and H. W. Berendes, Eds.), pp. 211–218. Raven Press, New York, 1977.
34. Peritz, E., Ramcharan, S., Frank, J., Brown, W. L., Huang, S., and Ray, R. The incidence of

cervical cancer and duration of oral contraceptive use. *Amer. J. Epidemiol* **106**, 462–469 (1977).

35. Pike, M. C., Casagrande, J., and Smith, P. G. Statistical analysis of individually matched case–control studies in epidemiology: Factor under study a discrete variable taking multiple values. *Brit. J. Prev. Soc. Med.* **29**, 196–201 (1975).
36. Prentice, R. L., and Pyke, R. Logistic disease incidence models and case–control studies. *Biometrika* **66**, 403–411 (1979).
37. Stellman, S. D., and Garfinkel, L. Artificial sweetener use and 1-year weight change in women. *Prev. Med.* **15**, 195–202 (1986).
38. Stern, E., Forsythe, A. B., Youkeles, L., and Coffelt, C. F. Steroid contraceptive use and cervical dysplasia: Increased risk of progression. *Science* **196**, 1460–1462 (1977).
39. Swan, S. H., and Brown, W. L. Oral contraceptive use, sexual activity, and cervical carcinoma. *Amer. J. Obstet. Gynecol.* **139**, 52–57 (1981).
40. Swan, S. H., and Petitti, D. B. A review of problems of bias and confounding in epidemiologic studies of cervical neoplasia and oral contraceptive use. *Amer. J. Epidemiol.* **115**, 10–18 (1982).
41. Thomas, D. B. Relationship of oral contraceptives to cervical carcinogenesis. *Obstet. Gynecol.* **40**, 508–518 (1972).
42. Vessey, M. P., Doll, R., Peto, R., Johnson, B., and Wiggins, P. A long-term follow-up study of women using different methods of contraception—An interim report. *J. Biosoc. Sci.* **8**, 373 (1976).
43. Walker, A. M., Dreyer, N. A., Friedlander, E., Loughlin, J., Rothman, K. J., and Kohn, H. I. An independent analysis of the National Cancer Institute Study on non-nutritive sweeteners and bladder cancer. *Amer. J. Public Health* **72**, 376–381 (1982).
44. Worth, A. J., and Boyes, D. A. A case–control study into the possible effects of birth control pills on pre-clinical carcinoma of the cervix. *J. Obstet. Gynecol. Brit. Commonw.* **79**, 673–679 (1972).
45. Wright, N. H., Vessey, M. P., Kenward, B., McPherson, K., and Doll, R. Neoplasia and dysplasia of the cervix uteri and contraception: A possible protective effect of the diaphragm. *Brit. J. Cancer* **38**, 273–279 (1978).
46. Wynder, E. L., and Goldsmith, R. The epidemiology of bladder cancer: A second look. *Cancer* **40**, 1246–1268 (1977).
47. Wynder, E. L., and Stellman, S. D. Artificial sweetener use and bladder cancer: A case–control study. *Science* **207**, 1214–1216 (1980).