# T H E B FILES

## Case studies of bias in real life epidemiologic studies

Bias File 4. The early controversy over estrogen and endometrial 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

## Bias File 3. The early controversy over estrogen and endometrial cancer

**The story**

Exogenous unopposed estrogen (i.e. without progestin) use is now known to substantially increase the risk of endometrial cancer. But in the 1970s and early 80s, this was a very contentious and controversial issue. Several case-control studies reported a strong association between estrogen use and endometrial cancer, especially in women taking estrogen regularly for a number of years. Most investigators were convinced that this was a causal association. However, a few investigators disagreed. They argued that estrogens were merely causing the cancers to be diagnosed rather than to occur (Horwitz & Feinstein, 1978). In other words, they argued that "detection bias" explained the strong associations that were found in these studies. Estrogens induce uterine bleeding, even in healthy women. Therefore, women who regularly took estrogen are probably more likely to seek medical attention because of bleeding, therefore more likely to be worked up by physicians, thus causing a variety of gynecological conditions (including sub-clinical, symptomless or occult endometrial cancer) to be detected earlier or in some cases detected when they otherwise would have remained undetected. This was referred to as detection or diagnostic surveillance bias. Who was correct and how did the controversy get resolved? Several textbooks have nice descriptions of this controversy, including Kelsey (1996), Weiss (2006), and Rothman, Greenland & Lash (2008)

**The study**

As an example of a case-control study that found a strong association between estrogen and endometrial cancer, see the paper by Mack et al (NEJM 1976) which used community controls. In this study "*all cases of endometrial cancer occurring among the residents of an affluent retirement community were compared with controls chosen from a roster of all women in the same community. Evidence of estrogen and other drug use and of selected medical conditions was obtained from three sources: medical records of the principal care facility, interviews, and the records of the local pharmacy. The risk ratio for any estrogen use was estimated from all available evidence to be 8.0 (95% CI 3.5 to 18.1). and the for conjugated estrogen use to be 5.6 (95% CI, 2.8 to 11.1). Increased risk from estrogens was shown for invasive as well as noninvasive cancer, and a dose-response effect was demonstrated*."

Similar strong associations were also seen in case-control studies that used hospital controls. As an example, Kelsey et al. conducted a hospital-based case-control study of the epidemiology of endometrial cancer in women aged 45-74 years in Connecticut from 1977 to 1979. In total, 167 cases and 903 controls were included. Controls were chosen from among patients admitted to inpatient surgical services, excluding gynecology. Estrogen therapy was strongly associated with endometrial cancer. The odds ratio was 8.2, when estrogen was used up to 10 years.

**The bias**

In 1978, Ralph Horwitz and Alvan Feinstein published a paper in the *New England Journal of Medicine*, arguing that detection bias may have led to an overestimation of the effect in many case-control studies. What potential solution did Horwitz and Feinstein propose? They proposed an alternative method of control sampling, by letting cases and controls emerge from a group of women referred to the hospital for the same intra-endometrial diagnostic procedure (dilatation and curettage [D&C] and biopsy). In other words, they suggested using as controls, women who were worked up for benign

gynecological disorders. They reasoned that benign conditions would also be subject to detection bias, and therefore recommended a control series of women undergoing D & C procedures and did not have endometrial cancer. The purpose, according to Horwitz and Feinstein is to "try and equalize the forces of 'diagnostic surveillance' that might otherwise create major 'detection bias' in conventional case-control studies." By using the alternative control selection approach, they showed the effect estimate approached a value closer to null (OR 1.7) when cases were compared with controls who had all received D & C or hysterectomy because of uterine bleeding.

Horwitz and Feinstein's NEJM paper was accompanied by an editorial by Hutchison and Rothman (1978), in which they disagreed with the authors. Instead of correcting the detection bias, the proposed control selection process actually introduces another strong selection bias, Hutchison and Rothman argued. They suggested that some of the benign gynecological conditions that cause bleeding among control women could actually be induced by estrogen use. This would result in a falsely highly frequency of estrogen exposure in the control group and grossly under-estimate the estrogen-cancer association. In short, the control selection approach suggested by Horwitz and Feinstein addressed one source of error (detection bias), but created a new bias that previously did not exist (selection bias).

Another remedy that Horwitz and Feinstein proposed was to examine the association with women who had presented with vaginal bleeding or had undergone treatment for such bleeding. This was also problematic. As pointed out by Greenland and Neutra (1981), because both the exposure (estrogens) and the disease (endometrial cancer) strongly increased the risk of vaginal bleeding, restricting the study to women with bleeding results in a "Berksonian bias" that can easily diminish the observed odds ratio.

As the story evolved, investigators attempted using multiple control groups, to confirm or refute the arguments of Horwitz and Feinstein. Hulka et al (1980) conducted a case-control study to address the issue of detection bias among endometrial cancer cases and controls. In this study, "women admitted to the North Carolina Memorial Hospital for dilatation and curettage (D&C) during 1970-1976 were selected as one of three control groups in a study of endometrial cancer and exogenous estrogen. Study subjects included 256 cases, 316 D&C controls, 224 gynecology controls and 321 community controls. The D&C controls had a higher frequency of estrogen use than either of the other control groups or the cases. These differences existed for both blacks and whites. When white cases were compared to either gynecology or community controls, relative risks were increased for long duration estrogen use and for recent use prior to diagnosis. With D&C controls, relative risks were not significantly different from unity irrespective of duration or recency of estrogen use. Exclusion of hyperplasias from the D&C controls had no substantive effect on these results. Bleeding was a presenting complaint for 92% of cases, 82% of D&C controls and 22% of gynecology controls. Both among cases and gynecology controls, there was no statistically significant association between bleeding and estrogen use, whereas this association was evident among D&C controls, and specifically among those who did not have pathologic evidence of endometrial hyperplasia." Hulka and colleagues concluded that their study supported the presence of detection bias among D&C controls but did not provide evidence of this bias among endometrial cancer cases. This study, therefore, reinforced the belief in the estrogen-cancer association. The study is also a nice example of use of different control groups and how to reconcile discrepant results when multiple control groups are used in case-control studies.

Ultimately, as pointed out by Weiss, the importance of the proposed detection bias hinges on the prevalence of endometrial cancer in postmenopausal women that goes undetected in the absence of D&C or endometrial biopsy. In a review of an autopsy series, Horwitz et al (1981) asserted that they had

identified a large proportion of occult cases, relative to those diagnosed during life. However, in their study, they compared the prevalence at autopsy with the incidence during life, without realizing that these two measures do not have the same units. Taking into account the hypothesized long duration of the asymptomatic cases they identified in their prevalence sample, their own data actually support the argument that most cases of endometrial cancer that develop indeed do wind up getting diagnosed during life (Weiss NS, 2009, personal communication,).

Since the 1980s, evidence has accumulated, showing a strong positive association between estrogen use and endometrial cancer. It is interesting, that even in 1986, Horwitz and Feinstein (1986) continued to argue the position they took in 1978. In their 1986 paper they stated that "after considering all of the pertinent data, we have no reason to modify our conclusions in 1987, "that the strength of the much-publicized association between estrogens and endometrial cancer has doubtlessly been exaggerated and needs reevaluation.""

**The lesson**

There are several lessons to be learnt from this controversy. First is the critical issue of control selection in case-control studies, especially when hospital controls are recruited. There is a risk of inducing or worsening selection bias whenever we use specific criteria such as presence or absence of certain conditions among the control groups, especially if such conditions are associated with the exposure under study. If those criteria are also related to the study disease, severe Berksonian bias is likely to ensue (Rothman et al. 2008, pp 115-122). Berksonian bias is a type of selection bias that occurs when both the exposure and the disease affect selection probabilities and specifically because they affect selection (Rothman et al 2008, pp. 134-137).

The other lesson from the controversy is the need to think about detection bias, where the exposure can lead to more intensive work-up or diagnostic surveillance, and result in over-estimation of the underlying effect. Horwitz and Feinstein did highlight a valid concern, and detection bias may have explained at least some of the observed association. However, the approach to defining controls that they proposed almost certainly leads to a substantial degree of another type of bias (see above). Nonetheless, as Weiss points out, "despite the fact that the strategy (of choosing controls from patients undergoing the diagnostic test(s) for the suspected adverse effect) did not "work" for endometrial cancer, there probably are some situations in which its use has merit. One might be the study of a possible estrogen-gallstone disease association, for there is no reason to believe that an association exists between estrogen use and the presence of symptoms that led to a negative cholecystogram or ultrasound test." (Weiss 2006).

Lastly, this case study also illustrates the value of considering multiple control groups, especially when there is lack of clarity on the single best group to use as controls. However, if the effect estimate vary considerably by the control group used, then much effort must be put into trying to understand and explain the discrepancies.

Robins (2001) provides an explanation of this controversy using directed acyclic graphs (DAGs), which help to clarify the two sources of bias, and the reasons that the proposed Horwitz and Feinstein case-control study would trade in one type of bias for another.

**Sources and suggested readings***

1. Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. N Engl J Med. 1978 Nov 16;299(20):1089-94.
2. Rothman K, Greenland S, Lash T. Modern epidemiology. Lippincott Williams & Wilkins, 3rd edition, 2008.
3. Kelsey J et al. Methods in observational epidemiology. 2nd edition. Oxford Univ Press, 1996.
4. Weiss NS. Clinical epidemiology. Oxford Univ Press, 2006.
5. Mack TM, et al. Estrogens and endometrial cancer in a retirement community. N Engl J Med. 1976 Jun 3;294(23):1262-7.
6. Kelsey JL, et al. A case-control study of cancer of the endometrium. Am J Epidemiol. 1982 Aug;116(2):333-42.
7. Hutchison GB, Rothman KJ. Correcting a bias? N Engl J Med. 1978 Nov 16;299(20):1129-30.
8. Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. J Chronic Dis. 1981;34(9-10):433-8.
9. Hulka BS, et al. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. Am J Epidemiol. 1980 Sep;112(3):376-87.
10. Horwitz RI, Feinstein AR, Horwitz SM, Robboy SJ. Necropsy diagnosis of endometrial cancer and detection-bias in case/control studies. Lancet. 1981 Jul 11;2(8237):66-8.
11. Horwitz RI, Feinstein AR. Estrogens and endometrial cancer. Responses to arguments and current status of an epidemiologic controversy. Am J Med. 1986 Sep;81(3):503-7.
12. Robins JM. Data, design, and background knowledge in etiologic inference. Epidemiology. 2001 May;12(3):313-20.

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

**Acknowledgement**

8. Siiteri PK, Schwarz BE, MacDonald PC: Estrogen receptors and the es-
   trone hypothesis in relation to endometrial and breast cancer. Gynecol
   Oncol 2:228-238, 1974
9. Wynder EL, Escher GC, Mantel N: An epidemiological investigation of
   cancer of the endometrium. Cancer 19:489-520, 1966
10. MacMahon B: Risk factors for endometrial cancer. Gynecol Oncol
    2:122-129, 1974
11. MacDonald PC, Siiteri PK: The relationship between the extraglandular
    production of estrone and the occurrence of endometrial neoplasia. Gy-
    necol Oncol 2:259-263, 1974
12. Meissner WA, Sommers SC, Sherman G: Endometrial hyperplasia, en-
    dometrial carcinoma, and endometriosis produced experimentally by es-
    trogen. Cancer 10:500-509, 1957
13. Gusberg SB: Precursors of corpus carcinoma estrogens and adenoma-
    tous hyperplasia. Am J Obstet Gynecol 54:905-927, 1947
14. Gusberg SB, Hall RE: Precursors of corpus cancer. III. The appearance
    of cancer of the endometrium in estrogenically conditioned patients.
    Obstet Gynecol 17:397-412, 1961
15. Gusberg SB, Kaplan AL: Precursors of corpus cancer. IV. Adenomatous
    hyperplasia as stage O carcinoma of the endometrium. Am J Obstet
    Gynecol 87:662-678, 1963
16. Jensen EI, Østergaard E: Clinical studies concerning the relationship of
    estrogens to the development of cancer of the corpus uteri. Am J Obstet
    Gynecol 67:1094-1102, 1954
17. Smith DC, Prentice R, Thompson DJ, et al: Association of exogenous
    estrogen and endometrial carcinoma. N Engl J Med 293:1164-1167,
    1975
18. Ziel HK, Finkle WD: Increased risk of endometrial carcinoma
    among users of conjugated estrogens. N Engl J Med 293:1167-1170,
    1975
19. Mack TM, Pike MC, Henderson BE, et al: Estrogens and endome-
    trial cancer in a retirement community. N Engl J Med 294:1262-1267,
    1976
20. Cutler BS, Forbes AP, Ingersoll FM, et al: Endometrial carcinoma after
    stilbestrol therapy in gonadal dysgenesis. N Engl J Med 287:628-631,
    1972
21. Pharmaceutical Preparations, Except Biologicals, 1973 (US Bureau of
    the Census Current Industrial Reports, Series Ma-28G[73]-1). Washing-
    ton, DC, Government Printing Office, 1975
22. Stadel BV, Weiss N: Characteristics of menopausal women: a survey of
    King and Pierce counties in Washington, 1973-1974. Am J Epidemiol
    102:209-216, 1975

# ESTROGENS AND ENDOMETRIAL CANCER IN A RETIREMENT COMMUNITY

Thomas M. Mack, M.D., M.P.H., Malcolm C. Pike, Ph.D., Brian E. Henderson, M.D.,
Robert I. Pfeffer, M.D., Vibeke R. Gerkins, R.N., M.P.H., Mary Arthur, B.S.,
and Sandra E. Brown, B.S.

**Abstract** All cases of endometrial cancer occurring among the residents of an affluent retirement community were compared with controls chosen from a roster of all women in the same community. Evidence of estrogen and other drug use and of selected medical conditions was obtained from three sources: medical records of the principal care facility, interviews, and the records of the local pharmacy. The risk ratio for any estrogen use was estimated from all available evidence to be 8.0 (95 per cent confidence interval, 3.5 to 18.1), and that for conjugated estrogen use to be 5.6 (95 per cent confidence interval, 2.8 to 11.1). Increased risk from estrogens was shown for invasive as well as noninvasive cancer, and a dose-response effect was demonstrated. For an estrogen user, the risk from endometrial cancer appeared to exceed by far the base-line risk from any other single cancer. ( N Engl J Med 294:1262-1267, 1976)

T HE results of two recent reports,[1,2] each based on data from individual cases and controls, are compatible with a causal link between conjugated-estrogen use and endometrial carcinoma. A third report affirms the credibility of this link in terms of chronology and geography.[3]

We now report an attempt to confirm these findings using various measures of drug use, measuring the invasiveness of the endometrial cancer, and choosing both cases and controls from the entire population of a retirement community of uniformly high affluence.

## MATERIALS AND METHODS

The population studied consists of a residential retirement community located near Los Angeles. In December, 1975, its residents numbered more than 18,000. Nearly all are white and relatively affluent, since expensive dwellings must be purchased. At least one member of each resident family must have attained the age of 52 years; the median age is about 70. Residents are provided with a closed community, and a single comprehensive medical-care facility

is used as the major source of care by about 85 per cent of them. A pharmacy is located within the medical complex, and has for a decade been the preponderant source of medications.[4]

Newcomers to the community routinely receive, from the staff of the care system, an entry questionnaire with general medical questions. More than 80 per cent of these questionnaires are completed and returned within a few months of entry. Upon receipt of this form or when health services are first requested, a medical chart is prepared.

We performed a survey in this population in 1971[5] and have collected information about patients with cancer from pathology records since that time. Notification comes either from the local facility, from surrounding hospitals, from our program of cancer surveillance in Los Angeles County[6] or from routine review of all Orange and Los Angeles County death certificates. Approximately 39,000 female person-years had been under surveillance as of November 15, 1975, and 65 newly diagnosed cases of adenocarcinoma of the endometrium had been collected by that date. One of these cancers appeared to be metastatic from another site, and one appeared in a woman who entered the community after the preparation of the most recent list of residents. These two cases were excluded, leaving 63 cases for study.

For each case, we chose controls from a roster of the community prepared in sequence of identifying number (usually but not always assigned sequentially at entry) by listing, in order of proximity, the four women whose names preceded and followed the name of the patient, who were matched to the patient within one year of birth date, who were of the same marital status (ever-married or single), and who were alive and living in the community at the time of diagnosis of the patient's disease. Replacement names were chosen for the 134 controls (representing about 37 per cent of the popula-

tion) who turned out to have had a hysterectomy before the date of diagnosis in the case, and for the 10 controls who both lacked a clinic record and were unavailable for (or refused) interview. In all, 396 control names were chosen.

Using current and back editions of the *Physicians' Desk Reference*,[7] we prepared lists of systemic medications under the headings of conjugated estrogens, other estrogens, rauwolfia preparations, thiazides, other antihypertensive drugs,[8] barbiturates, phenothiazines, propanediol carbamates, benzodiazepines, and miscellaneous antianxiety preparations (chlormezanone, tricyclic antidepressants and sedative antihistamines). Each chart was searched for mention of any of the listed drugs, and the date of first use, the date of first chart note, and the best (minimum) estimate of duration of use was recorded. The average dose of the conjugated estrogen preparations was also noted. Although detailed indications for estrogen therapy were not always available, most if not all women were post-menopausal at time of entry into the community.

Other abstracted items included dates of birth, menarche, marriage, first pregnancy, first full-term pregnancy, menopause, and death, number of full-term pregnancies, height, usual and current weight, and any mention of obesity, past pelvic irradiation, hysterectomy, gallbladder disease, diabetes, hypertension, benign breast disease or past diagnosis of cancer. When obesity was not specifically mentioned, it was inferred when the usual weight was in excess of a standard upper limit for height in medium-frame women.[9] For the cases, the chief complaint and duration of the presenting illness, and the grade, stage and degree of invasiveness of endometrial cancer were noted. Clinic charts were available and abstracted for 55 cases and 191 controls. For one patient and two controls who had died without a clinic record, the above information was obtained from the attending physician by telephone.

Permission for a telephone interview was solicited from each living patient and control. A few subjects expressed a preference for personal interviews and were accommodated. Those without available telephone numbers were visited at their homes and, when possible, interviewed. The questions covered the above-mentioned milestone dates, operations and diagnoses, height, usual and current weight, usual regularity and volume of menstrual flow, symptoms of menopause, birthplace, education and employment history, and the periods of use of "female hormones" (specifying type and method of administration), other hormones (steroids and thyroid), "prescription sleeping pills" and "tranquilizers." The last interview took place in November, 1975, before the possible relation between estrogen use and endometrial cancer had been publicized. Interviews were obtained from 46 patients and 189 controls, making abstracts or interviews (or both) available for all. For 39 cases and 70 of their matched controls who resided in the community in April, 1972, survey items, including sibship size and family history of cancer, were available.[5]

Finally, pharmacy records for the years 1964 through 1975 had been screened for other purposes,[4] and all prescriptions for estrogens and oral hypoglycemic drugs abstracted for name and identifying number, date, type of drug, pattern of administration, number of pills, and refill dates and numbers of pills. From these data the prescription history of each case and control was obtained. Mean dose (mean size of pill, not mean daily dose), duration of use, interval since last use and principal method of administration (number of drug-free days in sequence at specified intervals) were noted or computed from the prescription information.

Exposures evidenced only by notes written less than six months before the date of diagnosis in the case and in the same period for the matched controls were ignored. (In only two patients were symptoms present for as long as six months.) Unless otherwise specified, the matched sets were retained in the analysis, and risk ratios (relative risks) were computed in the usual way[10] with 95 per cent confidence limits.

## RESULTS

As previously described, rates of drug use in this population are high.[8] Of the women who used conjugated estro-

gens, over 90 per cent used preparations without methyltestosterone, and only a very few women had taken progesterone. A variety of estrogen preparations other than conjugated estrogens were specified and placed in a separate category. They included diethylstilbestrol, ethinyl estradiol, estradiol, esterified estrogens, chlorotrianisene and a few unidentifiable parenteral estrogen preparations. The distribution of preparations within the categories of rauwolfia, thiazide, other antihypertensive drug and barbiturate was as previously reported.[8] When all nonbarbiturate sedatives were lumped, they were termed "tranquilizers." Thyroid was usually taken in the form of thyroid hormone. Adrenocortical steroids were usually glucosteroids taken in short courses by injection. Oral hypoglycemic drugs included both sulfonylureas and biguanides.

### Estrogen Use

No matter which source of drug information was chosen, conjugated estrogen use was strongly associated with endometrial carcinoma (Table 1). When all sources were pooled, the association was very strong.

Use of other estrogens was also associated with endometrial carcinoma. When this use was examined in the cases and matched controls happening to be concordant for use of conjugated estrogens, the risk ratio for other estrogens remained high among both conjugated-estrogen users and, especially, nonusers. No single preparation predominated among these cases; diethylstilbestrol and ethinyl estradiol occurred with about equal frequency and accounted for the majority of other estrogens specified. The risk ratio for any estrogen use was higher and more incompatible with chance than that for conjugated estrogens alone.[*]

### Use of Other Drugs

No other drug preparations were strongly, or significantly, associated with the disease (Table 2). The risk-ratio point estimates for barbiturates and rauwolfia preparations were slightly high, and that for any use of non-estrogen drugs was quite high at 3.9. From previous work[8] we expected that this measure of medical services consumption would be associated with estrogen use. When estrogen users only were examined, the high risk ratio for any non-estrogen drug use was markedly reduced; the data were insufficient to permit stratification on dose as well.

### Other Risk Factors

Gallbladder disease was also significantly associated with endometrial carcinoma (Table 2). Although the frequencies were small, the association appeared to remain in the estrogen-using patients with their estrogen-using matched con-

---

[*]If the 10 controls who were excluded because neither abstracts nor interviews were available are all assumed to have taken estrogen (none had obtained prescriptions at the pharmacy) and retained in the analysis, only one control in each of two matched sets converts from negative to positive, and risk-ratio estimates are essentially unaffected. If the converse assumption is made, all risk ratios increase slightly.

Similarly, if the fact of hysterectomy in potential controls is ignored, and replacements not made on that basis, the risk-ratio estimates based on pharmacy records increase, indicating that fewer of the post-hysterectomy controls had used estrogens than their age-matched replacements.

**Table 1. Risk Ratios for Extrogen Use for Patients with Endometrial Carcinoma and Matched Controls.**

| Drug | Exposure Source* | Matched Cases/ Controls | Frequency in Controls | Risk Ratio | 95% Confidence Limits |
|---|---|---|---|---|---|
| **Conjugated estrogens:** | | | | | |
| Entire group | A | 63/252 | 0.29 | 4.1 | 2.2-7.7 |
| Interviewed† | I | 46/148 | 0.40 | 2.8 | 1.4-5.5 |
| Entire group | P | 63/252 | 0.25 | 2.6 | 1.4-5.1 |
| Entire group | AIP | 63/252 | 0.43 | 5.6 | 2.8-11.1 |
| **Other estrogens:** | | | | | |
| Conjugated-estrogen users† | AIP | 47/90 | 0.26 | 2.5 | 1.0-6.0 |
| Conjugated-estrogen nonusers† | AIP | 11/29 | 0.07 | 11.6 | 1.3-103.3 |
| Entire group | AIP | 63/252 | 0.18 | 3.3 | 1.7-6.4 |
| **Any estrogens:** | | | | | |
| Entire group | AIP | 63/252 | 0.50 | 8.0 | 3.5-18.1 |
| Obese† | AIP | 38/91 | 0.62 | 4.5 | 1.5-13.5 |
| Nonobese† | AIP | 15/29 | 0.66 | ∞ | 1.5-∞ |
| Nulliparous† | AIP | 15/27 | 0.59 | 2.3 | 0.6-9.6 |
| Parous† | AIP | 38/108 | 0.59 | 22.0 | 2.9-165.4 |
| Gallbladder disease† | AIP | 7/9 | 0.67 | 3.0‡ | 0.2-183.3 |
| | | 17/24§ | 0.67 | 1.6§ | 0.3-9.0 |
| No gallbladder disease† | AIP | 46/169 | 0.50 | 14.1 | 1.2-46.9 |
| Hot flashes† | AIP | 6/7 | 0.55 | ∞‡ | 1.0-∞ |
| | | 16/38§ | 0.55 | ∞§ | 3.3-∞ |
| No hot flashes† | AIP | 41/129 | 0.47 | 6.1 | 2.4-15.2 |

*A denotes abstract, I interview, P pharmacy, & S survey; combinations signify 1 or more.

†Cases & variable no. of concordant matched controls.

‡Unmatched analysis on cases with matched controls (matched analysis precluded by low frequencies).

§Unmatched analysis on all cases & controls.

trols (gallbladder disease preceded estrogen use in most instances) and in those who did not take estrogens with their estrogen-free matched controls.

There was an association with vasomotor symptoms at menopause ("hot flashes") that also appeared to be independent of the risk from estrogens. Although not statistically significant, a similar association was found with a positive family history for cancer. Risk ratios for obesity, nulliparity, age at menopause and menarche, hypertension and diabetes were all below 2.0.

Characteristics of menstrual flow, benign breast disease, age at marriage, age at first pregnancy, education, sibship size and occupation were not associated with endometrial cancer. Few instances of prior cancer or pelvic irradiation were found.

### Modes of Estrogen Use

The effects of conjugated estrogen dose and duration on risk were evaluated (Table 3). Because of small numbers, it was necessary that this calculation be done by an unmatched analysis (the results of all the foregoing analyses were essentially unchanged when the matching was ignored). Much higher risks followed high dose within all but the shortest duration category (summary chi-square test for the three long duration categories, P = 0.01). Risk ratios always exceeded 3.0, even for low doses.

Data from the pharmacy permitted evaluation of various modes of administration. At smaller doses, administration in monthly cycles with drug-free intervals of four or more days appeared to decrease the risk, whereas no such effect

was apparent at larger doses. The same data analyzed in groups according to the duration of drug-free period before diagnosis in the case showed a consistent, slight, decrease in risk after an interval of two or more years. The incompleteness of the pharmacy record can be roughly taken into account if each of these risks is multiplied by a factor of 2.2 (5.6/2.6 from Table 1). Thus, for low doses, the best estimate of relative risk after cyclic administration is 1.8, and after a two year drug-free period it is 3.1.

### Risk Modification

Because the effect of dose was strong, and because we could not finely categorize dose, the identification of other factors that modify the risk from estrogens must be tentative. Moreover, since 56 of the 63 patients had taken estrogens, there was limited ability to distinguish statistically the factors that influenced the size of the risk ratio. A pattern was visible, however. Very high risk ratios (above 13.0) were seen in the cases and their matched controls that happened to be concordant for any of a number of characteris-

**Table 2. Risk Ratios for Use of Selected Drugs and Conditions for Patients with Endometrial Carcinoma and Matched Controls.**

| Exposure | Exposure Source* | Entire Group | | | Matched Users of Any Estrogens | | |
|---|---|---|---|---|---|---|---|
| | | Matched Cases/ Controls | Frequency in Controls | Risk Ratio | Matched Cases/ Controls | Frequency in Controls | Risk Ratio |
| Rauwolfia | A | 63/252 | 0.13 | 1.9 | | | |
| Barbiturates | AI | 63/252 | 0.25 | 1.8 | | | |
| Propanediol carbamates | A | 63/252 | 0.17 | 1.6 | | | |
| Benzodiazepines | A | 63/252 | 0.22 | 0.9 | | | |
| Any "tranquilizers" | AI | 63/252 | 0.50 | 1.5 | | | |
| Thyroid | I | 46/148 | 0.30 | 1.7 | | | |
| Adrenocortical steroids | I | 46/148 | 0.28 | 1.1 | | | |
| Oral hypoglycemic drugs | P | 63/252 | 0.02 | 1.4 | | | |
| Any non-estrogen drug use | AIP | 63/252 | 0.70 | 3.9† | 53/114 | 0.87 | 1.6 |
| Gallbladder disease | AI | 63/252 | 0.10 | 3.7† | 53/114 | 0.13 | 2.3 |
| Hot flashes | I | 39/109 | 0.20 | 2.7† | 29/55 | 0.22 | 2.3 |
| Obesity | AI | 57/192 | 0.63 | 1.5 | 48/102 | 0.58 | 1.6 |
| Hypertension | A | 63/252 | 0.33 | 1.5 | 53/114 | 0.46 | 0.8 |
| Nulliparity | AI | 61/215 | 0.27 | 1.4 | 52/109 | 0.24 | 1.4 |
| Menopause at >50 yr | AI | 50/147 | 0.54 | 1.1 | 40/76 | 0.54 | 1.0 |
| Menarche at <13 yr | AI | 42/116 | 0.37 | 1.0 | 34/62 | 0.37 | 1.2 |
| Diabetes | AIP | 63/252 | 0.09 | 0.9 | 53/114 | 0.11 | 0.6 |
| Family history of cancer | S | 29/70 | 0.31 | 1.6 | 24/37 | 0.27 | 2.0 |
| Family history of breast, ovary, endometrial cancer | S | 29/70 | 0.06 | 1.6 | 24/37 | 0.03 | 3.2 |

*A denotes abstract, I interview, P pharmacy, & S survey; combinations signify 1 or more. Small variations in no. reflect information about normal characteristics.

†95% confidence limits exclude 1.0.

Table 3. Risk Ratios (RR) for Various Doses, Durations and Modes of Administration of Conjugated Estrogen for Patients with Endometrial Carcinoma and (Unmatched) Controls.

| | MEAN DOSE | | | | | |
| | <0.625 MG | | >0.625 MG | | TOTALS† | |
| Duration of Known Use (All Sources) | cases/ con- trols | RR* | cases/ con- trols | RR* | cases/ con- trols | RR* |
|---|---|---|---|---|---|---|
| 1-11 mo | 5/9 | 6.6 | 0/8 | 0.0 | 6/26 | 2.8 |
| 12-59 mo | 8/28 | 3.4 | 5/2 | 29.8 | 15/40 | 4.5 |
| 60-95 mo | 3/6 | 6.0 | 2/2 | 11.9 | 7/9 | 9.3 |
| 96+ mo | 4/10 | 4.8 | 8/5 | 19.1 | 17/23 | 8.8 |
| Total† | 23/55 | 5.0 | 15/19 | 9.4 | 51/109 | 5.6 |
| Mode of Use (Pharmacy) | cases/ con- trols | RR‡ | cases/ con- trols | RR‡ | cases/ con- trols | RR‡ |
| Free interval, 0-1 days | 5/14 | 1.9 | 2/5 | 2.1 | 7/19 | 4.4 |
| Free interval, 2-3 days | 7/19 | 1.9 | 6/5 | 6.3 | 13/24 | 6.5 |
| Free interval, 4+ days | 2/13 | 0.8 | 3/5 | 3.2 | 5/18 | 3.3 |
| Estrogen-Free Interval before Diagnosis | cases/ con- trols | RR‡ | cases/ con- trols | RR‡ | cases/ con- trols | RR‡ |
| 0-23 mo | 9/24 | 2.0 | 9/6 | 7.9 | 18/30 | 7.2 |
| 24+ mo | 6/23 | 1.4 | 3/9 | 1.8 | 9/32 | 3.4 |

*Maximum likelihood estimate as compared with nonusers: cases/controls = 12/143.

†Including those for which dose or duration (or both) was not known.

‡Maximum likelihood estimate compared with nonusers by pharmacy records: cases/controls = 36/190.

tics, most of which are known to be closely related: married before age 25, currently married, parous, non-obese, without gallbladder disease, without tranquilizer use, with "hot flashes" and with benign breast disease. Conversely, lower risk ratios (below 3.0) for any estrogen use appeared among women who were married after the age of 25, highly educated, nulliparous, with menopause after the age of 50, and consumers of tranquilizers and of thyroid preparations. In the obese and in those reporting no "hot flashes" at menopause, the risk ratios were lower than in the entire group, but still substantially and significantly elevated above 1.0. The strong association with any estrogens was true throughout the limited age range in the community (risk ratios of 9.3, 6.3 and 9.1 for <70, 70-74 and 75+ years respectively).

### Histology, Stage and Clinical Presentation

A description of the histology was available for all tumors (Table 4), and the degree of invasiveness could be documented in 61. Three (4.8 per cent) were called adenoacanthomas; squamous elements were described in four others (6.3 per cent), and one (1.6 per cent) was described as "carcinosarcomatous." Eighteen (28.6 per cent) were described as "Grade III" or "poorly differentiated." Forty-one (67.2 per cent) were invasive — 26 (42.6 per cent) beyond the proximal third of the myometrium. This distribution of stage and grade is comparable to that in reported series.[11]

Risk ratios were high for poorly differentiated and for invasive tumors, although not as high as for well differentiated and noninvasive tumors. Early death did not modify the risk.

The presenting illness in almost all cases consisted of bleeding, and the duration of symptoms was described in 38 cases (60.3 per cent). Most of the remaining women were being followed regularly and can therefore be assumed to have had symptoms for only a short time. Risks for estrogen use appeared very high in the latter group and in those with symptomatic periods known to be short, whatever the histologic or clinical prognosis.

### DISCUSSION

This study supports, at a high level of statistical significance, the hypothesis that exogenous estrogens cause endometrial cancer. Since the line between hypothesis and fact is especially hard to identify on the basis of purely observational data, each alternative explanation for the findings should be explicitly considered.

### Case Selection

Misleading conclusions from case-control studies may be inherent in the selection of cases, especially if selection is not directly based on a standard diagnostic process. In the context of cancer, problems may also result from confusion of hyperplasia with neoplasia, or inclusion of noninvasive lesions found by a cytologic screening survey or routine section of autopsy specimens.

Endometrial cancer must be diagnosed by biopsy or autopsy. In the present study, the cases include all those and only those newly diagnosed from a specific population and

Table 4. Risk Ratios (RR) for Estrogen Use, According to Characteristics of Disease, for Patients with Endometrial Carcinoma and Matched Controls.

| CHARACTERISTIC | CASES/ CONTROLS | FREQUENCY IN CONTROLS | RR | 95% CONFIDENCE LIMITS |
|---|---|---|---|---|
| Noninvasive | 20/80 | 0.53 | ∞ | 3.9-∞ |
| Invasive: | 41/164 | 0.49 | 5.0 | 2.1-11.8 |
| To proximal ⅓ of myometrium | 15/60 | 0.50 | 14.5 | 1.8-118.8 |
| To middle or distal ⅓ of myometrium | 14/56 | 0.48 | 3.8 | 1.0-14.9 |
| Direct extension or metastasis | 12/48 | 0.50 | 3.0 | 0.7-12.4 |
| Grades I & II | 45/180 | 0.48 | 12.9 | 3.9-43.0 |
| Grade III | 18/72 | 0.53 | 4.0 | 1.2-13.5 |
| Decedents | 9/36 | 0.55 | 9.0 | 1.0-77.4 |
| Survivors | 54/216 | 0.49 | 7.8 | 3.2-19.0 |
| Duration of symptoms, >1 mo: | | | | |
| Grades I & II, & noninvasive | 6/24 | 0.58 | ∞ | 0.7-∞ |
| Grade III or deep invasive (or both) | 10/40 | 0.50 | 1.5 | 0.4-6.1 |
| Totals | 16/64 | 0.53 | 2.7 | 0.8-9.5 |
| Duration of symptoms, other: | | | | |
| Grades I & II, & noninvasive | 25/100 | 0.51 | ∞ | 5.6-∞ |
| Grade III or deep invasive (or both) | 20/80 | 0.48 | 6.5 | 1.7-24.1 |
| Totals | 45/180 | 0.51 | 14.4 | 4.3-48.2 |

time span, whether or not these women had been hospitalized, or even under care. Diagnostic conventions were constant over the interval, the incidence rates were high, and it is not probable that patients ranged sufficiently far afield to be missed. None of the tumors were incidental pathological findings. Only two of the patients were known to have had cancer diagnosed after an abnormal "routine" cytologic test, and one of those tumors had invaded the myometrium. The distribution of cases according to grade and stage is consistent with published series. Finally, the risk ratios for advanced stages and grades were consistently high. They were higher for noninvasive disease but also for invasive disease when biopsy was done soon after the onset of symptoms.

## Control Selection

In case-control studies the choice of controls having a specific diagnosis, or even having various diagnoses but predictably different patterns of interaction with the medical-care system, can produce spuriously positive results.[8] When healthy persons are chosen as controls, selective non-co-operation may still result in unreal findings.

Our controls were chosen from the roster of all persons at risk of being diagnosed. They were explicitly comparable to the patients in age, socioeconomic status and ease of access to medical care. Although there were some women for whom information was not available, the maximum possible impact was negligible.

## Measures of Drug Consumption

Although recall bias usually results in overestimation, our risk ratio as calculated from the interview alone appears to be a biased underestimate. When known from other sources to use estrogens, a larger proportion of patients than of controls gave negative (including "can't recall") histories. Since this lack of recall was true for many items, it seems to have been due to the tendency of some patients to tire easily.

We minimized the effect of the clinical illness on the abstraction process by ignoring all notes, on cases or controls, made within six months of diagnosis of the case. Moreover, high risks were found in long-term users for whom the records are most comparable. These facts, in addition to the standardized nature of the records, and the absence of a similar bias with respect to either the measures of other drug use in endometrial cancer cases or, in a previous study,[8] the measures of estrogen and other drug use in cases of breast cancer, would seem to make overzealous abstraction an unlikely explanation for such a strong association. In fact, the most accurate estimate of the true risk ratio is probably that derived from the pooling of all sources of drug information; each of the individual sources is demonstrably insensitive, and gives a low estimate of the risk ratio. Only the combined action of biases in all three sources could spuriously produce this effect.

## Causality of the Association

Bradford Hill,[12] when enumerating the features that enable one to judge the causality of associations seen in observational studies, lists strength, specificity, a dose-response effect, a sequence of events compatible with causality, bio-logic plausibility and coherence, and consistency with other studies.

The strength of the association found here is large. It appears in all subgroups, with all measures of drug use, and with pathological criteria of increasing specificity.

Drug-use patterns in the controls reproduce those found previously in the same population,[8] and no single other drug was strongly or significantly associated with the disease. Low associations between non-estrogen drugs and endometrial cancer are consistent with a propensity on the part of estrogen users to take other drugs.

Increasing risk with larger dose is consistently apparent within each duration-of-use category except the lowest, although no combination of dose and duration appears safe. The sequence from exposure to disease is clear. There is diminution in risk with the passage of time free of exposure.

On the basis of clinical and laboratory work,[13-16] estrogens are a credible component in the causation of endometrial carcinoma, and the risk-modifying effect of obesity and nulliparity is consistent with at least some overlap between the etiologic effects of exogenous and endogenous estrogens. The next most important risk factor for endometrial carcinoma that could be demonstrated was gallbladder disease. The latter finding was true even in the absence of estrogen use, was unexpected, and explains none of the association with exogenous estrogens. None of these patients are known to have had functional liver disease, and the gallbladder disease usually preceded the first use of estrogens. Associations have been observed between use of oral contraceptives and the composition of human bile[17] and between use of both contraceptive[18] and menopausal[19] estrogens and the incidence of gallbladder disease. That this condition is a risk modifier for the risk from estrogens also therefore suggests a relation between endogenous estrogen production and endometrial cancer. So does the fact that menopausal "hot flashes" seem both to modify the risk from estrogens and to act as a risk factor independent of estrogens.

The crude annual incidence of endometrial cancer in this community was about $152/10^5$. That is 2.1 times the Third National Cancer Survey rate[20] for comparable* white women (rates of cancer of the cervix, breast, and colon in the community are 0.5, 1.4, and 1.0 times the equivalent Third National Cancer Survey rates), and 1.3 times the equivalent Los Angeles County rate for non-Spanish-surnamed white women in the same period. When the denominator is decreased by the 37 per cent of the women in the community who had no uterine corpus, the rate becomes $241/10^5$. Assuming our community to be representative, the approximate incidence in nonusers of estrogens can be calculated to be about $53/10^5$, and that in users about $424/10^5$. The difference, $371/10^5$, represents the attributable risk of endometrial cancer, the extra burden shouldered by a woman who takes estrogens. For a white woman 65 to 74 years of age that figure is about one-third the entire Third National Cancer Survey cancer risk for all sites, and approximates by itself the combined base-line risk from cancer of the breast, cervix, lung and stomach.

Because it is often difficult to distinguish between the ef-

*Adjusted to the age distribution of the study community.

fects of a treatment and those of the condition for which it is given, the high incidence rate of endometrial cancer in the community is of special interest. There is no reason to believe that the prevalence of specific menopausal symptoms in the community is unusually high, whereas the prevalence of treatment clearly is. Moreover, estrogens provoked a high risk whether or not the most specific symptom, "hot flashes," had occurred.

We attribute our failure to confirm unequivocally such previously identified risk factors as obesity, diabetes, nulliparity and age at menopause[21] to the great preponderance of estrogen-associated cases. Except for that failure, our findings are consistent with previous epidemiologic studies, including the two recent case-control studies.[1,2]

### Recommendations

In the face of a large, consistent and specific risk ratio, a credible sequence, biologic plausibility in the face of information from the laboratory, the bedside and the community, and consistency with other epidemiologic studies, prudence dictates that we tentatively assume the association to be causal, and act on that basis.

However, estrogen use should not be compared to an optional threat like smoking, but to the use of valuable but potentially dangerous tools like insulin and digitalis. The benefits of estrogen use must be carefully measured, and the search continued for predictors of endometrial cancer in estrogen users. In this context cohort studies will be extremely valuable. When estrogens are indicated, they should be given at the lowest effective dose for the shortest possible time. No estrogen can at present be singled out as safe, but many clinicians advocate cyclic rather than continuous estrogen therapy, in the belief that a monthly respite, even in the absence of progesterone and endometrial sloughing, will exert some ameliorating effect.[22] At low doses, this expectation is consistent with our data. The nonobese parous women usually considered to be at low risk are those seemingly most endangered by estrogens, although from these data no group can be treated with impunity.

### Outlook

Estimates of the annual numbers of new drug prescriptions according to category and preparation are commissioned regularly by the industry and obtained by pharmacy sampling. Annual estimates for the category of estrogens were made available.* In 1958 about 1.6 million new prescriptions for the most popular conjugated-estrogen preparation were written. This figure changed little until 1965, but by 1966, it had doubled. Another period of stability ended in 1971, and in 1974 about five million new prescriptions were issued. This trend is thought to reflect nationwide consumption and must therefore be reflected, after a latent period of unknown length, by a trend in the incidence of any disease caused by estrogens. Weiss'[3] has concluded that between 1969 and 1973, American women experienced an increase in the incidence of endometrial cancer. The magnitude of the increase was summarized according to age and region by risk ratios, most of which ex-

*Courtesy of IMS, Limited, July, 1975.

ceeded 1.4. One can examine the compatibility between the magnitude of this increase and the magnitude of the trend in new prescriptions over four-year intervals, using various assumptions about the average latent period between the beginning of exposure and the diagnosis of cancer. An average latent period of anywhere from four to eight years appears most compatible with the observed increase in incidence. Such a latent period is consistent with the appearance of risk after a few years of use and the failure of the risk ratio to return to unity after a drug-free interval of two or more years. If the association is causal, if the new drug prescription trends are reliable, and if the mean latent period is four years or more, one can expect still another increase in the incidence of endometrial cancer corresponding to the increase in new prescriptions over the period 1970-1974, whether or not a few years of stable rates precede it, and probably whether or not estrogen consumption declines, as expected, in the next few years. This increase in incidence can be expected to continue past 1980.

### REFERENCES

1. Smith DC, Prentice R, Thompson DJ, et al: Association of exogenous estrogen and endometrial carcinoma. N Engl J Med 293:1164-1167, 1975
2. Ziel HK, Finkle WD: Increased risk of endometrial carcinoma among users of conjugated estrogens. N Engl J Med 293:1167-1170, 1975
3. Weiss NS, Szekely DR, Austin DF: Increasing incidence of endometrial cancer in the United States. N Engl J Med 294:1259-1262, 1976
4. Pfeffer RI, Van Den Noort S: Estrogen use and stroke risk in postmenopausal women. Am J Epidemiol (in press)
5. Henderson BE, Bogdanoff E, Gerkins VR, et al: Evaluation of cancer risk factors in a retirement community. Cancer Res 34:1045-1048, 1974
6. Hisserich JC, Martin SP, Henderson BE: An areawide cancer reporting network. Public Health Rep 90:15-17, 1975
7. Physicians' Desk Reference. Twenty fourth and twenty seventh editions. Edited by BB Huff. Oradell, New Jersey, Medical Economics Company, 1970, 1973
8. Mack TM, Henderson BE, Gerkins VR, et al: Reserpine and breast cancer in a retirement community. N Engl J Med 292:1366-1371, 1975
9. Fifty years of life conservation. Stat Bull Metropol Life Ins Co 40:1-4, 1959
10. Miettinen OS: Estimation of relative risk from individually matched series. Biometrics 26:75-86, 1970
11. DiSaia PJ, Morrow CP, Townsend DE: Synopsis of Gynecologic Oncology. New York, John Wiley and Sons, 1975
12. Hill AB: Principles of Medical Statistics, Ninth edition. New York, Oxford University Press, 1971
13. MacDonald PC, Siiteri PK: The relationship between the extraglandular production of estrone and the occurrence of endometrial neoplasia. Gynecol Oncol 2:259-263, 1974
14. Siiteri PK, Schwarz BE, MacDonald PC: Estrogen receptors and the estrone hypothesis in relation to endometrial and breast cancer. Gynecol Oncol 2:228-238, 1974
15. Silverberg SG, Makowski EL: Endometrial carcinoma in young women taking oral contraceptive agents. Obstet Gynecol 46:503-506, 1975
16. Cutler BS, Forbes AP, Ingersoll FM, et al: Endometrial carcinoma after stilbestrol therapy in gonadal dysgenesis. N Engl J Med 287:628-631, 1972
17. Royal College of General Practitioners, Oral Contraception Study: Oral Contraceptives and Health: An interim report. New York, Pitman Medical Publishing Company, 1974
18. Bennion LJ, Ginsberg RL, Garnick MB, et al: Effects of oral contraceptives on the gallbladder bile of normal women. N Engl J Med 294:189-192, 1976
19. Boston Collaborative Drug Surveillance Program: Surgically confirmed gallbladder disease, venous thromboembolism, and breast tumors in relation to postmenopausal estrogen therapy. N Engl J Med 290:15-19, 1974
20. Biometry Branch, National Cancer Institute: Third National Cancer Survey: Incidence data. Natl Cancer Inst Monogr 41, 1975
21. MacMahon B: Risk factors for endometrial cancer. Gynecol Oncol 2: 122-129, 1974
22. The Pharmacologic Basis of Therapeutics, Fifth edition. Edited by LS Goodman, A Gilman. New York, Macmillan Publishing Company, 1975

v. 9, 1978

E

im on com-
nd children,
'lorida.
ger, Depart-
uffalo, New

SURGERY

1g in Pediat-

1 Niño DIF,

Cesar Beach
.
iles, Tampa

SURGICAL

1ncies" is the
, January 25-

. Thompson,

JRED

e Center will
ured patient,

f Continuing
on, P.O. Box
i671.

AGEMENT

ty of Clinical
ions in Infec-
Arizona, on

d, 810 West
: to (602) 246-

JE

sity School of
Vhole Person
: University.

line, Office of
South Lewis,

nsiveness to
us Leukemia
inder heading
ed only with
mg, given by

# The New England
# Journal of Medicine

## ALTERNATIVE ANALYTIC METHODS FOR CASE-CONTROL STUDIES OF ESTROGENS AND ENDOMETRIAL CANCER

RALPH I. HORWITZ, M.D., AND ALVAN R. FEINSTEIN, M.D.

**Abstract** In a case-control study of estrogens and endometrial cancer, alternative sampling methods were used to eliminate the detection bias that arises from the increased diagnostic attention received by women with uterine bleeding after estrogen exposure. In a set of cases and controls chosen by conventional procedures the odds ratio was 11.98. In an alternative set of cases and controls at the same institution, consisting of patients who had all received dilatation and curettage or hysterectomy because of uterine bleeding, the odds ratio was 1.7.

A methodologic analysis demonstrates detection bias arising from the pattern of hospital referral and shows the way in which the bias is neglected or increased by conventional sampling procedures, but reduced by the alternative procedure. The magnitude of the association between estrogens and endometrial cancer has been greatly overestimated because of detection bias; when an appropriate compensation for the bias is introduced, the odds ratio approaches a value much closer to 1. (N Engl J Med 299:1089-1094, 1978)

AN association between exogenous estrogens and endometrial cancer has been reported for postmenopausal women in five recent investigations[1-5] that all employed the conventional methods of the "retrospective case-control" study.

In a case-control study the investigator works in a hospital or other medical setting and follows people backward from the effect toward the cause. In that setting he collects a group of the diseased people called "cases," and from the available nondiseased people he chooses a separate group called "controls." He then finds out which people were exposed or not exposed to the alleged causal agent. With this information, the patients are divided into four groups, whose numbers are counted as a, b, c and d, as shown in Figure 1.

The odds ratio used to express the results of a case-control study is obtained from these numbers as a ratio of two ratios: the exposure ratio in the cases, divided by the exposure ratio in the controls $(a/c \div b/d)$. The *odds ratio* is used as a substitute for the *risk ratio*, which is the rate with which the disease occurs in exposed people, divided by the rate of the disease's occurrence in nonexposed people. If these two rates of occurrence are very small and if no distor-

tions have occurred in the four groups that comprise the case-control study, the odds ratio will be approximately equal to the risk ratio. If the odds ratio exceeds 1 and is statistically significant, either in a direct test of significance or by demonstration that the value of 1 is not contained in the associated confidence interval, the investigator concludes that a causal relation may exist between the agent and the disease.

With the conventional methods, the results of the five recent case-control investigations of estrogens and endometrial cancer produced odds ratios of 7.5, 7.6, 8.0, 4.9 and 3.1, respectively. These data were in conflict with two older case-control studies[6,7] in which the respective odds ratios were 1.1 and 0.5. The two older studies had been performed with a different method of patient selection in which the population under investigation consisted of women with postmenopausal bleeding receiving dilatation and curettage. The cases and controls were determined according to the results of that procedure.

In contemplating the discrepancy in results between the older and the more recent studies, we wondered about the role of estrogens in provoking bleeding as an adverse side effect. With an asymptomatic endometrial cancer, women who did not take estrogen might not have any symptoms that would provoke an intra-endometrial diagnostic examination, but the use of estrogens might cause the bleeding that would lead to referral for the diagnostic test whose results detect a cancer that might otherwise be undiscovered. The increase in detection might take place at two separate phases of a patient's medical itinerary. For a patient in the community, the bleeding might
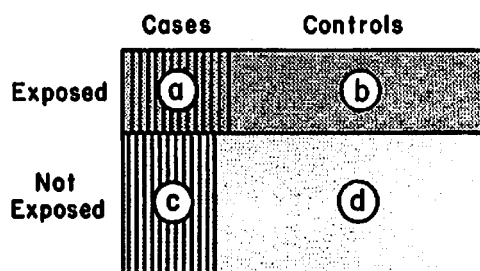
Figure 1. Assembly of the Case-Control Groups.

evoke increased medical surveillance, followed by the doctor's decision to seek the diagnostic test. For a patient hospitalized for whatever reason, the bleeding might evoke a direct ordering of the test. These two sources of an increased detection rate, rather than the pathogenetic effect of estrogens, might thus be responsible for the elevated odds ratio.

To test this possibility, we decided to perform two separate case-control studies at the same institution. In the first study, we would select cases and controls by the same conventional process used in one of the five recent investigations. In the second study, we would use an alternative method of sampling, by letting the cases and controls emerge from the results found in a group of women referred to the hospital for the same intra-endometrial diagnostic procedure. The purpose of the alternative approach was to try to equalize the forces of "diagnostic surveillance" that might otherwise create major "detection bias" in the conventional data of cases and controls.

In this paper, we report the results of those two studies.

## PATIENT SELECTION AND RESEARCH METHODS

For both studies the starting population was postmenopausal women hospitalized at the Yale–New Haven Medical Center. In the conventional study, which duplicated the case-control groups chosen by Smith et al.,[1] the source of the selected patients was 561 women with gynecologic cancer listed in the Yale Tumor Registry between July 1, 1974, and June 30, 1976. Of these women, 119 had a diagnosis of endometrial cancer and became the case group. From the remaining women, 119 were matched for age (within four years) and race to become the controls. Among the 119 controls, 60 had carcinoma of the cervix, 43 had carcinoma of the ovary, 15 had carcinoma of the vulva, and one had carcinoma of the vagina.

In the alternative study, the source of the population was 6869 consecutive women who underwent dilatation and curettage or hysterectomy between January 1, 1974, and June 30, 1976. Of these women, 149 were diagnosed as having endometrial cancer and became the case group. The women with diagnoses other than uterine cancer were the comparative patients, from whom the controls were randomly selected among those previously matched for age (within four years) and race with each member of the case group. The diverse histologic diagnoses received by these 149 controls included: uterine polyps and leiomyomas, 49; atrophic endometrium, 36; proliferative endometrium, 14; hyperplastic endometrium, 12; secretory endometrium, two; and basal or resting endometrium, 46. (More than 149 diagnoses are listed because more than one histologic diagnosis may have been made on a single specimen of endometrial tissue.)

To be accepted as a case in either study, a patient was required to have endometrial carcinoma of Grade 1 or higher. Patients with Stage 0 "in situ" carcinoma of the endometrium were not accepted as either cases or controls.

In all previous studies of this topic, the investigators have rarely stipulated what was meant by "exposure to estrogens." Before the current research began, a decision was made, after consultation with gynecologic authorities, that estrogen exposure would be defined as at least 0.3 mg of conjugated estrogens per day for at least six months at any time before hospitalization. For both studies, the data about pharmacologic exposure to estrogens were collected, from the patients' hospital charts and from office records of the appropriate physicians, by data abstractors who were kept "blind" to the research hypothesis.

For each study, the data were collected in the form of a two-by-two table relating estrogen exposure or nonexposure to the patients' status as a case or control. Because of the role of uterine bleeding as a stimulus to hospitalization, each two-by-two table was further subdivided according to the presence or absence of such bleeding. Since the control patients were chosen by a matching procedure, each set of results could be analyzed with either a matched or unmatched form of analysis. Because the results of the matched and unmatched analyses were essentially identical, we shall present the results of the unmatched analyses, which are conceptually simpler and easier to understand.

For each of the original two-by-two tables, we calculated the odds ratio, in the conventional manner, as an estimate of the overall relative risk. The "statistical significance" was calculated by the use of the Fisher exact test; the 95 per cent confidence interval around the odds ratio was calculated according to the method described by Thomas.[8] When the original two-by-two tables were stratified according to the presence or absence of bleeding, each of the subsequent tables could have its own odds ratio, Fisher exact test and confidence intervals calculated in a similar manner.

## RESULTS

Table 1 shows the pertinent clinical features of the cases and controls in both studies. The mean age of the patients at the time of diagnosis and at the time of menopause was similar in all four groups. In both studies, the case groups had significantly more patients with nulliparity and obesity — factors that have previously been noted in relation to endometrial cancer.[9] However, there was no difference between the two studies in the prevalence of these factors in the two case groups or in the two control groups. Hypertension and diabetes occurred with similar rates of frequency among the cases and controls of both studies.

The relation of estrogen to endometrial cancer as found by the conventional sampling method is shown in Tables 2 and 3. In Table 2, 29 per cent of the cases

Table 1. Clinical Characteristics of the Case and Control Groups.

| CHARACTERISTIC | CASE GROUP | | CONTROL GROUP | |
|---|---|---|---|---|
| | CONVENTIONAL METHOD | ALTERNATIVE METHOD | CONVENTIONAL METHOD | ALTERNATIVE METHOD |
| No. of patients | 119 | 149 | 119 | 149 |
| Mean age of patient (yr) | 61±9* | 62±9 | 62±9 | 61±9 |
| Mean age at menopause (yr) | 50±4 | 50±4 | 48±7 | 50±5 |
| % Nulliparous | 26 | 27 | 18 | 13 |
| % Obese | 30 | 33 | 22 | 21 |
| % Hypertensive | 33 | 40 | 28 | 36 |
| % Diabetic | 10 | 11 | 8 | 8 |

*± SD.

Table 2. Relation of Estrogen to Endometrial Cancer in the Conventional Study.

| Group | Cases | Controls |
|---|---|---|
| Estrogen takers | 35 | 4 |
| Non-estrogen takers | 84 | 115 |
| Totals | 119 | 119 |
| Odds ratio | 11.98 | |
| 95% confidence interval | 4.02–47.73 | |
| Fisher's exact test | P = 0.001. | |

but only 3 per cent of the controls were estrogen takers. The odds ratio (calculated as [35 times 115] divided by [4 times 84]) is 11.98. The 95 per cent confidence interval excludes 1, and the Fisher exact test yields P = 0.001. Table 3 shows the same results, but with the patients stratified for bleeding. Among the women with uterine bleeding, 30 per cent (34 of 113) were estrogen takers, as compared with 4 per cent (one of 26) of the controls. Among the women with nonbleeding complaints, 17 per cent (one of six) of the case group were exposed to estrogens, as compared with three per cent (three of 93) of the controls.

Table 3. Results of Table 2, Stratified for Bleeding.

| Group | Presenting Complaint | | | |
|---|---|---|---|---|
| | Uterine Bleeding | | No Bleeding | |
| | case | control | case | control |
| Estrogen takers | 34 | 1 | 1 | 3 |
| Non-estrogen takers | 79 | 25 | 5 | 90 |
| Totals | 113 | 26 | 6 | 93 |
| Odds ratio | 10.76 | | 6.00 | |
| 95% confidence interval | 1.60–454.57 | | 0.09–89.81 | |
| Fisher's exact test | P = 0.005 | | P = 0.224 | |

For the group with uterine bleeding, the odds ratio is 10.76, and the 95 per cent confidence interval excludes 1. For the group with complaints other than uterine bleeding, the odds ratio is 6.00, and the 95 per cent confidence interval includes 1.

The results of the alternative sampling method are presented in Tables 4 and 5. The unstratified data, shown in Table 4, indicate that 30 per cent of the cases were estrogen takers (44 of 149), as compared with 15 per cent of the controls (23 of 149). The odds ratio is 2.3, with a 95 per cent confidence interval that excludes 1 and a Fisher exact test with P = 0.005. This ratio, although strikingly smaller than what was obtained by conventional sampling methods, may continue to reflect the consequence of estrogen-influenced detection bias.

To adjust for this important source of bias, the results are stratified according to the reason for hospitalization. As demonstrated in Table 5, among the women with uterine bleeding, 30 per cent (43 of 142) of the case group were estrogen takers, as compared with 20 per cent (18 of 89) of the controls. Among those without bleeding 14 per cent (one of seven) of the case group were estrogen takers, as compared with 8 per cent (five of 60) of the controls. The individual odds ratios are 1.71 for the group with uterine bleeding and 1.83 for the group with non-bleeding complaints. The similarity of the values obtained in the bleeding and nonbleeding groups suggests that the risk ratio is being accurately appraised.

Table 4. Relation of Estrogen to Endometrial Cancer, Studied by the Alternative Sampling Method.

| Group | Cases | Controls |
|---|---|---|
| Estrogen takers | 44 | 23 |
| Non-estrogen takers | 105 | 126 |
| Totals | 149 | 149 |
| Odds ratio | 2.30 | |
| 95% confidence interval | 1.26–4.25 | |
| Fisher's exact test | P = 0.005 | |

In both groups, the 95 per cent confidence interval around the odds ratio includes 1.

Table 6 shows the results for the alternative sampling method after exclusion of controls with histologic diagnoses that are possibly related to estrogens (patients with proliferative or hyperplastic endometrium). The unstratified data indicate that 30 per cent of the case group (44 of 149) were estrogen takers, as compared with 13 per cent of the controls (17 of 126). The odds ratio is 2.69, with a 95 per cent confidence interval that excludes 1 (1.46 to 4.94). In the patients who had uterine bleeding, 43 of the 142 case group were estrogen takers (30 per cent), as compared with 13 of the 73 controls (18 per cent). The odds ratio decreases to 2.0, with a 95 per cent confidence interval that excludes 1 (1.01 to 4.14).

Table 5. Results of Table 4, Stratified for Bleeding.

| Group | Presenting Complaint | | | |
|---|---|---|---|---|
| | Uterine Bleeding | | No Bleeding | |
| | case | control | case | control |
| Estrogen takers | 43 | 18 | 1 | 5 |
| Non-estrogen takers | 99 | 71 | 6 | 55 |
| Totals | 142 | 89 | 7 | 60 |
| Odds ratio | 1.71 | | 1.83 | |
| 95% confidence interval | 0.88–3.42 | | 0.03–20.97 | |
| Fisher's exact test | P = 0.123 | | P = 0.498 | |

t accepted

have rarely
Before the
onsultation
would be
day for at
For both
ogens were
ice records
were kept

a two-by-
e patients'
bleeding as
as further
bleeding.
procedure,
hed or un-
tched and
resent the
lly simpler

d the odds
the overall
by the use
val around
scribed by
atified ac-
the subse-
t test and

es of the
n age of
e time of
In both
ly more
ors that
ometrial
between
rs in the
groups.
similar
ntrols of

ancer as
is shown
he cases

d Control

| Group |
|---|
| Alternative Method |
| 149 |
| 61±9 |
| 50±5 |
| 13 |
| 21 |
| 36 |
| 8 |

**Table 6. Results for Alternative Sampling Method: Controls with Estrogen-Related Disorders Excluded.**

| GROUP | CASE | CONTROL |
|---|---|---|
| Unstratified for presenting complaint: | | |
| Estrogen takers | 44 | 17 |
| Non-estrogen takers | 105 | 109 |
| Totals | 149 | 126 |
| Odds ratio | 2.69 | |
| 95% confidence interval | 1.46–4.94 | |
| Fisher's exact test | P = 0.002 | |
| | | |
| Stratified for uterine bleeding as presenting complaint: | | |
| Estrogen takers | 43 | 13 |
| Non-estrogen takers | 99 | 60 |
| Totals | 142 | 73 |
| Odds ratio | 2.00 | |
| 95% confidence interval | 1.01–4.14 | |
| Fisher's exact test | P = 0.034 | |

## DISCUSSION

With a conventional procedure to select cases and controls, our results are similar to those of the five previous studies that reported an association between estrogens and endometrial cancer.[1-5] With the alternative method of selecting cases and controls, however, the results are quite different, resembling those of the two older studies,[6,7] with individual odds ratios that did not significantly differ from 1.

The alternative method tests a hypothesis that depends on two assumptions: that many cases of endometrial cancer are asymptomatic; and that the rates of diagnostic surveillance are higher in the estrogen group because estrogens provoke increased bleeding and referral for diagnosis. With these assumptions, estrogens may lead to an increased detection of the cancer but need not be a causal agent.

The assumptions are supported by at least two types of evidence. One set of evidence is the high proportion ($\geq$20 per cent) of uterine cancer that was found to be asymptomatic in three different investigations. In 1964, Hofmeister and Barbo,[10] performing routine endometrial sampling for more than 19,000 women in private gynecologic practice, found 66 women with uterine cancer, of whom 13, or 20 per cent, were symptom-free at the time of diagnosis. In 1966, Abramson and Driscoll,[11] after routine endometrial sampling for 1540 patients, reported five cases of uterine cancer, of which three, or 60 per cent, were asymptomatic. In 1970, Ng and Reagan,[12] working in the gynecology clinic of the Case–Western Reserve Medical Center, found 363 cases of endometrial cancer among patients who were routinely advised to have endometrial aspiration. Of those patients, 20 per cent were asymptomatic.

A second set of supporting evidence, found during our study, appears in Table 7. Consistent with the assumption that women with low-grade (Grade 1) cancer exposed to estrogens are likely to have uterine bleeding and to be referred to the hospital, the results show that Grade 1 cancer occurred in 63 per cent of the nonexposed bleeding women but in 88 per cent of those who bled and were exposed to estrogens (chi-square = 9.5; P<0.005). Among women without bleeding, the numbers are consistent with the assumption, but are too small to allow any meaningful statistical conclusions.

### The Validity of the Control-Group Selection

As noted earlier, the odds ratio will approximate the risk ratio only if the four constituent groups of the case-control study are selected without distortion. These four groups are referred to the hospital from among the people in the community who are exposed and diseased, exposed and nondiseased, nonexposed and diseased, and nonexposed and nondiseased. If these four community groups have been referred to the hospital at similar rates, the four hospital groups will suitably represent the exposure ratios that exist in the community. If the rates of referral are disparate, the exposure ratios found in the hospital groups may be distorted.

A source of such disparity is the uterine bleeding that commonly occurs as a side effect of exposure to estrogens. Since women with uterine cancer who take estrogens are more likely than non-estrogen takers to bleed and to be referred to the hospital for diagnostic testing, the value of $a$ in the exposed and diseased hospital case group will be elevated to a value, $a'$, that exceeds its corresponding value in the community. The exposure ratio for cases will be calculated as $a'/c$ and will thus be higher than the correct proportion, $a/c$. Consequently, since the odds ratio will be calculated as $a'/c \div b/d$, it will be falsely elevated because of the inevitable bias in the selection of the case group.

This bias will be incorporated into any case-control study that is conducted in the conventional manner, making no provision for the "detection bias" in the case group. An important role for the control group, therefore, is to counteract the effects of this bias.

A control group consisting of "other uterine dis-

**Table 7. Rate of Occurrence of Grade 1 Endometrial Cancer in Relation to Estrogen Exposure and Uterine Bleeding.**

| GROUP | NO. OF WOMEN* |
|---|---|
| Bleeders: Estrogen takers | 38/43 (88%) |
| Bleeders: Non-estrogen takers | 62/99 (63%) |
| Non-bleeders: Estrogen takers | 1/1 (100%) |
| Non-bleeders: Non-estrogen takers | 3/6 (50%) |
| Subtotal: Bleeders | 100/142 (70%) |
| Subtotal: Non-bleeders | 4/7 (57%) |
| Subtotal: Estrogen takers | 39/44 (89%) |
| Subtotal: Non-estrogen takers | 65/105 (62%) |
| Totals | 104/149 (70%) |

*Denominators represent all women in this category with endometrial cancer, & numerators women with Grade 1 endometrial cancer.

ease" will include many women who were also referred to the hospital for the diagnostic evaluation of uterine bleeding that may have been produced by estrogens. The value of b in the exposed and nondiseased members of this control group will thus be raised upward to a value of b'. When the odds ratio is calculated as a'/c ÷ b'/d, the effects of the elevated a' and b' will counteract each other, so that the result will more closely approximate the correct value of a/c ÷ b/d.

In a control group of women with other gynecologic cancers, uterine bleeding is not a common symptom, and the diverse clinical phenomena that lead to hospital referral are unaffected by estrogen usage. In this situation, the value of b/d, which is the exposure ratio for the controls, will be unaffected, and the odds ratio will be falsely elevated when calculated as a'/c ÷ b/d.

A control group of "all others," consisting of people without gynecologic cancer or other uterine disease, will include women in the community or women referred to the hospital for all other reasons. Since the proportionate composition of this group will also be unaffected by estrogen usage, the value of b/d as the exposure ratio of controls will again be unaffected, and the odds ratio will continue to be falsely elevated when calculated as a'/c ÷ b/d.

The true risk ratio will thus be inflated if the controls emerge either from a community group of "all others" or from patients with other gynecologic cancer. Consequently, the best chance of minimizing the influence of hospital referral bias, and of getting an undistorted value for the odds ratio, is to choose controls from among women with "other uterine disease." Even here, however, the exact effects of bleeding will be uncertain and best compensated for by stratification of the odds ratio according to the presence or absence of bleeding as a stimulus for hospital referral.

In the five recent studies[1-5] reporting a causal association between estrogens and endometrial cancer, the control groups were selected either from women with gynecologic cancer or from general-hospital or community groups. Since these control groups do not compensate for estrogen-influenced hospital referral bias in the case group, the high odds ratios found in those studies are artificially elevated. Similarly, in the study in which we replicated conventional methods (using women with other gynecologic cancers as controls), we also found an extremely high odds ratio. When the controls were selected from women with other uterine disease, as in our second sampling procedure, the elevation of the exposure ratio in the control group compensated for the elevation in the case group, and the odds ratio was more likely to approximate the true value of the risk ratio.

Our alternative sampling procedure was actually used, inadvertently, in two previous investigations. Dunn and Bradbury,[6] selecting their cases from among women with untreated endometrial cancer and their controls from women with postmenopausal bleeding, found an odds ratio of 1.1. Pacheco and Kempers,[7] after choosing cases and controls from among all patients seen for postmenopausal bleeding who underwent dilatation and curettage, found an odds ratio of 0.5. In both studies the selection of women with uterine bleeding as a comparative group led to greater similarity of diagnostic surveillance between cases and controls.

Believers in the customary sampling procedure for case-control studies would argue that our alternative method of selecting controls is improper because estrogen usage leads both to postmenopausal bleeding and subsequent dilatation and curettage. Rather than representing a weakness in patient selection, our alternative approach has the important scientific merit of compensating for the estrogen-induced detection bias that is irremediable in the case group when sampling is performed with the customary methods. If no such compensation is made in the control group, the major bias that exists in the case group will be neglected and will grossly distort the results.

A second counterargument is that our alternative-method control group should not contain patients with such possibly estrogen-related histologic diagnoses as hyperplastic or proliferative endometrium. This argument is answered by a comparison of the results shown in Tables 5 and 6. Regardless of whether such patients were included or excluded from the control group, the alternative sampling method produces odds ratios that are substantially lower than what is found with conventional procedures.

Except for noting this distinction, we are reluctant to draw any conclusions from Table 6, which was prepared to answer critics who believe that "estrogen-related disorders" should not be permitted in the control group. We believe that this table contains an unfair comparison. Since uterine bleeding attributable to estrogen-related endometrial hyperplasia and proliferation is largely responsible for the bias in the case group, the exclusion of such disorders from the control group removes a major mechanism that compensates for the bias. If these disorders do not belong in the control group, carcinomas accompanied by hyperplasia or proliferation do not belong in the cases.

The fact that the five recent case-control studies all "confirmed" one another does not constitute scientific proof that their conclusion is correct. The late Harold Dorn, in writing about case-control research, once said that "reproducibility does not establish validity, since the same mistake can be made repeatedly."[13]

Epidemiologic authorities have often stated that estimates[4] of the odds ratio could be distorted by the influence of exposure on the composition of the case and control groups. McMahon and Pugh, in their standard textbook, *Epidemiology: Principles and methods*, assert "...computation of relative risk involves two assumptions: (1) that the disease under study is relatively infrequent in both exposed and unexposed persons; and (2) that neither cases nor controls are selected in favor of either exposed or non-exposed

individuals...."[14] This statement does not indicate what should be done when "detection bias" produces a case group that is distorted in favor of exposed individuals. Our alternative method of selecting controls provides a compensation for this bias, by allowing the control group an opportunity to have received the same type of selection.

Regardless of the method used to adjust for bias in the case group, no such adjustments were performed in studies reporting a high risk ratio between estrogens and endometrial cancer. Because the ratios were greatly overestimated in those studies and because the ratios are much closer to 1 when a compensating control group is used, we conclude that the strength of the much publicized association between estrogens and endometrial cancer has doubtlessly been exaggerated and needs re-evaluation.

The type of problem we have cited is particularly disturbing because it is not unique to the case-control relation of estrogens and endometrial cancer. The absence of suitable attention to detection bias casts doubt on the odds ratios found for many other etiologic associations that have been explored with case-control studies. The detection-bias problem will arise whenever a target disease that can occur in asymptomatic or other subclinical forms is likely to be preferentially diagnosed in persons exposed to the alleged etiologic agent. Finding an effective solution to this problem offers a major scientific challenge to case-control investigators. We believe the alternative method proposed here offers one such solution.

## REFERENCES

1. Smith DC, Prentice R, Thompson DJ, et al: Association of exogenous estrogens and endometrial carcinoma. N Engl J Med 293:1164-1167, 1975
2. Ziel HK, Finkle WD: Increased risk of endometrial carcinoma among users of conjugated estrogens. N Engl J Med 293:1167-1170, 1975
3. Mack TM, Pike MC, Henderson BE, et al: Estrogens and endometrial cancer in a retirement community. N Engl J Med 294:1262-1267, 1976
4. McDonald TW, Annegers JF, O'Fallon WM, et al: Exogenous estrogen and endometrial carcinoma: case-control and incidence study. Am J Obstet Gynecol 127:572-580, 1977
5. Gray LA, Christopherson WM, Hoover RN: Estrogens and endometrial carcinoma. Obstet Gynecol 49:385-389, 1977
6. Dunn LJ, Bradbury JT: Endocrine factors in endometrial carcinoma: a preliminary report. Am J Obstet Gynecol 97:465-471, 1967
7. Pacheco JC, Kempers RD: Etiology of postmenopausal bleeding. Obstet Gynecol 32:40-46, 1968
8. Thomas DG: Exact confidence limits for the odds ratio in a 2 × 2 table. Appl Stat 20:105-110, 1971
9. Wynder EL, Escher GC, Mantel N: An epidemiological investigation of cancer of the endometrium. Cancer 19:489-520, 1966
10. Hofmeister FJ, Barbo DM: Cancer detection in private gynecologic practice: a concluding study. Obstet Gynecol 23:386-391, 1964
11. Abramson D, Driscoll SG: Endometrial aspiration biopsy. Obstet Gynecol 27:381-391, 1966
12. Ng ABP, Reagan JW: Incidence and prognosis of endometrial carcinoma by histologic grade and extent. Obstet Gynecol 35:437-443, 1970
13. Dorn HR: Philosophy of inferences from retrospective studies. Am J Public Health 43:677-683, 1953
14. McMahon B, Pugh TF: Epidemiology: Principles and methods. Boston, Little, Brown, 1970, pp 241-282

GUENTHER POHLMANN, M.D.

## CORRECTING A BIAS?

HORWITZ and Feinstein, in an article in this issue of the *Journal*, suggest that a selection bias explains the strong association between exogenous estrogens and endometrial cancer reported by many investigators. To correct this bias, they propose a method that was abandoned by others precisely because it introduced a large selection bias into a research design in which selection bias was not inherently an important feature.

Horwitz and Feinstein believe that much of the strong association observed between estrogen use and endometrial cancer results from increased medical attention and therefore greater detection of the cancer among estrogen users, in whom estrogen-induced vaginal bleeding commonly develops and who consequently come under close medical surveillance. In their study, they attempt to compensate for the presumed greater detection of endometrial cancer among estrogen users by choosing as a comparison group women with benign gynecologic conditions, which are also subject to greater detection among estrogen users. They argue against the use of women with other gynecologic cancers as a comparison group, because these neoplasms are likely to come to medical attention promptly, independently of estrogen use, and therefore would not compensate for the selection bias in the case series.

The selection bias to which they refer results from the selective screening of estrogen users for uterine abnormality because of vaginal bleeding induced by the hormone. To evaluate this bias, it is necessary to determine what effect such screening has on the number of people detected with a given pathologic condition. If the condition is a progressive one that would ultimately produce symptoms prompting medical attention, the number of new cases detected in a given period is very little influenced by a continuing screening activity. After an initial increase in the number of detected cases when the screening activity begins, the number of new cases found each year in a screened population will be exactly the same as the number found annually before screening. Because nearly all women with invasive endometrial cancer will ultimately have the disease diagnosed, screening will advance the date of diagnosis, but will have little effect on the total number of cases ultimately found. Therefore, estrogen users, especially long-term users, would be very little over-represented among a series of women with endometrial cancer. Since Horwitz and Feinstein did not consider short-term (less than six months) users to be estrogen users, their case series should be essentially free of selection bias.

The situation is quite different for the benign conditions that Horwitz and Feinstein prefer to use for the control series, including uterine polyps, leiomyomas and various phases of uterine overdevelopment. The effect of screening is to detect these conditions in some women who otherwise would not come to medical attention. Consequently the control series preferred by

Horwitz and Feinstein will be subject to the selection bias that they describe, though the case series will be essentially unbiased. Furthermore, it is possible that some of the benign conditions listed above are induced by exogenous estrogens. Both the screening effect and the induction of these benign conditions by estrogen would tend to yield falsely high estimates of estrogen use, as compared with what would be obtained from a valid control series. Thus, the analysis recommended by Horwitz and Feinstein compares a case series that has minimal selection bias with a control series that has a bias in the direction of exaggerating the frequency of estrogen use. The net effect is to underestimate the estrogen-cancer association.

On the other hand, the control series consisting of women with other gynecologic cancers is not subject to selection bias from screening, not only for the reason given by Horwitz and Feinstein (less of a tendency for vaginal bleeding), but also for the reason cited above for the case series (these conditions nearly always progress to become symptomatic and therefore are detected even without screening). Contrary to what Horwitz and Feinstein argue, a comparison of cases of endometrial cancer with a control series of women with other gynecologic cancers would be essentially without selection biases. From such a comparison they estimated a risk of endometrial cancer among estrogen users 12 times that of nonusers, which is the best estimate of the estrogen effect available from their data.

Among women with endometrial cancer, the selection bias that concerns Horwitz and Feinstein, if it exists at all, would be most apparent among short-term estrogen users. These women, when beginning estrogen use, would tend to have otherwise asymptomatic endometrial cancers detected early. The result would be an excess of short-term users among women with endometrial cancer; long-term users, on the other hand, would not be over-represented. An effect of estrogens on *selection* of cases would be most apparent for short-term users and least apparent for long-term users. This is the reverse of the usual dose-response relation that would be seen as an effect of estrogens on *induction* of endometrial cancer. Although Horwitz and Feinstein do not present their data according to duration of estrogen use, other investigators who have done so do not find the effect to be concentrated among short-term users.[1-3] To the contrary, the greatest risk has been consistently found among long-term users, refuting the contention of Horwitz and Feinstein that the associations reported could be explained by selection bias.

Case-control studies are not the only source of evidence linking estrogen use with endometrial cancer. Weiss et al.[4] have described a substantial increase in the incidence of endometrial cancer in the United States during the first half of this decade, a trend that may be largely attributable to increased use of exogenous estrogens. Instead of providing reassurance about the safety of estrogens, the data presented by Horwitz and Feinstein only add to the accumulating evidence that exogenous estrogens induce endometrial cancer.

Harvard University School         GEORGE B. HUTCHISON, M.D.
      of Public Health
Boston, MA 02115             KENNETH J. ROTHMAN, DR. P.H.

### REFERENCES

1. Ziel HK, Finkle WD: Increased risk of endometrial carcinoma among users of conjugated estrogens. N Engl J Med 293:1167-1170, 1975
2. Mack TM, Pike MC, Henderson BE, et al: Estrogens and endometrial cancer in a retirement community. N Engl J Med 294:1262-1267, 1976
3. Gray LA, Christopherson WM, Hoover RN: Estrogens and endometrial carcinoma. Obstet Gynecol 49:385-389, 1977
4. Weiss NS, Szekely DR, Austin DF: Increasing incidence of endometrial cancer in the United States. N Engl J Med 294:1259-1262, 1976

## WHAT DOES LABORATORY "QUALITY CONTROL" REALLY CONTROL?

IN 1947 Belk and Sunderman[1] presented the results of a study of the quality of performance of clinical laboratories. Since that time a plethora of papers have appeared in the medical and scientific literature verifying, amending and extrapolating the observations made in that publication. These studies differ from each other in approach, scope and statistical manipulation, but the conclusions, for the most part, are the same; clinical laboratory test results, unfortunately, are not always reliable. Another addition to the voluminous literature on this subject appears elsewhere in this issue of the *Journal*. One conclusion to be drawn from the article by McCormick et al. is that laboratory data are of better quality for specimens designated as controls than for blind specimens — a fact that any experienced laboratory professional knows intuitively. This situation is hardly surprising, for if each specimen could receive special attention — that is, could be analyzed more than once by the most highly skilled personnel — the results would surely be of better quality. Unfortunately, pressures for production and economy do not permit this level of performance on a regular basis.

As applied to clinical laboratories, the term "quality control" is widely misused. Most often it refers to a system whereby samples of known composition are analyzed and the results subjected to statistical analysis to determine analytical accuracy or precision or both. At best, this process assesses the quality of performance, but does nothing to control or improve it. Thus, "quality control," in that sense, probably should be replaced by "quality assessment," "proficiency testing" or "performance evaluation."

Maintenance and improvement of the quality of testing are always the responsibility of laboratory management. Assessment of performance, on the other hand, is often imposed on the laboratory by approval or licensing agencies and by study groups. The assessment process is usually carried out on test specimens that are virtually always identified as such. The acceptable limits set for proficiency-testing scor-

# "ALTERNATIVE" CONTROLS IN A CASE-CONTROL STUDY OF ENDOMETRIAL CANCER AND EXOGENOUS ESTROGEN

BARBARA S. HULKA, ROGER C. GRIMSON, BERNARD G. GREENBERG, DAVID G. KAUFMAN, WESLEY C. FOWLER, JR., CAROL J. R. HOGUE, GARY S. BERGER AND CHARLES C. PULLIAM

Hulka, B. S. (School of Public Health, U. of North Carolina, Chapel Hill, NC 27514), R. C. Grimson, B. G. Greenberg, D. G. Kaufman, W. C. Fowler, Jr., C. J. R. Hogue, G. S. Berger, and C. C. Pulliam. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. *Am J Epidemiol* 112:376–387, 1980.

To address the issue of detection bias among endometrial cancer cases and controls, women admitted to the North Carolina Memorial Hospital for dilatation and curettage (D&C) during 1970–1976 were selected as one of three control groups in a study of endometrial cancer and exogenous estrogen. Study subjects included 256 cases, 316 D&C controls, 224 gynecology controls and 321 community controls. The D&C controls had a higher frequency of estrogen use than either of the other control groups or the cases. These differences existed for both blacks and whites. When white cases were compared to either gynecology or community controls, relative risks were increased for long duration estrogen use and for recent use prior to diagnosis. With D&C controls, relative risks were not significantly different from unity irrespective of duration or recency of estrogen use. Exclusion of hyperplasias from the D&C controls had no substantive effect on these results. Bleeding was a presenting complaint for 92% of cases, 82% of D&C controls and 22% of gynecology controls. Both among cases and gynecology controls, there was no statistically significant association between bleeding and estrogen use, whereas this association was evident among D&C controls, and specifically among those who did not have pathologic evidence of endometrial hyperplasia. These data support the presence of detection bias among D&C controls but they do not provide evidence of this bias among endometrial cancer cases.

endometrial cancer; epidemiologic methods; estrogen

---

The appropriate method for selection of cases and controls in studies of endometrial cancer and exogenous estrogen has been strongly debated in recent publications (1–10). Horwitz and Feinstein (1) have argued that "detection bias" among cases is responsible for the elevated cancer risks reported in association with the use of exogenous estrogen. This bias is purported to come about as a result of estrogen-induced bleeding among women with previously asymptomatic endometrial cancer, when, owing to the bleeding, these women are referred for diagnostic evaluation, which then results in a diagnosis of endometrial cancer. It is presumed that if estrogen had not been administered, bleeding, referral and diagnosis would not have occurred at the same high rate of probability. This difference in the probability of referral among previously asymptomatic cases using estrogen and persisting asymptomatic cases not using estrogen forms the basis of the detection bias hypothesis.

To eliminate the differential effect of detection bias, Horwitz and Feinstein (1) performed a study in which both cases and controls were women who had received dilatation and curettage (D&C) and/or hysterectomy because of uterine bleeding. When endometrial cancer cases were compared with non-cancer cases (controls), the odds ratio was 1.7, suggesting only a small increase in the risk of endometrial cancer among users of estrogen. In a comparison study by the same authors, endometrial cancer cases and controls were selected from a tumor registry. The controls were women with other gynecologic cancers. When these groups were compared, the odds ratio was almost 12, suggesting a very large cancer risk due to estrogen. The authors stated that this large risk was fallacious because the cases had been subjected to detection bias whereas the controls had not.

On reviewing these data, we noted that the cases from both of their studies were very similar with respect to estrogen use (29 per cent and 30 per cent), and in terms of bleeding as a presenting complaint (95 per cent in both case series). It was the control groups that differed. Among controls drawn from the tumor registry, only 3 per cent had used estrogen and 22 per cent had bled. Among the D&C controls, 15 per cent had used estrogen and 60 per cent had bled. These differences between control groups suggested the possibility that detection bias was affecting the D&C controls to a much greater extent than either of the case groups. Since the D&C controls were proposed as the appropriate "alternatives," and they had been used in an earlier study (11), we decided to evaluate a similar comparison group within the context of a larger study.

In a case-control study of endometrial cancer and exogenous estrogen in North Carolina (12), three different comparison groups were used. Two of the groups were chosen according to traditional thinking; one included hospital admissions and the second comprised a probability sample of women in the community. The third group was composed of women undergoing D&C or endometrial biopsy. Since endometrial cancer is usually diagnosed by these procedures, equal diagnostic surveillance among cases and these controls was assured. In addition, the clinical symptoms experienced by these controls are likely to be similar to those of the cases, and these symptoms might reflect underlying constitutional similarities.

The purpose of this report is to evaluate the extent of detection bias among D&C controls and endometrial cancer cases. Endometrial cancer risk estimates using D&C controls will be compared to those obtained with the two other control groups, and a rationale will be provided to clarify the unbiased nature of these latter groups. The magnitude of the detection bias effect on the case series will be illustrated both with observed data and a hypothetical numerical example.

METHODS AND MATERIALS

The methods for this study have been presented elsewhere (12), and will be reviewed briefly here. The endometrial cancer cases were all those receiving their initial therapy at North Carolina Memorial Hospital (NCMH) between 1970 and 1976, and residing in the state at that time. The original histologically confirmed case series was reduced from 290 to 256 after histologic review was conducted independently by two internationally recognized gynecologic pathologists (Dr. James W. Reagan, Professor of Pathology and Reproductive Biology, Case Western Reserve U., Cleveland, OH, and Dr. Ralph M. Richart, Professor of Pathology, Division of Obstetric and Gynecologic Pathology, Columbia U., New York, NY). Only invasive cancers—adenocarcinomas, adenoacanthomas, adenosquamous carcinomas, clear cell carcinomas and undifferentiated carcinomas—were retained in the case series.

Hospital controls were selected from the pool of all gynecology admissions and consults on surgical or medical services of the NCMH during 1970 through 1976. Admissions to the gynecologic oncology service and women admitted primarily for D&C or endometrial biopsy were excluded. Controls were chosen to have frequency distributions comparable to those of the cases for age and year of admission within each racial group (black and white). Women who had had hysterectomies prior to the admission date, i.e., "index date" in this study, were initially retained for analysis. Their drug histories prior to hysterectomy and other attributes were found to be similar to those of the remaining gynecology controls. Since the date of hysterectomy was frequently 10 to 15 years prior to the index date, the women without intact uteri were excluded, leaving 224 gynecology controls.

Community controls were obtained

from a two-stage stratified sampling procedure which was designed to obtain a scientific selection of women residing in the major referral area of NCMH so that sample estimates would be valid predictors of population parameters. Fifty-two contiguous counties among the 100 counties in North Carolina formed the geographic boundaries for the sample. Among eligible women identified, 88 per cent were interviewed and these formed the pool from which 321 controls were selected based on the age distribution of the cases. Each control was randomly assigned an "index date" based on the distribution of hospital admission dates for the cases. Only data about events occurring prior to the index date were included in the analysis.

The D&C/endometrial biopsy controls (henceforth referred to as D&C controls) were obtained through review of all pathology reports in the NCMH Pathology Department from January, 1970, through December, 1976. Women were excluded from the pool of possible controls if they were under age 30 years, if the indication for the procedure was related to infertility or pregnancy, or if the procedure was "incidental" to another surgical procedure, e.g., hysterectomy, sterilization or conization. The remaining women were distributed by age and index date within racial group so that a control series similar in composition to the cases could be randomly selected. The final number of D&C controls was 316.

Outside pathological review was obtained on a sample consisting of one seventh of the gynecology and the D&C controls. Slides from these subjects were intermixed with those from the cases, and all were reviewed by the same pathologists, who had no knowledge of whether the slides came from cases or controls. No control subject received an endometrial cancer diagnosis by either pathologist.

Data on estrogen use, sources of medi-

cal care since 1960, and reproductive and medical histories were obtained from multiple sources. Living subjects were interviewed as were relatives of those who were deceased. Hospital records were abstracted for cases, gynecology and D&C controls. Office records of the usual physician seen by each woman were abstracted, and additional physicians completed a questionnaire on estrogen use. In summary, interview data were obtained on 100 per cent of community controls, 88 per cent of cases, 88 per cent of D&C controls and 87 per cent of gynecology controls. Medical record abstracts from referring physicians' offices were obtained for 73 per cent of cases, 62 per cent of D&C controls, 65 per cent of gynecology controls and 76 per cent of community controls. Among women with additional physicians, drug history questionnaires were received for 83 per cent of cases, 79 per cent of D&C controls, 75 per cent of gynecology controls, and 91 per cent of community controls. Medical record abstracts or drug history questionnaires were obtained from a total of 660 physicians.

The factors screened as potential confounders or effect modifiers included age, obesity, hypertension, diabetes, gall bladder disease, parity, social class, index year and characteristics of estrogen use. A series of logistic discriminant analyses (13–15) were used to determine those variables that significantly contributed to endometrial cancer risk or to estrogen use. There was no single variable or set of variables which could accurately differentiate cases from D&C controls or estrogen users from non-users. Relative risks, as estimated from the odds ratios, are presented in this report. Although age was not a significant effect modifier in the D&C control data, these controls were slightly younger on average than the cases. Therefore, summary odds ratios and confidence intervals were obtained by weighted averages, wherein the weights were determined by the inverses of the variances of the age-specific strata (16). Odds ratios adjusted by the Mantel-Haenszel method (16) produced essentially identical results.

## RESULTS

Table 1 shows clinical and personal characteristics of the cases and of each control group. The D&C controls are similar to the gynecology controls except for a lower mean age and a lower per cent with history of gall bladder disease. The high

TABLE 1

*Characteristics of cases and three control groups: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Characteristic | Cases | Controls | | |
| --- | --- | --- | --- | --- |
| | | D&C* | Gynecology | Community |
| No. of subjects | 256 | 316 | 224 | 321 |
| Mean age (years) | 61.2 ± 11.5† | 58.6 ± 10.2 | 61.4 ± 13.6 | 56.3 ± 11.4 |
| % black | 27.3 | 34.2 | 31.7 | 26.5 |
| % nulliparous | 24.8 | 12.5 | 13.5 | 15.6 |
| % obese | 52.3 | 43.0 | 43.8 | 35.8 |
| % hypertensive | 59.8 | 56.0 | 57.1 | 39.9 |
| % diabetic | 21.9 | 22.5 | 20.1 | 8.1 |
| % with gall bladder disease | 14.8 | 15.2 | 23.2 | 10.9 |
| Mean age (years) at menopause | 47.1 ± 5.4 | 47.5 ± 4.4 | 46.5 ± 4.6 | 46.4 ± 5.2 |
| Mean index year | 1973.5 ± 2.0 | 1973.3 ± 2.0 | 1973.0 ± 1.9 | 1973.5 ± 2.0 |

\* D&C = dilatation and curettage.

† One standard deviation.

frequency of gall bladder disease among the gynecology controls is probably due to their source of ascertainment; about 40 per cent were gynecology consults on medical or surgical services. The D&C controls differ from the cases in having a slightly lower mean age, a larger per cent black and a lower per cent nulliparous and obese. An excess of the latter two characteristics among cases has been reported frequently (11, 17, 18).

The distributions of the first listed discharge diagnoses for D&C and gynecology controls are shown in table 2. Bleeding, hyperplasia and other endometrial or cervical lesions account for 76 per cent of primary diagnoses among the D&C con-

trols. Medical diagnoses, stress incontinence or urinary tract problems, and other cancers account for 67 per cent of the primary diagnoses among the gynecology controls. The differences in the types of diagnoses experienced by each control group are consistent with their different sources of identification.

In table 3, the frequency of any reported estrogen use is shown separately for whites and blacks. Among white D&C controls, 35 per cent reported estrogen use compared to 33 per cent of cases, 23 per cent of gynecology controls and 27 per cent of community controls. Black women exhibited a low frequency of estrogen use except for the D&C controls, among whom 24 per cent reported estrogen use.

Endometrial cancer risks by duration of estrogen use appear in table 4 with cases compared to each of the three control groups. Data are shown for white women only since duration of estrogen use was known for only three of the seven black cases who had ever used estrogen; among the three, only one had used estrogen for more than six months. With all control groups, the relative risks are less than unity for estrogen use duration of less than 3.5 years. With the gynecology and community controls, the relative risks increase with increasing duration of use until the relative risks significantly exceed unity. By 9.5 years or more, the risks are in excess of five-fold with either the gynecology or community controls. When cases are compared with the D&C con-

TABLE 2

*Discharge diagnoses for two control groups: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Diagnoses* | Controls | | | |
|---|---|---|---|---|
| | D&C† | | Gynecology | |
| | No. | % | No. | % |
| Uterine bleeding | 166 | 52.5 | 10 | 4.5 |
| Fibroids | 13 | 4.1 | 15 | 6.7 |
| Hyperplasia or atypical hyperplasia | 26 | 8.2 | 1 | 0.4 |
| Other endometrial or cervical lesions | 47 | 14.9 | 12 | 5.4 |
| Stress incontinence or urinary tract problems | 6 | 1.9 | 42 | 18.7 |
| Other gynecologic diagnoses | 19 | 6.0 | 24 | 10.7 |
| Other cancers | 17 | 5.4 | 35 | 15.6 |
| Medical diagnoses | 16 | 5.1 | 73 | 32.6 |
| Miscellaneous | 6 | 1.9 | 12 | 5.4 |
| Total | 316 | 100.0 | 224 | 100.0 |

* The first listed diagnosis is presented.
† D&C = dilatation and curettage.

TABLE 3

*Per cent of cases and controls reporting any estrogen use, by race: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Race | Cases | | | Controls | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | D&C* | | | Gynecology | | | Community | | |
| | Total no. | Estrogen | | Total no. | Estrogen | | Total no. | Estrogen | | Total no. | Estrogen | |
| | | No. | % | | No. | % | | No. | % | | No. | % |
| White | 186 | 61 | 32.8 | 208 | 72 | 34.6 | 153 | 35 | 22.9 | 236 | 64 | 27.1 |
| Black | 70 | 7 | 10.0 | 108 | 26 | 24.1 | 71 | 9 | 12.7 | 85 | 7 | 8.2 |

* D&C = dilatation and curettage.

TABLE 4

*Effect of duration of estrogen use on relative risks (RRs),\* using three control groups among white women: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Duration of use | No. of cases | Controls | | | | | | | | |
| | | D&C† | | | Gynecology | | | Community | | |
| | | No. | RR | 95% CI‡ | No. | RR | 95% CI | No. | RR | 95% CI |
|---|---|---|---|---|---|---|---|---|---|---|
| None used | 125 | 136 | | | 118 | | | 172 | | |
| <6 months | 8 | 13 | 0.7 | (0.3,1.8) | 12 | 0.7 | (0.3,1.8) | 20 | 0.8 | (0.3,1.9) |
| 6 months – <3.5 years | 9 | 14 | 0.7 | (0.3,1.7) | 9 | 0.9 | (0.3,2.6) | 21 | 0.7 | (0.3,1.6) |
| 3.5 years – <6.5 years | 9 | 16 | 0.8 | (0.3,1.8) | 1 ⎱ 3.8 | | (1.2,12.1) | 7 | 1.7 | (0.7,4.7) |
| 6.5 years – <9.5 years | 9 | 11 | 1.2 | (0.5,3.1) | 2 ⎰ | | | 5 | 2.5 | (0.8,7.4) |
| ≥9.5 years | 19 | 10 | 2.0 | (0.8,4.7) | 2 | 5.1 | (1.4,18.5) | 4 | 5.5 | (1.9,16.2) |
| No data on duration | 7 | 8 | | | 9 | | | 7 | | |

\* Age-adjusted with four age groups: <50, 50–59, 60–69 and 70+ years.

† D&C = dilatation and curettage.

‡ CI = confidence interval.

trols, however, there is no statistically significant elevated endometrial cancer risk, irrespective of estrogen use duration.

The relationship between the estrogen-free interval prior to diagnosis (i.e., the recency of estrogen use) and the relative risk appears in table 5. If estrogen use is continued within the six months prior to diagnosis, case comparisons with gynecology and community controls produce relative risks of almost three, whereas with D&C controls, the relative risk is about unity. When estrogen has been discontinued for six months or more prior to diagnosis, risks are not significantly different from unity for case comparisons with any of the control groups.

It might be argued that women with hyperplasia should be excluded from the D&C control group. Since some of the hyperplasias may represent premalignant lesions, their inclusion could result in controls that are overmatched to the cases. In view of this possible bias, we repeated the analysis excluding the hyperplasias from the D&C controls. Among the 316 D&C controls, 107 women had hyperplasia on the index date or on a prior D&C. These included 66 whites and 41 blacks with endometrial hyperplasia, adenomatous hyperplasia or atypical hyperplasia. Excluding these women from the D&C controls had only a minor effect

on the risks presented in tables 4 and 5. The risks for each duration of estrogen use were not significantly altered with the exception of the longest duration of use (9.5 years or more) for which the relative risk became 2.7 with a 95 per cent confidence interval touching 1.0 at the lower boundary. The risks for recency were not significantly different from those shown in table 5, and none was significantly different from unity. Combining blacks and whites without hyperplasia and comparing them to the combined black and white cases produced almost identical results to those described for whites alone.

Frequency of uterine bleeding as a presenting complaint at the index date of admission or procedure appears in table 6. Ninety-two per cent of cases, 82 per cent of D&C controls and 22 per cent of the gynecology controls presented with bleeding. The bleeding occurred only *after* estrogen administration except for two cases and three D&C controls who admitted to bleeding prior to the use of estrogen. History of estrogen use produced a statistically significant increase in the per cent of bleeders among D&C controls but no increase among gynecology controls. Among cases, a suggestive association between bleeding and estrogen use appeared, but it was not even close to being statistically significant.

TABLE 5

*Effect of recency of estrogen use on relative risks (RRs),\* using three control groups among white women: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Interval from last use | No. of cases | Controls | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | D&C† | | | Gynecology | | | Community | | |
| | | No. | RR | 95% CI‡ | No. | RR | 95% CI | No. | RR | 95% CI |
| None used | 125 | 136 | | | 118 | | | 172 | | |
| <6 months§ | 31 | 37 | 1.1 | (0.6,1.9) | 9 | 2.9 | (1.2,6.9) | 18 | 2.8 | (1.4,5.5) |
| 6 months– <2.5 years | 9 | 9 | 1.3 | (0.5,3.3) | 3 | 2.1 | (0.6,7.1) | 10 | 1.4 | (0.5,3.5) |
| ≥2.5 years | 7 | 15 | 0.6 | (0.2,1.5) | 9 | 0.7 | (0.2,2.2) | 26 | 0.4 | (0.2,0.9) |
| No data on recency | 14 | 8 | | | 8 | | | 6 | | |

  \* Age-adjusted with four age groups: <50, 50–59, 60–69 and 70+ years.
  † D&C = dilatation and curettage.
  ‡ CI = confidence interval.
  § Only women who started estrogen more than six months prior to diagnosis and continued to take it within the six months of diagnosis are included in this category.

TABLE 6

*Frequency of uterine bleeding as a presenting complaint by estrogen use among cases and controls: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Estrogen used | Cases | | | Controls | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | D&C* | | | Gynecology | | |
| | Total no. | Bleeding | | Total no. | Bleeding | | Total no. | Bleeding | |
| | | No. | % | | No. | % | | No. | % |
| No | 188 | 170 | 90.4 | 218 | 172 | 78.9 | 180 | 39 | 21.7 |
| Yes | 68 | 65 | 95.6 | 98 | 87 | 88.8 | 44 | 10 | 22.7 |
| Total | 256 | 235 | 91.8 | 316 | 259 | 82.0 | 224 | 49 | 21.9 |
| | $\chi^2 = 1.148$ | | | $\chi^2 = 3.818$ | | | $\chi^2 = 0.003$ | | |
| | $p = 0.28$ | | | $p = 0.05$ | | | $p = 0.96$ | | |

  \* D&C = dilatation and curettage.

Hyperplasia was diagnosed in 34 per cent (107/315) of D&C controls. (Status of hyperplasia was unknown for one D&C control who did not use estrogen.) When disaggregated by estrogen usage, 42 per cent (41/98) of estrogen users and 30 per cent (66/217) of non-users had hyperplasia. This difference between 42 per cent and 30 per cent is significant at the 0.06 probability level.

The relationship between hyperplasia, bleeding and estrogen use among the D&C controls is shown in table 7. Among women without hyperplasia, bleeding is more common among estrogen users (93 per cent) than among non-users (78 per cent) and the difference is significant at

the 0.02 probability level. The most common pathologic findings in this group were unremarkable endometrium, atrophic endometrium and endometrial polyps. Among women with hyperplasia, the per cent of bleeders is similar for prior estrogen users and non-users (83 and 82 per cent, respectively). In these data, estrogen use is not associated with the frequency of bleeding among women with hyperplasia.

DISCUSSION

*Community and gynecology controls: Lack of bias*

The use of D&C controls produced relative risks which were not consistent with

TABLE 7

*Uterine bleeding and hyperplasia in relation to estrogen use among D&C\* controls: Case-control study of endometrial cancer and exogenous estrogen, North Carolina, 1970–1976*

| Estrogen used | Hyperplasia† | | | Not hyperplasia | | |
|---|---|---|---|---|---|---|
| | Total no. | Bleeding | | Total no. | Bleeding | |
| | | No. | % | | No. | % |
| No | 66 | 54 | 81.8 | 151 | 117 | 77.5 |
| Yes | 41 | 34 | 82.9 | 57 | 53 | 93.0 |
| Total | 107‡ | 88 | 82.2 | 208‡ | 170 | 81.7 |
| | $\chi^2 = 0.013$ | | | $\chi^2 = 5.66$ | | |
| | $p = 0.91$ | | | $p = 0.02$ | | |

\* D&C = dilatation and curettage.

† Hyperplasia includes endometrial hyperplasia, adenomatous hyperplasia and atypical hyperplasia.

‡ Missing data on one woman who had not used estrogen.

those produced with community or gynecology controls. Attributing this difference in relative risks to detection bias in the D&C controls is justifiable only if the other control groups can be shown to be unbiased.

The community sample was selected to be statistically representative of the female population over age 30 years residing in the major referral areas of NCMH. From this pool of women, a stratified random sample of controls was selected based on the race and age distribution of the cases. Data on estrogen use and medical conditions back to 1960 were collected from the personal physicians of each woman and from the woman herself through personal interview. These controls provide an accurate description of estrogen use patterns among women in the relevant age groups in the Piedmont and Eastern regions of North Carolina.

The gynecology controls could be subject to objections. Because of the nature of their complaints, gynecology admissions and consults might be more likely than other women to have had estrogen prescribed. Even the exclusion of women admitted primarily for D&C or endometrial biopsy does not preclude this possibility. In fact, however, the white gynecology controls reported less estrogen use than the community controls, a finding which does not support the notion of excess estrogen use among these gynecology controls.

Furthermore, a consistent pattern of results, using a variety of analytic methods (13–16), was obtained from both the gynecology and the community controls. Although similar selection biases could produce similar but erroneous results, this is most unlikely when the two control groups were selected from completely different reference populations as was the situation in this study.

## Results using D&C controls

The most prominent differences between the D&C group and the other two control groups were in their patterns of estrogen use. The D&C controls had a higher frequency of estrogen use than either of the other control groups or the cases. These differences existed for both black and white women.

When white cases were compared to white D&C controls, the age-adjusted relative risks showed no significant increase with increasing duration of use. For each category of duration, the relative risk estimates were not significantly different from unity. In other reported case-control studies (10, 19–22) and in our study with community and gynecology controls (12), endometrial cancer risk was greatest for long-duration estrogen use, whereas short-duration use did not increase the

risk. These findings are just the opposite from those required to support the hypothesis of detection bias among cases. Presumably, it would not take long for an existing but previously undiagnosed cancer to bleed after exposure to estrogen, so high risks for short-duration estrogen use should be evident.

If cancers bleed soon after estrogen administration and women respond promptly by seeking medical attention, we would expect to find high risks for women who first started taking estrogen within a short time interval prior to diagnosis. That women do respond promptly to abnormal bleeding, whether they use estrogen or not, was reported by Antunes et al. (10). Therefore, the finding that none of our cases, white or black, had first started estrogen within six months of their hospital admission is not consistent with detection bias. Among the cases using estrogen for whom the date of first administration was known, 88 per cent (53/60) started the estrogen 3.5 years or more prior to hospital admission.

When D&C controls were compared to cases, relative risks were not elevated for recent estrogen users as compared to those who terminated use at a time more remote from the diagnosis. All relative risks were close to unity or less, irrespective of the recency interval. With our other control groups (12) and in the report of Mack et al. (19), risks were greatest for recent estrogen use prior to diagnosis and dropped as the estrogen-free interval increased. These findings using more traditional controls are consistent with the detection bias hypothesis. They are also consistent with the concept that estrogen acts as a promotional agent whose effects are dissipated within two years after stopping estrogen.

An argument can be made for removing hyperplasias from the control group in order to avoid the possibility of overmatching. If some hyperplasias are cancer precursors, the two conditions may have similar risk factors, some of which may be linked to estrogen use. Hyperplasias were excluded from the D&C control group used by Dunn and Bradbury (11), and they found equal proportions of estrogen users among cases and controls. Since 34 per cent of our D&C controls had diagnosed hyperplasia, they were excluded and the analysis was repeated. This modification of the control group had no substantive effect on the results.

*Estrogen and endometrial bleeding*

The D&C controls were decidedly different from the gynecology controls and were very similar to the cases in frequency of uterine bleeding as a presenting complaint. The frequency of bleeding was 92 per cent for cases, 82 per cent for D&C controls, and 22 per cent for gynecology controls. Among the D&C controls, estrogen users had a greater percentage of bleeders than non-users, whereas the gynecology controls exhibited no such relationship. That bleeding was not related to estrogen use among the gynecology controls indicates different causes for the bleeding from those affecting the D&C controls and is consistent with the differences in discharge diagnoses for the two groups as presented in table 2. The increased frequency of bleeding associated with estrogen use among the D&C controls provides evidence for bias in this control group; women who bleed and get D&C are more likely to have received estrogen than other women.

Our data confirm the association between estrogen use and hyperplasia of the endometrium. Among the D&C group, 42 per cent of estrogen users had hyperplasia, compared to 30 per cent of the non-users. However, within the hyperplasia group, estrogen use was not related to the frequency of bleeding. Since some hyperplasias are thought to be cancer precursors, one might expect estrogen to induce bleeding from hyperplastic lesions in analogous fashion to the pos-

tulated estrogen-induced bleeding from endometrial carcinoma. Our data do not support the concept of estrogen-induced bleeding in either hyperplasia or endometrial cancer. We do find a high frequency of bleeding in conjunction with hyperplasia and an even higher frequency among diagnosed endometrial carcinomas, but this bleeding does not appear to be a function of exogenous estrogen stimulation. Without a strong association between estrogen use and bleeding from carcinoma (or its hyperplastic precursor), the probability of an important role for detection bias among cases is greatly reduced.

### Detection bias or earlier diagnosis

An important issue requiring clarification is the distinction between detection bias and earlier diagnosis. The concept underlying detection bias is that cases are identified through the mechanism of estrogen administration, bleeding and referral for diagnostic evaluation, and that, in the absence of this sequence of events, these cases would not be diagnosed. As an alternative, this same sequence of events could merely promote the earlier diagnosis of cases which would otherwise have come to medical attention at a later stage in their development. The literature (10, 12, 19, 20, 23) confirms the greater frequency of in-situ, minimally invasive and stage IA carcinomas among the estrogen-associated cancers as compared to the non-estrogen-associated cancers. These findings are consistent with the operation of either detection bias or earlier diagnosis. Two types of events would favor detection bias. First, there could be a significant number of women dying, or otherwise leaving the population permanently, between the time of early diagnosis and the time of usual diagnosis. In an elderly population, high death rates from causes other than endometrial cancer would remove women from the population

after an early diagnosis but before the time of usual diagnosis. A second mechanism would rely on the existence of a large pool of asymptomatic cancers with a very low malignant potential. If such cancers exist, early diagnosis through estrogen administration could make them manifest, whereas, without estrogen, their existence would remain unknown throughout the women's normal lifespan. Existing knowledge about the natural history and variant biological forms of endometrial cancer is insufficient either to confirm or preclude the existence of such indolent forms of endometrial cancer.

### Detection bias among cases: Hypothetical example

For detection bias to affect the cases, it is necessary to assume that estrogen use causes bleeding from previously asymptomatic endometrial cancers—although, in our data, estrogen use was not associated with a significantly increased frequency of bleeding in the diagnosed case series. However, given the assumption of estrogen-induced bleeding, the proportion of diagnosed cases affected by detection bias can be estimated as can the effect of this proportion on the risk estimates. The magnitude of the detection bias effect will be determined by the size of the intersection among three groups of women: women with asymptomatic cancer, women using estrogen and women who make up the diagnosed endometrial cancer cases. This can be illustrated with a hypothetical population of 100,000 women age 50 years and over, as shown in the Venn diagram in figure 1. Estimates of rates and proportions obtained from a number of sources have been used to obtain the numbers in the diagram, as indicated below the figure.

The remaining numbers in the subsets of the Venn diagram can be obtained arithmetically. Only five cases appear in
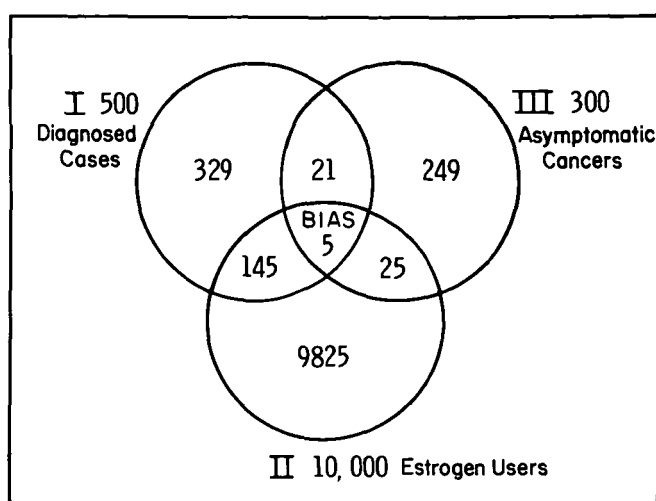
FIGURE 1. Detection bias and endometrial cancer: Diagnosed cases and asymptomatic cases. Assuming a hypothetical population of 100,000 women age ⩾50 years, three groups are formed: I = the 5-year cumulative incidence of diagnosed cancer; II = the 5-year period prevalence of estrogen use; III = the 5-year period prevalence of asymptomatic cancer.

*Sources of estimates of rates and proportions used to obtain numbers in diagram*

1) Incidence of diagnosed endometrial cancer = 1/1000/year* × 5 years × 100,000 women = 500 diagnosed cases (Group I).
2) 5-year period prevalence of estrogen use ≈ 10%† of 100,000 women = 10,000 women having used estrogen (Group II).
3) 5-year period prevalence of asymptomatic cancers ≈ 3/1000 (27, 28) × 100,000 women = 300 asymptomatic cancers (Group III).
4) 30% (1) of diagnosed cases used estrogen = 0.30 × 500 = 150.
5) 10%† of asymptomatic cases used estrogen = 0.10 × 300 = 30.
6) 20%‡ of estrogen users with previously asymptomatic cancer bled and became diagnosed cases = 0.20 × 30 ≈ 5.
7) 6% (1) of non-estrogen-using diagnosed cases were asymptomatic = 0.06 × 350 = 21.

* This figure is based on the highest age-specific rates reported from the Third National Cancer Survey (24) plus a small increment which allows for the increasing incidence reported in the 1970s from several state and local cancer registries (25, 26).

† 10% appears to be a reasonable estimate based on the proportion of controls using estrogen for 6 months or more in the case-control studies published since 1975 (1, 10, 19–22).

‡ Estimated incidence of uterine bleeding is 0.3% per cycle among postmenopausal women using conjugated estrogen (29). Assuming a 28-day cycle for estrogen administration and ≈ 13 cycles/year × 5 years = 65 cycles × 0.003 ≈ 20% incidence of bleeding in 5 years.

the intersection of the three groups—diagnosed cancer cases, estrogen users and asymptomatic cancer cases. Estimates of the relative risk may be obtained from the numbers in the diagram. Including the cases subject to detection bias, the relative risk ($RR$) is estimated as follows:

$$RR = \frac{P(D \mid E)}{P(D \mid \bar{E})} = \frac{\frac{150}{10,000}}{\frac{350}{100,000 - 10,000}} = 3.9.$$

Detection bias is removed from the relative risk by removing the five subjects in the intersection of the three groups:

$$RR = \frac{P(D \mid E)}{P(D \mid \bar{E})} = \frac{\frac{145}{10,000}}{\frac{350}{100,000 - 10,000}} = 3.7.$$

Both relative risks are consistent with those reported elsewhere (12, 20, 22) and the two relative risks are very similar.

*Additional studies*

Although we find very little evidence of detection bias in either an observed case series or in a hypothetical one, focusing on this possible bias may have served an important function in highlighting the need for additional research on the natu-

ral history of endometrial cancer. More data are needed about cancer precursors, those that are truly destined to progress and those which may be reversible. To what extent endometrial cancer is asymptomatic, and for how long, is not known. The biologic characteristics of the cancer and its variants, including the possibility of a less aggressive form of the disease, require study. Once cancer is established, the prognostic effects of clinical stage, myometrial invasion and tumor grade, have been studied (30), but these have not been evaluated in conjunction with exogenous estrogen which may also affect the prognosis.

## REFERENCES

1. Horwitz RI, Feinstein AR: Alternative analytic methods for case-control studies of estrogens and endometrial cancer. N Engl J Med 299:1089–1094, 1978
2. Hutchinson GB, Rothman KJ: Correcting a bias? Editorial. N Engl J Med 299:1129, 1978
3. DuPont WD: Letter to the editor. N Engl J Med 300:495, 1979
4. Franks P: Letter to the editor. N Engl J Med 300:495, 1979
5. Soutter WP: Letter to the editor. N Engl J Med 300:496, 1979
6. Kay CR: Letter to the editor. N Engl J Med 300:496, 1979
7. Cole F: Letter to the editor. N Engl J Med 300:496, 1979
8. Donelson RK: Letter to the editor. N Engl J Med 300:496, 1979
9. Feinstein AR, Horwitz RI: Letter to the editor. N Engl J Med 300:497, 1979
10. Antunes CMF, Stolley PD, Rosenshein NB, et al: Endometrial cancer and estrogen use: Report of a large case-control study. N Engl J Med 300:9–13, 1979
11. Dunn LJ, Bradbury JT: Endocrine factors in endometrial carcinoma. Am J Obstet Gynecol 97:465–471, 1967
12. Hulka BS, Kaufman DG, Fowler WC, et al: Estrogen and endometrial cancer: Cases and two control groups from North Carolina. Am J Obstet Gynecol 137:92–101, 1980
13. Prentice R: Use of the logistic model in retro-
spective studies. Biometrics 32:599–606, 1976
14. Prentice RL, Breslow, NE: Retrospective studies and failure time models. Biometrika 65:153–158, 1978
15. Breslow N, Powers W: Are there two logistic regressions for retrospective studies? Biometrics 34:100–105, 1978
16. Fleiss JL: Statistical Methods for Rates and Proportions. John Wiley and Sons, New York, 1973
17. MacMahon B: Risk factors for endometrial cancer. Gynecol Oncol 2:122–129, 1974
18. Synder EL, Escher GC, Mantel N: An epidemiological investigation of cancer of the endometrium. Cancer 191:489–520, 1966
19. Mack TM, Pike MC, Henderson BE, et al: Estrogens and endometrial cancer in a retirement community. N Engl J Med 294:1262–1267, 1976
20. McDonald TW, Annegers JF, O'Fallon WM, et al: Exogenous estrogen and endometrial carcinoma: Case-control and incidence study. Am J Obstet Gynecol 127:572–580, 1977
21. Ziel HK, Finkle, WD: Increased risk of endometrial carcinoma among users of conjugated estrogens. N Engl J Med 293:1167–1170, 1975
22. Gray LA Sr, Christopherson WM, Hoover RN: Estrogens and endometrial carcinoma. J Am Coll Obstet Gynecol 49:385–389, 1977
23. Smith DC: Report Presented at the Meeting of The Obstetrics and Gynecology Advisory Committee, Food and Drug Administration. Washington DC, December 16, 1975
24. Third National Cancer Survey: Incidence Data. National Cancer Institute Monograph 41, March, 1975. DHEW Pub. No. (NIH) 75-787. Bethesda, MD, US DHEW, 1975, pp 110, 122 and 134
25. Weiss NS, Szekely DR, Austin DE: Increasing incidence of endometrial cancer in the United States. N Engl J Med 294:1259–1262, 1976
26. Austin DF, Roe KM: Increase in cancer of the corpus uteri in the San Francisco-Oakland Standard Metropolitan Statistical Area, 1960–75. JNCI 62:13–16, 1979
27. Hofmeister FJ, Barbo DM: Cancer detection in private gynecologic practice: A concluding study. Obstet Gynecol 23:386–391, 1964
28. Abramson D, Driscoll SG: Endometrial aspiration biopsy. Obstet Gynecol 27:381–391, 1966
29. Shoemaker ES, Forney JP, MacDonald PC: Estrogen treatment of postmenopausal women—benefits and risks. JAMA 238:1524–1530, 1977
30. Malkasian GD: Carcinoma of the endometrium: Effect of stage and grade on survival. Cancer 41:996–1001, 1978

# AN ANALYSIS OF DETECTION BIAS AND PROPOSED CORRECTIONS IN THE STUDY OF ESTROGENS AND ENDOMETRIAL CANCER

SANDER GREENLAND and RAYMOND NEUTRA

Division of Epidemiology, School of Public Health, University of California,
Los Angeles, CA 90024, U.S.A.

**Abstract**—We present an algebraic analysis of detection bias and the Horwitz–Feinstein correction in the study of estrogens and endometrial cancer. This analysis sets forth precisely how the validity of conventional study results depends on the cancer detection rates, and how the validity of the Horwitz–Feinstein 'alternative' approach depends on both the detection rates and the rates of the medical intervention employed as the matching factor. Numerical examples demonstrate that the observed results of both conventional and alternative studies are mutually compatible with a broad range of true values of the association under study, and that there is no *a priori* basis for assuming that either set of results is closer to the true value. Nevertheless, estimates of exposure-specific disease detection rates can be used to evaluate the extent of bias present in conventional results.

## INTRODUCTION

WHEN A disease outcome is subject to serious underdiagnosis or underreporting, an epidemiologic study of the disease must address the possibility that the rate of diagnosis or reporting, and thus disease detection, varied between exposure categories. If either of these rates did vary with exposure, the observed magnitude of association between the exposure and the disease will be biased [1]. This 'detection bias' will occur whenever:

(1) The exposure induces a symptom or sign that leads to an increased diagnosis of the disease (in which case the bias has been termed 'unmasking bias') [1];
(2) Awareness of exposure history increases disease detection efforts in the exposed (in which case the bias has been termed 'diagnostic suspicion bias') [1]; or
(3) Awareness of exposure history increases reporting of disease among the exposed (in which case the bias may be termed 'disease reporting bias').

The above list, though not exhaustive, represents the general problem of bias due to differential detection of cases in the target population. A method of correcting for such bias would be extremely useful.

In the recent controversy concerning studies of replacement estrogen therapy and endometrial cancer, it has been suggested that it is possible to correct for bias due to exposure-induced differential case detection by matching or stratifying on symptoms or signs affected by the exposure, or by matching or stratifying on treatments for such symptoms or signs [2, 3]. This position has been seriously disputed [4, 5], partly on the logical grounds that the proposed correction will introduce a selection bias (which has been described as 'overmatching' between the case and control series) if the symptoms or signs are also influenced by the disease. Sackett [1] has commented that the net effect of the proposed correction 'has yet to be resolved.' In an attempt to contribute toward resolution of the logical aspects of the problem, we present an analysis of how detection bias can spuriously exaggerate an association in both follow-up and case-control studies, but also how attempts to remedy this problem by stratifying on factors that affect

TABLE 1. TARGET POPULATION FOR STUDIES OF EXOGENOUS ESTROGEN USE AND ENDOMETRIAL CANCER AFTER 2-YEAR PERIOD

|  | Estrogen users | Nonusers |
|---|---|---|
| Cases* | $2000\,r_E$ | $20{,}000\,r_{\bar{E}}$ |
| Noncases | $2000\,(1 - r_E)$ | $20{,}000\,(1 - r_{\bar{E}})$ |
| Totals | 4000 | 40,000 |
| Incidence rate | $r_E$ | $r_{\bar{E}}$ |

$$\text{Risk ratio} = \frac{r_E}{r_{\bar{E}}}$$

$$\text{Odds ratio} = \frac{r_E(1 - r_{\bar{E}})}{r_{\bar{E}}(1 - r_E)} \doteq \frac{r_E}{r_{\bar{E}}} \quad \begin{array}{l}\text{if the incidence in both groups is low (i.e.}\\ \text{both } r_E \text{ and } r_{\bar{E}} < 0.02)\end{array}$$

*Developed endometrial cancer over the study period.

detection can indeed introduce a severe deflationary bias in the observed association. These points have been noted before, but they have not been illustrated or analyzed in any generality. For purposes of brevity and clarity, the example below involves a number of important simplifications relative to the actual estrogen controversy, and we do not wish to imply that our analysis addresses all aspects of the controversy. Rather, we wish to derive certain methodologic points from the estrogen–endometrial cancer example that can be applied to the general situations listed previously. The actual numbers employed below were suggested by two recent conflicting reports [Refs 3 and 6].

## AN ANALYTIC EXAMPLE

For our analysis, let us suppose the following about a target population of 44,000 women age 50–64 on July 1, 1975, and observed from that date until June 30, 1977:

(1) On July 1, 1975, the women were free of endometrial cancer and had intact uteri, and
(2) Replacement estrogen therapy did in fact produce an increase in the 2-year incidence rate of endometrial cancer, from $r_{\bar{E}}$ in nonusers to $r_E$ in users.

Table 1 presents the distribution of exposure and the expected appearance of the target population at the end of the follow-up period on June 30, 1977, given an estrogen use prevalence of 9.1% (the actual estrogen use prevalence turns out to be irrelevant to the algebraic results).

In a real study situation, the observed results would undoubtedly differ from Table 1. Had we done a follow-up study on this population, we would have failed to detect all the cases of endometrial cancer, especially if we did not add any case-detection efforts to the background level of screening. Furthermore, the detection of cancer might have been increased by prior use of estrogen. To illustrate such a process, suppose that:

(1) The probability of having a hysterectomy or dilatation and curettage (D & C) is increased by estrogen use as well as by endometrial cancer;
(2) Endometrial cancer cases are detected whenever a hysterectomy or D & C is performed, and
(3) A hysterectomy is performed whenever endometrial cancer is detected.

Specifically, suppose the probability of having a hysterectomy or D & C during the study period was:

$P_{11}$ if one developed cancer and used estrogen;
$P_{12}$ if one developed cancer but did not use estrogen;
$P_{21}$ if one did not develop cancer but used estrogen, and
$P_{22}$ if one did not develop cancer and did not use estrogen.

TABLE 2. EXPECTED FOLLOW-UP STUDY DATA AFTER INCOMPLETE
DETECTION OF CASES

|  | Estrogen users | Nonusers |
|---|---|---|
| Cases | $2000\,r_E P_{11}$ | $20{,}000\,r_{\bar{E}} P_{12}$ |
| Apparent noncases | $2000\,(1 - r_E P_{11})$ | $20{,}000\,(1 - r_{\bar{E}} P_{12})$ |
| Totals | 4000 | 40,000 |
| Apparent incidence rate | $r_E P_{11}$ | $r_{\bar{E}} P_{12}$ |

Probability of detection if case uses estrogen $= P_{11}$.
Probability of detection if case doesn't use estrogen $= P_{12}$.

$$\text{Apparent risk ratio} = \frac{r_E P_{11}}{r_{\bar{E}} P_{12}} = \frac{r_E}{r_{\bar{E}}}\left(\frac{P_{11}}{P_{12}}\right).$$

Assumptions 2 and 3 imply that the set of detected cancer cases and hysterectomized cancer cases are identical. It follows that the probability of being detected for estrogen-using cases would be equal to $P_{11}$, and the probability of being detected for cases not using estrogen would be equal to $P_{12}$. Table 2 shows how the target population would be expected to appear in the study with the resulting incomplete case-detection: the misclassification due to differential case detection would have biased the risk ratio estimate by a factor equal to $P_{11}/P_{12}$. In this study example, we would expect $P_{11} > P_{12}$, giving a net positive bias to the study results.

If a conventional case-control study had been performed instead of a follow-up study, the observed odds ratio would have been biased to the same degree as the risk ratio in the follow-up study. To see this, suppose we took all the detected cases and some fraction $S_A$ of apparent noncases to serve as controls for our case-control study. Table 3 shows the expected results of such a study; the odds-ratio estimate is biased by a factor of approximately $P_{11}/P_{12}$. Thus, the biased case-detection would have produced an identical degree of bias in both the follow-up and conventional case-control estimates of association. Note, however, that according to the usual classification scheme for internal validity [7], the bias in the follow-up study and the bias in the conventional case-control study fall into different conceptual categories, despite their common ultimate cause. The bias in the follow-up study is a 'misclassification' bias: all cancer cases are subjects in the study, although some enter the analysis misclassified as noncases. The bias in the case-control study is a case selection bias, since the original incomplete detection manifests itself in the study as differential selection probabilities $P_{11}$ and $P_{12}$ for exposed and unexposed cases, respectively; this in turn results in the introduction of a 'selection bias factor' of $P_{11}/P_{12}$ into the conventional case-control study odds ratio.

Suppose now we did an 'alternative' case-control study on the population of Table 1, taking controls only from among those noncases who had a hysterectomy or D & C. Table 4 shows how this case-control study would be expected to appear, using all

TABLE 3. CONVENTIONAL CASE-CONTROL STUDY, USING A SAMPLE FROM ALL APPARENT
NONCASES AS THE CONTROL GROUP

|  | Estrogen users | Nonusers |
|---|---|---|
| Cases | $2000\,r_E P_{11}$ | $20{,}000\,r_{\bar{E}} P_{12}$ |
| Controls | $2000\,(1 - r_E P_{11})S_A$ | $20{,}000\,(1 - r_{\bar{E}} P_{12})S_A$ |

$S_A$ = sampling fraction for all apparent noncases.

$$\text{Apparent odds ratio} = \frac{r_E P_{11}(1 - r_{\bar{E}} P_{12})S_A}{r_{\bar{E}} P_{12}(1 - r_E P_{11})S_A}$$

$$\doteq \frac{r_E P_{11}}{r_{\bar{E}} P_{12}} = \frac{r_E}{r_{\bar{E}}}\left(\frac{P_{11}}{P_{12}}\right) \text{ if the incidences are low.}$$

TABLE 4. ALTERNATIVE CASE-CONTROL STUDY, USING AS THE CONTROL GROUP A SAMPLE FROM NONCASES WHO HAD HYSTERECTOMY OR D & C OVER THE STUDY PERIOD

|          | Estrogen users | Nonusers |
|----------|----------------|----------|
| Cases    | $2000\, r_E P_{11}$ | $20{,}000\, r_{\bar{E}} P_{12}$ |
| Controls | $2000\,(1 - r_E) P_{21} S_H$ | $20{,}000\,(1 - r_{\bar{E}}) P_{22} S_H$ |

$S_H$ = sampling fraction for noncases with hysterectomy or D & C.

$$\text{Apparent odds ratio} = \frac{r_E P_{11}(1 - r_{\bar{E}}) P_{22} S_H}{r_{\bar{E}} P_{12}(1 - r_E) P_{21} S_H}$$

$$= \frac{r_E (1 - r_{\bar{E}})}{r_{\bar{E}}(1 - r_E)} \left(\frac{P_{11}}{P_{12}}\right)\left(\frac{P_{22}}{P_{21}}\right)$$

$$\doteq \frac{r_E}{r_{\bar{E}}} \left(\frac{P_{11}}{P_{12}}\right)\left(\frac{P_{22}}{P_{21}}\right) \quad \text{if the incidences are low.}$$

detected cases and some fraction $S_H$ of the noncases who had hysterectomy or D & C. The selection bias induced in the control series by this design would bias the odds ratio by an additional factor of $P_{22} S_H / P_{21} S_H = P_{22}/P_{21}$. This bias results from matching (in the selection) on a variable (hysterectomy or D & C) which, because it can be caused by the disease under study, should not ordinarily be controlled in an analysis of the etiology of disease. (In an algebraically similar fashion, another bias term would be necessary if we conducted the alternative study with additional matching on the occurrence of uterine bleeding). Since we would expect in our situation that $P_{22} < P_{21}$, the control selection bias would be negative, cancelling out the case selection bias to some degree. Much of the real-life controversy revolves around the relative strengths of the case selection bias and the control selection bias. The total bias in the 'alternative' case-control study odds ratio may be written as

$$\frac{P_{11}}{P_{12}}\frac{P_{22}}{P_{21}},$$

which may be viewed as the 'selection bias factor' inherent in the 'alternative' study. The 'alternative' results will be unbiased if and only if this factor is equal to one. Note that, under the above scheme, the difference in results between the conventional study and the 'alternative' study will be entirely due to the control selection bias factor $P_{22}/P_{21}$. Thus, the studies will differ only in the proportion of estrogen users in their control groups; a similar feature may be noted in the original published comparison of a conventional and an alternative study (cf. Tables 2 & 4, Ref. 2).

Table 5 presents the results of fourteen numeric examples using the algebraic structure given above. For different values of $r_E$, $r_{\bar{E}}$, and the selection probabilities, it presents the true target population risk ratio, the apparent odds ratio expected from a conventional case-control study, and the apparent odds ratio expected from an 'alternative' case-control study. Some of the examples are unrealistic relative to the particulars of the estrogen–endometrial cancer situation, but have been included to provide a more complete illustration of the functional relationship between the quantities of interest. It is important to note, though, that there are many combinations of selection probabilities for which the conventional results are only slightly biased, others for which the alternative results are only slightly biased and still others for which neither result seems very good. Perhaps more importantly, we see that a wide variety of combinations of true risk ratios and selection probabilities produce results that closely parallel the pattern seen in the estrogen–endometrial cancer controversy. For example, the true risk ratios in the examples vary from 8.0 to 1.0, yet the conventional odds ratios are all within range of those reported by the conventional studies of exogenous estrogens and endometrial cancer (including Horwitz & Feinstein's), while the 'alternative'-study odds ratios are all

TABLE 5. RESULTS OF EXAMPLES USING VARIOUS INCIDENCES AND SELECTION PROB-
ABILITIES IN THE FORMULAE OF TABLES 1–4

| Target population | | Risk ratio | Selection probabilities | | | | Conventional case-control odds ratio | Alternative case-control odds ratio |
|---|---|---|---|---|---|---|---|---|
| $r_E$ | $r_{\bar{E}}$ | | $P_{11}$ | $P_{12}$ | $P_{21}S_H$ | $P_{22}S_H$ | | |
| 16* | 2* | 8.0 | 0.800 | 0.700 | 0.040 | 0.010 | 9.3 | 2.3 |
| 16 | 2 | 8.0 | 0.800 | 0.550 | 0.040 | 0.008 | 11.8 | 2.4 |
| 16 | 2 | 8.0 | 0.320 | 0.280 | 0.020 | 0.005 | 9.3 | 2.3 |
| 16 | 2 | 8.0 | 0.320 | 0.220 | 0.020 | 0.004 | 11.8 | 2.4 |
| 20 | 4 | 5.0 | 0.800 | 0.700 | 0.040 | 0.014 | 5.8 | 2.0 |
| 20 | 4 | 5.0 | 0.800 | 0.435 | 0.040 | 0.010 | 9.3 | 2.3 |
| 20 | 4 | 5.0 | 0.320 | 0.280 | 0.020 | 0.007 | 5.8 | 2.0 |
| 20 | 4 | 5.0 | 0.320 | 0.175 | 0.020 | 0.005 | 9.3 | 2.3 |
| 20 | 8 | 2.5 | 0.800 | 0.500 | 0.040 | 0.018 | 4.0 | 1.8 |
| 20 | 8 | 2.5 | 0.800 | 0.350 | 0.040 | 0.014 | 5.8 | 2.0 |
| 20 | 8 | 2.5 | 0.320 | 0.200 | 0.020 | 0.009 | 4.0 | 1.8 |
| 20 | 8 | 2.5 | 0.320 | 0.140 | 0.020 | 0.007 | 5.8 | 2.0 |
| 10 | 10 | 1.0 | 0.800 | 0.200 | 0.040 | 0.018 | 4.0 | 1.8 |
| 10 | 10 | 1.0 | 0.320 | 0.080 | 0.020 | 0.009 | 4.0 | 1.8 |

*Incidence per 1000 (2-yr period).

similar to those found by Horwitz & Feinstein when they used only patients that had received hysterectomy or D & C [3]. Although other factors have undoubtedly influenced the results of studies in this area [8], the phenomena illustrated in Tables 1–5 could be an important source of the controversy.

Many other examples that are compatible with the reports in the literature can be constructed using the formulae in Tables 1–4. Table 5 illustrates that use of the proposed 'alternative' design may successfully correct for the deleterious effects of the detection bias, or it may lead to a net selection bias with consequences more severe than the original distortion. Which situation holds depends heavily on the true values of the exposure-specific detection and selection probabilities and these are subject to much controversy.

(An illustration similar to Table 5 could be constructed for the follow-up study, with the 'alternative' study involving stratification on hysterectomy—D & C. The results from the conventional follow-up study would parallel those from the conventional case-control study, and the results from the 'alternative' follow-up study would parallel those from the 'alternative' case-control study.)

## DISCUSSION

Our discussion has considered one methodologic problem in the estrogen–endometrial cancer controversy in isolation. Nevertheless, the problem of detection bias is a general one, and we would think it inevitable that believers in the alternative study method would attempt to apply that method to other situations in which detection bias is felt to be a problem. Based on our illustration of the general structure of the problem and the 'alternative' method, it should be clear that attempts to correct detection bias by matching, restriction, or stratification based on special treatment or outcome categories must be validated by estimates of the true exposure-specific detection rates and the exposure-specific rates of the special treatment. The more unreliable or controversial these estimates are, the less likely it will be that any great confidence will develop in the results of either 'alternative' studies or conventional studies.

Note, however, that the bias in the 'alternative' odds ratio depends on four unknown quantities ($P_{11}, P_{12}, P_{21}$ and $P_{22}$), while the bias in the conventional odds ratio depends on only two ($P_{11}$ and $P_{12}$). When some ranges of estimates of the detection probabilities $P_{11}$ and $P_{12}$ are available, it will be possible to construct a series of tentative 'corrected' odds ratios. This can be done by multiplying the conventionally obtained odds ratio by a correction factor $P_{12}/P_{11}$, the inverse of the corresponding selection bias factor. (The same sort of correction can be performed for conventional follow-up studies as well,

using the risk ratio in place of the odds ratio.) Such an exercise, performed for various plausible values of $P_{11}$ and $P_{12}$, may illuminate how reliably the true association can be determined in light of incomplete case detection.* We believe that when, as is usually the case, very little is known about $P_{21}$ and $P_{22}$, the results of the foregoing sort of analysis will be more valid and reliable than results of 'alternative' studies.

In summary, we have discussed the problem of detection bias and found that it produces a selection bias in case-control studies and a misclassification bias in follow-up studies. In general, we cannot validly employ restriction, matching, or stratification to remove such bias. When we have empirical information on the detection probabilities, however, we will be able to estimate the degree of bias present in conventional results, and thus correct for it.

## REFERENCES

1.  Sackett DL: Bias in analytic research. **J Chron Dis** 32: 51–63, 1979
2.  Feinstein AR: Methodologic problems and standards in case-control research. **J Chron Dis** 32: 35–41, 1979
3.  Horwitz RI, Feinstein AR: Alternative analytic methods for case-control studies of estrogens and endometrial cancer. **N Engl J Med** 299: 1089–1094, 1978
4.  Hutchison GB, Rothman KJ: Correcting a bias? **N Engl J Med** 299: 1129–1130, 1978
5.  Vessey MP: Comment. **J Chron Dis** 32: 64–66, 1979
6.  Jick H, Watkins RN, Hunter JR, Dinan BJ, Madsen S, Rothman KJ, Walker AM: Replacement estrogens and endometrial cancer. **N Engl J Med** 300: 218–222, 1979
7.  Miettinen OS: Estimability and estimation in case-referent studies. **Am J Epidemiol** 103: 226–235, 1976
8.  Hulka BS, Hogue CJR, Greenberg BG: Methodologic issues in epidemiologic studies of endometrial cancer and exogenous estrogen. **Am J Epidemiol** 107: 267–276, 1978
9.  Hutchison GB: Case-control studies of estrogens and endometrial cancer (letter). **N Engl J Med** 300: 497, 1979

---

*Hutchison [9] has, in fact, examined data concerning detection rates of endometrial cancer in a manner similar to this and concluded that the case selection bias inherent in the conventional studies is severely overcorrected by the control selection bias inherent in the 'alternative' studies.

# Estrogens and Endometrial Cancer

## Responses to Arguments and Current Status of an Epidemiologic Controversy

RALPH I. HORWITZ, M.D.
ALVAN R. FEINSTEIN, M.D.
*New Haven, Connecticut*

Several years ago, we expressed doubts about the methods and interpretation of case-control studies in which replacement estrogen therapy was regarded as having significantly increased the risk of endometrial cancer in postmenopausal women [1]. We proposed instead that a strong association between estrogen use and endometrial cancer might arise if estrogen therapy led to the detection of symptomless endometrial cancers that might otherwise be undiscovered.

The "detection bias" could occur in either of two ways: passive or active. The passive effects could be produced by the increased medical surveillance and diagnostic examinations that would be carried out routinely in women receiving estrogen replacement therapy for menopausal symptoms. Since such women would be visiting a physician regularly, routine screening tests would be performed more often and would often be followed by the diagnostic evaluations that uncover otherwise asymptomatic disease. The active effects of detection bias would arise if bleeding, from the endometrial proliferation produced by estrogen therapy, evokes ad hoc diagnostic evaluations that detect co-existing but otherwise symptomless cancers.

The co-existence of proliferative and carcinomatous lesions has seldom been reported, because histopathologists who find a carcinoma in endometrial tissue usually concentrate on describing the cancer, and may not report details of any other lesion that is present. Nevertheless, in a recent special "blind" review of specimens [2], proliferative lesions were noted in 63 percent of women who had used estrogens and who had endometrial cancer. Since the source of bleeding is impossible to determine in such women, they might also have had symptomless cancers that were detected only because of the estrogen therapy.

If suitable adjustment or compensation is not made for the passive and active sources of "detection bias," a falsely increased association with antecedent therapy will be produced in any conventional retrospective study of a "case" group of women with endometrial cancer and a "control" group of women without it. To compensate for this bias, we proposed an alternative methodologic design [1] containing two strategies aimed at "equalizing" the intensity of diagnostic surveillance and disease detection in the case-control comparison. First, the case subjects and control subjects would all emerge from a group of women referred for the same pertinent diagnostic procedure: either dilatation and curettage or hysterectomy. Second, the data would be analyzed according to the clinical reason (with bleeding or without bleeding) for the diagnostic procedure. In a case-control study performed with this method, the over-all odds ratio (2.3; 95 percent confidence interval, 1.3 to

4.3), although still elevated, was substantially lower than the values reported in earlier studies. When analyzed according to the antecedent clinical manifestations, the odds ratio was 1.7 for patients with bleeding and 1.8 for patients without bleeding.

The type of problem we have cited has important scientific consequences because it is not unique to the relationship of estrogens and endometrial cancer. The unrecognized effects of detection bias can distort the odds ratios for many other etiologic associations that have been explored epidemiologically with the investigative case-control method. The detection bias problem will arise whenever exposure to an alleged etiologic agent may also lead to preferential diagnosis of a target disease that can go undetected or remain asymptomatic for protracted periods of time.

Perhaps because this methodologic problem casts doubt on many other epidemiologic relationships, a lively controversy has emerged over the cogency of the detection bias hypothesis. During this controversy, the epidemiologists who disputed either the "detection bias" hypothesis or the alternative methodologic design (or both) offered several distinctive arguments to support their contentions. Our purpose now is to respond to those arguments, to present some previously unpublished data, to discuss the impact of some new investigations, and to summarize the current state of the problem.

We shall note new evidence confirming the existence of symptomless endometrial cancer—a fundamental premise of the detection bias hypothesis. We shall also point out that the hypothesis can explain three sets of findings that allegedly support a causal relationship: the risk associated with the time of antecedent estrogen usage; the dose-response relationship; and the allegedly lower risk of estrogen-progesterone combination therapy. Finally, we shall indicate how the detection bias hypothesis is supported by the histologic stages in the estrogen-associated endometrial cancers and by additional evidence from an "alternative" case-control study.

## EXISTENCE OF SYMPTOMLESS UNDETECTED CASES OF ENDOMETRIAL CANCER

A fundamental premise of the detection bias hypothesis is that, in many cases, endometrial cancer is symptomless and can escape detection during the patient's lifetime. This premise has been challenged with assertions such as "Nearly all women with endometrial cancer will ultimately have the disease diagnosed" [3]. Although no specific evidence was offered to support this assertion, abundant data have been assembled to show that many other cancers are symptomless during life and first detected, if at all, at postmortem examination. For example, of the cancers found in 2,734 postmortem examinations at Boston City Hospital, 26 percent had been undetected previously [4]; the previously undetected proportions for

individual cancers were 86 percent for thyroid, 52 percent for prostate, 28 percent for colon, and 27 percent for lung.

Two additional studies have now provided specific evidence about the occurrence of symptomless endometrial cancer. In a review of 50,000 postmortem examinations performed at the Massachusetts General Hospital and Yale-New Haven Hospital from 1918 to 1978, the average annual rates of previously undetected endometrial cancer were found to be 31 and 22 per 10,000, respectively, at the two institutions [5]. These rates of "unreported" endometrial cancer are about five times higher than the "reported" rates with which endometrial cancer was found during life and cited during the same calendar period at the Connecticut Tumor Registry.

In a more recent study, Koss et al [6] used direct endometrial sampling as a diagnostic screening test in a cohort of 2,586 asymptomatic women. Of these women, 1,567 were reexamined one year later, and 187 were reexamined two years later. A total of 18 endometrial cancers were detected during the first examination, for an occurrence rate of 70 per 10,000 women, and three additional cancers were detected on reexamination, a rate of 17 per 10,000. These results of Koss et al confirm what was noted in the postmortem study and demonstrate that the occurrence rate of endometrial cancer in routinely examined asymptomatic women is about eight times higher than the rates reported by the state registries. The cancers found by Koss et al included both adenocarcinomas and adenoacanthomas, the histologic severity ranged from grade 1 to grade 3, and myometrial invasion extended from none to deep. In this group of women with uniform diagnostic surveillance, no association was noted between estrogen use and the risk of endometrial cancer. There were six endometrial cancers among 565 estrogen-users, and 15 cancers among 2,021 non-users. The associated odds ratio is only 1.3.

The data from Koss et al add conclusive new information that endometrial cancer is often symptomless, and that it can be detected during life if otherwise asymptomatic women undergo suitable diagnostic examination. The epidemiologic argument that endometrial cancer seldom escapes detection during life is thus refuted by empiric evidence showing that many cases are "silent," unsuspected, and first found, if at all, at postmortem examination or during a special diagnostic examination. Such cancers would readily be available for detection when the appropriate examination is evoked by estrogen-related surveillance or by estrogen-induced bleeding.

## RISK OF ENDOMETRIAL CANCER AND TIME OF ANTECEDENT ESTROGEN USAGE

A second major argument offered against the detection bias hypothesis is the contention that the risk of endometrial cancer was increased for both past and recent use of

estrogens. This result was found in four case-control studies [7–10] in which the risk of endometrial cancer was analyzed according to the time elapsed since the date of latest use of estrogens. An elevated risk was found for both the distant and recent use of estrogens in three studies [7–9]. In the fourth study [10], the risk was elevated only among patients who used estrogen at some time within the year before diagnosis. In all four studies, estrogens were regarded as having a causal role in endometrial cancer, but no attention was given to a distinctive time gradient in risk, and to the possible role of detection bias in producing the gradient.

If estrogens have a truly causal association with endometrial cancer, the risk should not differ substantially for recent users and for women who discontinued estrogens more than a year before diagnosis. Contrary to this expectation, the odds ratio in all four studies was substantially lower in women with distant use than in women with recent use, whose cancer was more likely to be detected because of estrogen-related bleeding. For example, in one study [7], the odds ratio was 6.5 if less than one year had elapsed since the latest use of estrogens, but it was only 1.9 for women who had last used estrogens more than a year before the diagnosis of endometrial cancer. The corresponding values for this same disparity were 3.7 versus 1.4 in a second study [8], and 8.7 versus 3.8 in a third study. In the fourth study [10], the odds ratio for women using estrogens for less than five years was 7.2 when the interval between last use and diagnosis was one year or less, but it was 1.0 when the interval was greater than one year.

The time gradient in risks for all four studies is entirely compatible with the detection bias hypothesis that women taking estrogens for postmenopausal symptoms are more likely to be examined, leading to detection of otherwise symptomless cancers. Whether long-discontinued use creates a small but distinctive increased risk of endometrial cancer is uncertain, since the odds ratio for "past users" of estrogen was only slightly elevated in three studies, and was not statistically significant in two [8,10].

### DOSE-RESPONSE RELATIONSHIP

Another argument offered for a causal relationship is the "dose-response" phenomenon in which an increasing odds ratio occurs with an increasing dose or duration of antecedent estrogen use. Such a relationship has been found in some, but not all [11], of the case-control studies. Although a dose-response relationship between estrogens and endometrial cancer is consistent with the causal hypothesis, the relationship is also consistent with the detection bias hypothesis. Larger doses or more prolonged use of estrogens could increase the incidence of uterine bleeding, thereby leading to the detection of otherwise symptomless cancers.

A different source of dose-response data is the rela-

tionship noted between changes in estrogen sales (or prescriptions) and changes in the incidence of endometrial cancer. When the incidence of endometrial cancer from 1969 to 1979 in the San Francisco region was compared with concomitant changes in estrogen sales and prescriptions [12], the results showed parallel trends. The incidence of endometrial cancer reported in that region for postmenopausal women rose from 25 per 10,000 in 1969 to a peak of 43 per 10,000 in 1975, and then dropped to 24 per 10,000 in 1979. At the same time, prescriptions of conjugated estrogens increased throughout the first half of the decade, peaked in 1976, and then decreased in 1977 and 1978.

The sharp drop in incidence with a concomitant reduction in estrogen usage is inconsistent with the accepted idea that carcinogens have a long latency period. The investigators therefore suggested that estrogens must act as a tumor promoter rather than a tumor initiator. Regardless of the role ascribed to estrogen therapy, the relationship can readily be explained by the detection bias hypothesis. As the number of estrogen prescriptions increased, the number of women with estrogen-related uterine bleeding would also increase, leading to the unmasking of many otherwise symptomless endometrial cancers.

### ESTROGENS AND STAGE OF ENDOMETRIAL CANCER

If the detection bias hypothesis is true—that otherwise asymptomatic cancers are particularly likely to be detected in estrogen-using women—then the estrogen-cancer association should be greatest for early-stage, low-grade tumors. This distinction has been confirmed in several studies showing that estrogen use is associated with earlier-stage, lower-grade tumors and with fewer instances of myometrial invasion [13–16]. In reviewing the evidence in their own study as well as the data in the literature, Kelsey et al [16] summarized the issue with the conclusion, ". . . the association is strongest with cancer of stage 1, probably indicating that women using estrogen replacement therapy have cancers diagnosed earlier than other women."

### RESULTS OF ESTROGEN-PROGESTERONE THERAPY

In several studies [17,18] reported since the beginning of the controversy, no substantial increase was found for the risk of endometrial cancer in postmenopausal women treated with estrogen-progesterone combinations. These results have helped remove the controversy from public dispute, since combination therapy has become widely accepted. Although some investigators have contended that the results also demonstrate the risk of "unopposed" estrogen therapy, the evidence is also consistent with the detection bias hypothesis. By helping reduce the occur-

rence of abnormal endometrial bleeding, the progesterone component of the therapy reduces the opportunity for symptomless cancers to receive increased diagnostic attention.

## VALIDITY OF THE "ALTERNATIVE" CASE-CONTROL STUDIES

Although detection bias is generally acknowledged as a problem, no attempts were made to deal with the problem in conventional case-control studies of the estrogen-endometrial cancer relationship. The case groups consisted of women with endometrial cancer, and the control groups were chosen with standard conventional methods that do not address the detection bias problem. When the problem was directly approached with an alternative sampling method [1], however, the validity of the new method was doubted because it allegedly produced a control group containing an excess of patients with estrogen-related disorders.

The new "alternative" method we proposed uses a "diagnostic sampling" strategy, in which both case and control groups are chosen from among patients who received the appropriate diagnostic examination (hysterectomy or dilatation and curettage). The argument against this method is that it might produce a spurious enrichment of estrogen-users in the control group, which would contain many patients undergoing testing to evaluate bleeding produced by benign, estrogen-induced endometrial proliferation. In response to these criticisms, which were offered by reviewers before our paper was published, we presented [1] two separate analyses of the data—one with the original control group and the other with subjects with proliferative disorders excluded from the control group. The odds ratio was 2.3 (95 percent confidence interval, 1.3 to 4.2) in the first analysis and 2.7 (95 percent confidence interval, 1.5 to 4.9) in the second. In the second set of analyses, proliferative disorders were removed from the control group as required by the reviewers, so that the manuscript could be accepted for publication. In the published paper, however, we stated (and we still believe) that such an analysis is improper. Since the *case* group is already biased by the effects of estrogen-induced bleeding, a statistical adjustment that removes endometrial proliferation from the control group but not from the case group is unfair.

At the time of our original publication, however, the occurrence of benign endometrial proliferation could not be determined in the case group, because the original pathologists had not described any lesions other than the cancer. Lacking the necessary data, we could not attempt to perform an unbiased adjustment by removing subjects with proliferative lesions from both the case group and the control group. We have now had the opportunity to perform the appropriate analysis, and we are herein reporting those results for the first time. In a "blind" histologic review of all available specimens, examined without clinical or other information, a gynecologic pathologist checked the available tissue slides for 233 of the 298 women who had constituted 112 case subjects and 121 control subjects in our original study. For each slide, the pathologist was asked to describe not only any neoplastic tissue but also any other endometrial abnormalities that might be present. As reported elsewhere [2], the results demonstrated that proliferative lesions often accompanied endometrial cancers and were found most frequently in patients with grade 1 (well-differentiated) cancers.

The results could also be used to perform the desired unbiased adjustment of our case-control data. The pathologist had identified hyperplastic or proliferative endometrium as a principal diagnosis in 90 of the 121 control subjects and as an additional diagnosis in 40 of the 112 case subjects. With these patients removed from the analysis, so that no proliferative lesions were contained in either the case or control group, the remaining patients showed the following results: case subjects with estrogen use, 19; case subjects without estrogen use, 53; control subjects with estrogen use, six; control subjects without estrogen use, 25. The odds ratio of 1.5 (95 percent confidence interval, 0.50 to 4.53) that emerges from these figures is considerably lower than the odds ratios of 2.7 and 2.1 obtained in our previous analyses.

## COMMENTS

After this paper was accepted for publication, a report was published of yet another case-control study that used conventional methods and that claimed to disprove the detection bias hypothesis [19]. However, this study was nearly identical to a previous report [7] that we have already discussed. The main distinction of the new report was in the expanded size of the case and control groups. As in the earlier report, the more recent paper described higher odds ratios for stage I or II cancer than for stage III or IV, and the odds ratio was considerably higher if less than one year had elapsed since the latest use of estrogens than if estrogens had last been used more than a year before the diagnosis of endometrial cancer. As noted in our earlier comments, both sets of observations are consistent with the detection bias hypothesis.

The authors of the recent case-control study [19] also suggested that the elevated odds ratios for advanced-stage disease and for distant estrogen use confirm the validity of the estrogen-endometrial cancer relationship. The basis for this contention is that estrogen-related detection bias should have its greatest effect on recent users of estrogen and early-stage disease. Although we agree with the theoretic basis for this argument, the evidence obtained in the recent study [19] is flawed by two important methodologic problems. First, much of the new investigation was carried out after the hypothesis that estrogens cause endometrial cancer had already received

enormous publicity in the lay press. Thus, it is possible that women with endometrial cancer were more likely to recall and report estrogen use than women without endometrial cancer. Second, the control subjects for this study were limited to women "who were admitted for conditions judged not to be related to estrogen use." If women who are likely to have used estrogens are deliberately and unilaterally excluded as potential control subjects, the prevalence of estrogen usage will be artificially low in the selected control group. This methodologic problem, which is known as *exclusion bias*, has already been shown [20] to be a source of major bias in the celebrated epidemiologic error, about 12 years ago, that incorrectly linked reserpine to the risk of breast cancer.

After considering all of the pertinent data, we have no reason to modify our conclusions in 1978 [1], "that the strength of the much-publicized association between estrogens and endometrial cancer has doubtlessly been exaggerated and needs re-evaluation." The evaluations performed in subsequent epidemiologic studies have not addressed the crux of the detection bias argument and have presented analyses that do not deal with the main point. Since the new sampling technique we proposed has not been accepted by conventional epidemiologists, and since their conventional sampling techniques do not deal with the cogent bias, the controversy currently stands at an impasse. It cannot be resolved until an accepted alternative sampling method is established to correct the bias in conventional case-control studies, or until data become available from cohort studies that arrange for equal diagnostic examinations in both the treated and untreated patients.

Perhaps the greatest scientific virtue of the current controversy is its demonstration of major problems in conventional epidemiologic case-control methods for dealing with the clinicopathologic realities of human cancers and other chronic diseases. Because so many instances of cancer and other chronic diseases can escape detection during life, the cases studied in epidemiologic research are only a part of the true occurrence of the disease. Since an agent that leads to increased diagnostic testing will also be suspected of causing the increased occurrence, the problems of detection bias must be recognized and suitably managed.

We believe the adjustments proposed in our new sampling method can provide a satisfactory solution for this problem in case-control studies, particularly when data are analytically stratified for the clinical stimuli that evoked diagnostic testing. This new method of compensating for bias in both the case and control groups seems more attractive than the current epidemiologic approach, in which the overt bias in the case group is ignored and left unadjusted. If the new sampling method is not acceptable, an important challenge for investigators concerned with scientific progress in epidemiology is to develop a satisfactory alternative.

## REFERENCES

1. Horwitz RI, Feinstein AR: Alternative analytic methods for case-control studies of estrogens and endometrial cancer. N Engl J Med 1978; 299: 1089–1094.
2. Horwitz RI, Feinstein AR, Vidone RA, Sommers SC, Robboy SJ: Histopathologic distinctions in the relationship of estrogens and endometrial cancer. JAMA 1981; 246: 1425–1427.
3. Hutchison GB, Rothman KJ: Correcting a bias? N Engl J Med 1978; 299: 1129–1130.
4. Bauer FW, Robbins SL: An autopsy study of cancer patients. JAMA 1972; 221: 1471–1474.
5. Horwitz RI, Feinstein AR, Horwitz SM, Robboy SJ: Necropsy diagnosis of endometrial cancer and detection-bias in case-control studies. Lancet 1981; II: 66–68.
6. Koss LG, Schreiber K, Oberlander SG, Moussouris HF, Lesser M: Detection of endometrial carcinoma and hyperplasia in asymptomatic women. Obstet Gynecol 1984; 64: 1–11.
7. Shapiro S, Kaufman DW, Slone D, et al: Recent and past use of conjugated estrogens in relation to adenocarcinoma of the endometrium. N Engl J Med 1980; 303: 485–488.
8. La Vecchia C, Franceschi S, Gallus G, et al: Oestrogens and obesity as risk factors for endometrial cancer in Italy. Int J Epidemiol 1982; 11: 120–126.
9. Weiss NS, Szekely DR, English DR, Schweid AI: Endometrial cancer in relation to patterns of menopausal estrogen use. JAMA 1979; 242: 261–264.
10. Stavraky KM, Collins JA, Donner A, Wells GA: A comparison of estrogen use by women with endometrial cancer, gynecologic disorders, and other illnesses. Am J Obstet Gynecol 1981; 141: 547–555.
11. Hulka BS, Fowler WC, Kaufman DG, et al: Estrogen and endometrial cancer: cases and two control groups from North Carolina. Am J Obstet Gynecol 1980; 137: 92–101.
12. Austin DF, Roe KM: The decreasing incidence of endometrial cancer: public health implications. Am J Public Health 1982; 72: 65–68.
13. Mack TM, Pike MC, Henderson BE, et al: Estrogens and endometrial cancer in a retirement community. N Engl J Med 1976; 294: 1262–1267.
14. McDonald TW, Annegers JF, O'Fallon WM, et al: Exogenous estrogen and endometrial carcinoma: case-control and incidence study. Am J Obstet Gynecol 1977; 127: 572–580.
15. Antunes CM, Stolley PD, Rosenshein NB, et al: Endometrial cancer and estrogen use. Report of a large case-control study. N Engl J Med 1979; 300: 9–13.
16. Kelsey JL, LiVolsi VA, Holford TR, et al: A case-control study of cancer of the endometrium. Am J Epidemiol 1982; 116: 333–342.
17. Gambrell RD: Postmenopausal bleeding. Clin Obstet Gynecol 1977; 4: 129–142.
18. Hammond CB, Jelovsek FR, Lee KL, Creasman WT, Parker RT: Effects of long-term estrogen replacement therapy. II. Neoplasia. Am J Obstet Gynecol 1979; 133: 537–547.
19. Shapiro S, Kelly JP, Rosenberg L, et al: Risk of localized and widespread endometrial cancer in relation to recent and discontinued use of conjugated estrogens. N Engl J Med 1985; 313: 969–972.
20. Horwitz RI, Feinstein AR: Exclusion bias and the false relationship of reserpine and breast cancer. Arch Intern Med 1985; 145: 1873–1875.

# Data, Design, and Background Knowledge in Etiologic Inference

*James M. Robins*

---

I use two examples to demonstrate that an appropriate etiologic analysis of an epidemiologic study depends as much on study design and background subject-matter knowledge as on the data. The demonstration is facilitated by the use of causal graphs. (Epidemiology 2001;11:313–320)

---

Greenland *et al*[1] discussed the use of causal graphs in epidemiologic research. A limitation of that paper was that it was lacking concrete examples designed to help the reader see how to take one's knowledge of study design, temporal ordering, basic biology, and epidemiologic principles to construct an appropriate causal graph. Here I present two epidemiologic thought experiments that make the point that the choice of an appropriate etiologic analysis depends as much on the design of the study and background subject-matter knowledge as on the data.

Specifically, in the first, I provide a single hypothetical dataset and three differing study designs, each of which plausibly could have given rise to the data. I show that the appropriate etiologic analysis differs with the design. In the second, I revisit a well-known epidemiologic controversy from the late 1970s. Horowitz and Feinstein[2] proposed that the strong association between postmenopausal estrogens and endometrial cancer seen in many epidemiologic studies might be wholly attributable to diagnostic bias. Others disagreed.[3–5] Part of the discussion centered on the issue of whether it was appropriate to stratify on vaginal bleeding, the purported cause of the diagnostic bias in the analysis. The goal here is to show, using causal graphs, that the answer depends on underlying assumptions about the relevant biological mechanisms.

## 1. Thought Experiment 1

Consider the data given in Table 1. $E$ is a correctly classified exposure of interest whose net causal effect on a disease outcome $D$ I would like to ascertain. $E^*$ is a misclassified version of $E$. We are interested in the effect of $E$ on $D$. Data on $E$, $E^*$, and $D$ are available on all study subjects. Sampling variability can be ignored. I will now describe the designs of three different studies. For each study, the data are the same. Only the designs are different. I wish to answer the following questions for each of the studies: Can one say whether exposure has an adverse, protective, or no causal effect on the outcome? What association measure is most likely to have a causal interpretation?

As a guide, I present some candidate association measures. In Table 2, I calculate the exposure-disease odds ratio $OR_{ED} = 1.73$. I can also calculate the conditional $ED$ odds ratio within strata of $E^*$, that is, $OR_{ED|E^* = 1} = OR_{ED|E^* = 0} = 3$. Similarly, I calculate that $OR_{E^*D} = 0.5$ and $OR_{E^*D|E = 1} = OR_{E^*D|E = 0} = 0.3$. I will report all associations on an odds ratio scale. This choice is dictated by the fact that in the case-control study described below, the only estimable population association measures are odds ratios.

CASE-CONTROL STUDY

Suppose the data arose from a case-control study of the effect of a particular nonsteroidal anti-inflammatory drug ($E$) on a congenital defect ($D$) that arises in the second trimester. Cases ($D = 1$) are infants with the congenital defect. Controls ($D = 0$) are infants without the defect. The control sampling fraction is unknown. The data $E^*$ were obtained 1 month postpartum by maternal self-report. The data $E$ were obtained from comprehensive accurate medical records of first trimester medications. All relevant preconception confounders and other drug exposures were controlled by stratification. The data in Table 1 are taken from a particular

From the Department of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115.

Address correspondence to: James M. Robins, Harvard School of Public Health, Boston, MA 02115.

**TABLE 1.   Data from a Hypothetical Study**

|        | D = 1 | | | | D = 0 | |
|--------|---------|---------|--------|--------|---------|---------|
|        | E* = 1 | E* = 0 | | | E* = 1 | E* = 0 |
| E = 1  | 180     | 200     | E = 1  | 600    | 200     |
| E = 0  | 20      | 200     | E = 0  | 200    | 600     |

stratum. Note that misclassification is differential, given that $OR_{E^*D|E=1} = OR_{E^*D|E=0} = 0.3 \neq 1$.

PROSPECTIVE COHORT STUDY

Suppose the data were obtained from a follow-up study of total mortality ($D$) in a cohort of short-term healthy 25-year-old uranium miners, all of whom only worked underground in 1967 for 6 months. The follow-up is complete through 1997. Suppose, for simplicity, there is a threshold pulmonary dose below which exposure to radon is known to have no effect on mortality. Let $E = 1$ ($E = 0$) denote above-threshold (below-threshold) exposure to radon as measured by lung dosimetry. Each miner was also assigned an estimated radon exposure $E^*$ on the basis of the air level of radon in his mine. Let $E^* = 1$ ($E^* = 0$) denote an estimate above (below) threshold radon exposure. The assignment of miners to particular mines was unrelated to lifestyle, demographic, or medical risk factors. A subject's actual exposure $E$ depends both on the level of radon in the mine and on the demands of the subject's job, such as the required amount of physical exertion and thus minute ventilation. Finally, it is known that 6 months of physical exertion at age 25 has no independent effect on later mortality.

RANDOMIZED CLINICAL TRIAL

Suppose the data were obtained from a randomized follow-up study of the effect of low-fat diet on death ($D$) over a 15-year follow-up period. Study subjects were randomly assigned to either a low-fat diet, educational, and motivational intervention arm ($E^* = 1$) or to a standard care arm ($E^* = 0$). Investigators were able to obtain accurate measures of the actual diet followed by the study subjects: $E = 1$ if a study subject followed a low-fat diet, and $E = 0$ otherwise. Assume $E^*$ has no direct effect on death ($D$) except through its effect on actual fat consumption $E$.

CAUSAL CONTRASTS

To determine which association measure is most likely causal, I need a formal definition of causal effects. Causal effects are best expressed in terms of counterfactual variables. Let the variable $D(1)$ denote a subject's

**TABLE 2.   Crude Data from a Hypothetical Study**

|       | E = 1 | E = 0 |
|-------|-------|-------|
| D = 1 | 380   | 220   |
| D = 0 | 800   | 800   |
|       | OR = 1.73 | |

outcome if exposed and $D(0)$ denote a subject's outcome if unexposed. For a given subject, the causal effect of treatment, measured on a difference scale, is $D(1) - D(0)$. If a subject is exposed ($E = 1$), the subject's observed outcome $D$ equals $D(1)$, and $D(0)$ is unobserved. If $E = 0$, $D$ equals $D(0)$, and $D(1)$ is unobserved. Let $pr[D(1) = 1]$ and $pr[D(0) = 1]$, respectively, be the probability that $D(1)$ is equal to 1 and $D(0)$ is equal to 1, where probabilities refer to proportions in a large, possibly hypothetical, source population. Then, the exposure-disease causal odds ratios is $OR_{causal,ED} = \{pr[D(1) = 1]/pr[D(1) = 0]\}/\{pr[D(0) = 1]/pr[D(0) = 0]\} = pr[D(1) = 1]pr[D(0) = 0]/\{pr[D(1) = 0]pr[D(0) = 1]\}$. For any variable $Z$, the exposure-disease causal odds ratio among the subset of subjects with $Z$ being $z$ is $OR_{causal,ED|Z = z} = pr[D(1) = 1|Z = z]pr[D(0) = 0|Z = z]/\{pr[D(1) = 0|Z = z]pr[D(0) = 1|Z = z]\}$

.

ANSWERS

In this subsection, we provide the appropriate answers. The justification for these answers is given after I have reviewed causal graphs below. In the case-control study, exposure is likely harmful and the best parameter choice is the crude odds ratio $OR_{DE} = 1.73$. The other measures are biased. In particular, the conditional odds ratio $OR_{ED|E^*} = 3$ is biased in the sense that it fails to equal the causal effect $OR_{causal,ED|E^*}$ of exposure on disease among subjects within a particular stratum of $E^*$.

In the prospective cohort study, exposure is likely beneficial, and the best parameter choice is the conditional odds ratio $OR_{DE|E^*} = 3$. In the randomized trial, exposure is likely beneficial, and the best parameter choice may be the crude $E^*D$ association $OR_{E^*D} = 0.5$, although it is likely that this association underestimates the true benefit of exposure. In this case, both the crude association $OR_{ED} = 1.73$, and the conditional association $OR_{ED|E^*} = 3$ are biased estimates of the causal effect of $E$ on $D$. These answers clearly show that the appropriate statistical analysis depends on the design.
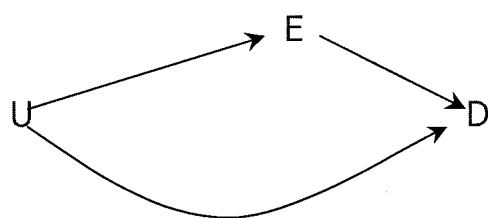
CAUSAL GRAPHS

To justify the answers, we review causal directed acyclic graphs (DAGs) as discussed by Pearl and Verma,[6] Spirtes et al,[7] Pearl,[8] Pearl and Robins,[9] and Greenland et al.[1]
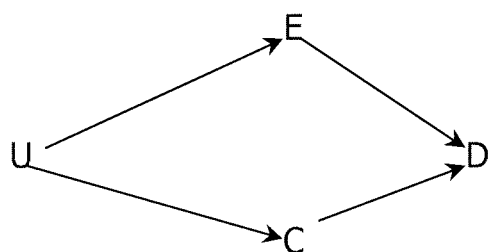
A causal graph is a directed acyclic graph (DAG) in which the vertices (nodes) of the graph represent variables; the directed edges (arrows) represent direct causal relations between variables; and there are no directed cycles, because no variable can cause itself (Figure 1). For a DAG to be causal, the variables represented on the graph must include the measured variables and additional unmeasured variables, such that if any two variables on the graph have a cause in common, that common cause is itself included as a variable on the graph. For example, in DAG 1, $E$ and $D$ are the measured variables. $U$ represents all unmeasured common causes of $E$ and $D$.
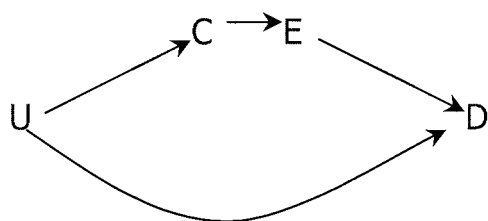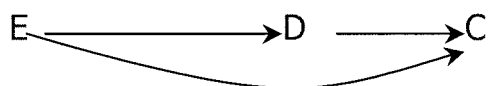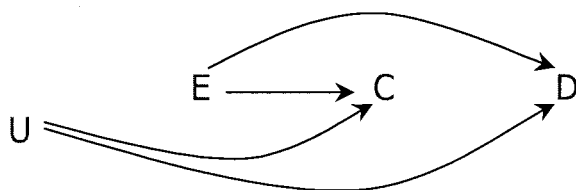
FIGURE 1.   Directed acyclic graphs for the sections Causal Graphs and Using Causal Graphs to Check for Confounding, in Thought Experiment 1. D = disease; E = exposure; U = an unmeasured potential confounder; C = a measured potential confounder.

A direct cause of a variable V on the graph is called a parent of V, and V is called the parent's child. The variables that can be reached starting from V by following a sequence of directed arrows pointing away from V are the descendants of V. The ancestors of V are those variables with V as a descendant. We will assume V is a cause of each of its descendants but a direct cause only of its children (where direct is always relative to the other variables on the DAG). Thus, V is caused by all its ancestors, but only its parents are direct causes.

Consider DAG 3. C is a cause of D through the pathway $C \rightarrow E \rightarrow D$ but is not a direct cause. The intuition is that intervening and manipulating C will affect E, and the change in E will in turn affect D. If we intervene and set each subject's value of E to the same level (say, exposed), however, then additionally manipulating C will no longer affect the distribution of E and thus that of D. Hence, we say that C has no direct effect on D when controlling for (in the sense of intervening and physically controlling or setting) the variable E. Note, however, that U is both a direct cause of D and an indirect cause through the causal pathway $U \rightarrow C \rightarrow E \rightarrow D$.

Our causal DAGs are of no use unless we make some assumption linking the causal structure represented by the DAG to the statistical data obtained in an epidemiologic study. Recall that if a set of variables X is statistically independent of (that is, unassociated with) another set of variables Y conditional on a third set of variables Z, then, within joint strata defined by the variables in Z, any variable in X is unassociated with any variable in Y. For example, suppose all variables are dichotomous and the set Z consists of the two variables $Z_1$ and $Z_2$. Then conditional independence implies that the odds ratio between any variable in X and any variable in Y is 1 within each of the $4 = 2^2$ strata of Z: $(Z_1, Z_2) = (0,0)$, $(Z_1, Z_2) = (0,1)$, $(Z_1, Z_2) = (1,0)$, and $(Z_1, Z_2) = (1,1)$. The following so-called causal Markov assumption (CMA) links the causal structure of the DAG with various statistical independencies.

CAUSAL MARKOV ASSUMPTION

On a causal graph, any variable that is not caused by a given variable V will be independent of V conditional on the direct causes of V.

Recall that the descendants of a variable V are those variables causally affected by V and that the parents of V are the variables that directly cause V. It follows that the CMA is the assumption that V is independent of its nondescendants conditional on its parents.

*Example 1*

On DAG 1, suppose that the arrow from E to D were absent so that neither E nor D causes the other. U represents all unmeasured common causes of E and D. Because U is the only parent of D, and E is not a descendant of D, the CMA implies that E and D are unassociated (that is, independent) given (that is, within strata of) U. That is, two variables that are not

causally related are independent conditional on their common causes.

It turns out that the CMA logically implies additional statistical independencies. Specifically, CMA implies that a set of variables $X$ is conditionally independent of another set of variables $Y$ given a third set of variables $Z$, if $X$ is "d-separated" from $Y$ given $Z$ on the graph, where "d-separation,"[10,11] described below, is a statement about the topology of the graph.

To describe d-separation, we first need to define the "moralized ancestral" graph generated by the variables in $X$, $Y$, and $Z$. In the following, a path between 2 variables is any unbroken sequence of edges (regardless of the directions of any arrows) connecting the two nodes.

The moralized ancestral graph generated by the variables in $X$, $Y$, and $Z$ is formed as follows:[11]

1. First, remove from the DAG all nodes (and corresponding edges) except those contained in the sets $X$, $Y$, and $Z$ and their ancestors.
2. Next, connect by an undirected edge every pair of nodes that both (a) share a common child and (b) are not already connected by a directed edge.

The graph is referred to as "moralized," because, in step 2, we marry (connect) all unmarried (unconnected) parents of a common child.

$X$ is d-separated from $Y$ given $Z$ if and only if on the moralized ancestral graph generated by $X$, $Y$, and $Z$, any path from a variable in $X$ to a variable in $Y$ intercepts (that is, goes through) some node in $Z$.

If $X$ and $Y$ are not d-separated given $Z$, we say they are d-connected given $Z$. Note that if there are no paths connecting variables in $X$ to variables in $Y$ on the moralized ancestral graph, then $X$ and $Y$ are d-separated.

To check for a crude (that is, unconditional or marginal) association, we make $Z$ the empty set. It is crucial that one perform step 1 before step 2 when forming the moralized ancestral graph.

*Example 2*

Consider causal graph DAG 4. Note that $E$ and $C$ can have no common cause, because, if they did, that common cause would have to be represented on the graph. Now, assume there is no arrow from $E$ to $D$ so $E$ does not cause $D$. Then, $E$ and $D$ are marginally independent (that is, have a crude odds ratio of 1). This statement follows either from the CMA or from the fact that $E$ is d-separated from $D$ given $Z$ equal to the empty set. Specifically, in step 1 of the moralized graph algorithm, $C$ and the arrows pointing into it are removed from the graph so that in step 2, $E$ and $D$ have no children. Thus, there is no path linking $E$ and $D$ on the moralized *ancestral* graph, so they are d-separated. This example tells us that two causally unrelated variables without a common cause are marginally unassociated (that is, independent).

In contrast, $E$ and $D$ are not d-separated given $C$. To see this, note that upon identifying $Z$ as $C$, $C$ is no longer removed in step 1 of the algorithm. Hence, in step 2, $E$ and $D$ have to be connected by an edge because they

have a common child $C$. Hence, $E$ and $D$ are d-connected given $C$, because there is a direct edge between them in the moralized *ancestral* graph that does not intercept $C$. This example tells us that if we condition on a common effect $C$ of two independent causes $E$ and $D$, we "usually" render those causes conditionally dependent. For instance, if we know a subject has the outcome $C$ (that is, we condition on that fact) but does not have the disease $D$, then it usually becomes more likely that the subject has the exposure $E$ (because we require some explanation for his or her having $C$). That is, among subjects with the outcome $C$, $E$ and $D$ are "usually" negatively associated (have an odds ratio less than 1).

The reason we included the word "usually" in the above is that although CMA allows one to deduce that d-separation implies statistical independence, it does not allow one to deduce that d-connection implies statistical dependence. However, d-connected variables will generally be independent only if there is an exact balancing of positive and negative causal effects. For example, in DAG 3, $U$ is a parent of and thus not d-separated from $D$. Yet if the direct effect of $U$ on $D$ is equal in magnitude but opposite in direction to the effect of $U$ on $D$ mediated through the variables $C$ and $E$, then $U$ and $D$ would be independent, even though they are d-connected. Because such precise fortuitous balancing of effects is highly unlikely to occur, we shall henceforth assume that d-connected variables are associated.[6,7]

USING CAUSAL GRAPHS TO CHECK FOR CONFOUNDING

We can use causal graphs and d-connection to check for confounding as follows. First, suppose, as on DAGs 1-5, $E$ is not an indirect cause of $D$. We begin by pretending that we know that exposure has no causal effect on the outcome $D$ by removing just those arrows pointing out of exposure necessary to make $D$ a nondescendant of $E$. If, under this causal null hypothesis, (1) $E$ and $D$ are still associated (that is, d-connected), then obviously the association does not reflect causation, and we say that the $E$-$D$ association is confounded, and (2) if $E$ and $D$ are associated (d-connected) conditional on (that is, within levels) of $Z$, we say there is confounding for the $E$-$D$ association within levels (strata) of $Z$. For example, the existence of an unmeasured common cause $U$ of $E$ and $D$ as in DAG 1 will make $E$ and $D$ associated under the causal null (because $E$ and $D$ will be d-connected). If data on $U$ have not been recorded for data analysis, confounding is intractable and we cannot identify the causal effect of $E$ on $D$. If data on $U$ are available, however, the conditional associations $OR_{ED \mid U}$ are unconfounded and will represent the causal effect of $E$ on $D$ within strata of $U$, that is, $OR_{ED \mid U} = OR_{causal, ED \mid U}$ at each level of $U$. This relation reflects the fact that under the causal null hypothesis of no arrow from $E$ to $D$, I showed in Example 1 that $E$ and $D$ are independent (d-separated) given $U$. Furthermore, suppose, as has been assumed, that we have not conditioned on a variable lying on a casual pathway from $E$ to $D$; then it is a general result that if $E$ is a time-independent exposure

and $E$ and $D$ are (conditionally) independent under the causal null, then, under the causal alternative, the (conditional) association between $E$ and $D$ will reflect the (conditional) causal effect of $E$ on $D$.[1]

Next we consider graphs 2 and 3, in which the variable $C$ has been measured. Thus, in DAG 3, $U$ remains an unmeasured common cause of $E$ and $D$, although it is not a direct cause of $E$. It follows that, in both DAGs 2 and 3, the marginal association $OR_{ED}$ is confounded, because $E$ and $D$ will be marginally associated (that is, d-connected) even under the causal null. However, the unmeasured variable $U$ will not function as a common cause of $E$ and $D$ within strata of $C$ because under the causal null $E$ and $D$ are d-separated given $C$. Thus, stratifying on $C$ in the analysis will control confounding and $OR_{ED|C} = 1$ and $OR_{ED|C} = 0$ will represent the causal effect of $E$ on $D$ within strata of $C$. The variable $U$ in DAGs 2 and 3 is referred to as a causal confounder, because it is a common cause of $E$ and $D$. DAG 3 shows that we can control confounding due to a causal confounder $U$ by stratifying on a variable $C$ that itself is not a cause of $D$. Note, however, that $C$ is an independent (but noncausal) risk factor for $D$ in the sense that $C$ and $D$ are associated (d-connected) within strata of $E$.

Consider next DAG 4. There are no unmeasured common causes of $E$ and $D$. As discussed in Example 2 above, under the causal null hypothesis of no arrow from $E$ to $D$, $E$ and $D$ will be independent. It follows that the marginal association $OR_{ED}$ is unconfounded and represents the causal effect of $E$ on $D$. In contrast, the conditional association $OR_{ED|C}$ will be confounded and thus will not be equal to the causal effect of $E$ on $D$ within strata of $C$, because we showed in Example 2 that, under the causal null of no arrow from $E$ to $D$, $E$ and $D$ will be conditionally associated within strata of $C$. This example shows that conditioning on a common effect $C$ of $E$ and $D$ introduces confounding within levels of $C$. This example also shows why, to check for confounding, we remove from the graph just those arrows necessary for the outcome $D$ to be a nondescendant of $E$; had we removed all arrows pointing out of $E$ (including that into $C$) we would not have recognized that conditioning on $C$ would cause confounding within levels of $C$.

An extension of this last example provides an explanation of the well-known adage that one must not adjust for variables affected by treatment. To see why, consider DAG 5, in which the exposure $E$ has a direct causal effect on $C$, and $C$ and $D$ have an unmeasured common cause $U$. Under the causal null with the arrow from $E$ to $D$ removed, $E$ and $D$ will be d-separated and thus unassociated. Thus, the marginal association $OR_{ED}$ will be unconfounded and represent causation. Nevertheless, the conditional associations $OR_{ED|C = 1}$ and $OR_{ED|C = 0}$ will be confounded and thus biased for the conditional causal effect within levels of $C$. This situation reflects the fact that, under the causal null, $E$ and $U$ will be associated once we condition on their common effect $C$. Thus, because $U$ itself is correlated with $D$, $E$ and $D$ will be conditionally associated (that is, d-connected) within

levels of $C$. Note the fact that the analysis stratifying on $C$ was confounded even under the causal null proves that adjusting for a variable $C$ affected by treatment can lead to confounding and bias even when $C$ is not an intermediate variable on any causal pathway from exposure to disease.

Finally, suppose, on a causal graph, $E$ is an indirect cause of $D$ through a directed path $E \rightarrow C \rightarrow D$ so that, among those with $C=c$, the net (overall) effect $OR_{causal,ED|C=c}$ differs from the direct effect of $E$ on $D$. We can still graphically test for confounding as described above, except that, now, regardless of our test results, we must never conclude that $OR_{ED|C = c}$ equals $OR_{causal,ED|C=c}$ for any variable $C$ on a causal pathway from $E$ to $D$.

With this background, we are ready to justify the answers given above.

### JUSTIFICATIONS OF ANSWERS
#### Case-Control Study

We first argue that the causal graph representing our case-control study is DAG 6 (Figure 2). By assumption, we need not worry about unmeasured preconception confounders. Furthermore, we know that if there is an arrow between $E$ and $D$, it must go from $E$ to $D$ because the medical records were created in the first trimester, before the development of the second trimester congenital defect. Also, actually taking a medicine will be a cause of a woman reporting that she took a medicine; hence, the arrow from $E$ to $E^*$. Finally, because a woman's self-report, $E^*$, is obtained after her child's birth, the defect $D$ will be a cause of $E^*$, if, as is likely, mothers whose children have a congenital defect are more prone to recall their medications than are other mothers. We can use the data to confirm the existence of an arrow from $D$ to $E^*$, because otherwise $E^*$ and $D$ would be independent (d-separated) within levels of $E$. But one can check from Table 1 that among subjects with $E = 1$, $D$ and $E^*$ are associated ($OR_{DE^*|E = 1} = 0.3$), so misclassification is differential. DAG 6 is isomorphic to DAG 4 with $E^*$ playing the role of $C$. Thus, as in DAG 4, we conclude that the marginal association $OR_{ED} = 1.7$ is causal but the conditional association $OR_{ED|E^*} = 3$ will differ from the conditional causal effect $OR_{causal,ED|E^*}$. Mistakenly interpreting $OR_{ED|E^*} = 3$ as causal could in principle lead to poor public health decisions, as would occur if a cost-benefit analysis determines that a conditional causal odds ratio of 2.9 is the cutoff point above which the risks of congenital malformation outweigh the benefits to the mother of treatment with $E$.

Finally, a possibility that we have not considered is that those mothers who develop, say, a subclinical infection in the first trimester are at increased risk both of a second trimester congenital malformation and of worsening arthritis, which they may then treat with the drug $E$. In that case, we would need to add to our causal graph an unmeasured common cause $U$ (subclinical infection) of both $E$ and $D$ that represents subclinical first trimester infection, in which case even $OR_{ED}$ would be confounded.

## PROSPECTIVE COHORT STUDY

In the prospective cohort study, sufficient information is given so that we know there is no confounding by unmeasured pre-employment factors. Yet, as noted above, $E^*$ is associated with $D$ given $E$. Now clearly $E^*$, which is a measure of the air-level of radon in mines, cannot itself directly cause death other than through its effect on a subject's actual pulmonary radon exposure $E$, so that there cannot be a direct arrow from $E^*$ to $D$. Nevertheless, because $E^*$ was measured before death, $D$ cannot be a cause of $E^*$ either. Furthermore, we are given that there is no arrow from any unmeasured confounder into $E$, because, although physical exertion is a cause of the pulmonary dose $E$, it is not a cause of $D$. The most reasonable explanation for these facts is that $E^*$ is a surrogate for some other unmeasured adverse causal exposure in the mine (say silica). Thus, we might consider the causal graph shown in DAG 7. In this figure, $MINE$ represents the particular mine in which the subject works. It is plausible that mines with high levels of radon may have low levels of silica-bearing rock (because silica-bearing rock is not radioactive). Therefore, $E^*$ and $SILICA$ will be negatively correlated. If DAG 7 is the true causal graph (with $MINE$ and $SILICA$ being unmeasured variables), then under the causal null hypothesis in which the arrow from $E$ to $D$ is removed, $E$ and $D$ will still remain correlated because $MINE$ is an unmeasured common cause of $E$ and $D$ but, by d-separation, $E$ and $D$ will be independent conditional on $E^*$. Thus, $OR_{ED}$ is confounded; however, $OR_{DE|E^*} = 3$ equals the causal effect $OR_{causal,DE|E^*}$ of exposure on disease within strata of $E^*0.3$. In contrast, the conditional association $OR_{E^*D|E} = 0.3$ represents not a protective effect of $E^*$ on $D$ but rather the negative correlation between $E^*$ and $SILICA$ conjoined with the adverse causal effect of $SILICA$ on $D$. DAG 7, however, probably does not tell the whole story. One would expect that physical exertion is a direct cause of a worker's actual (unrecorded) silica dose. Thus, physical exertion is an unmeasured common cause of $E$ and $D$, even when we condition on $E^*$, precluding unbiased estimation of the causal effect of $E$ on $D$.

## RANDOMIZED CLINICAL TRIAL

The study is a typical randomized trial with noncompliance and is represented by the causal graph in DAG 8.[12] Because $E^*$ was randomly assigned, it has no arrows into it. Given assignment, however, both the decision to comply and the outcome $D$ may well depend on underlying health status $U$. $E^*$ has no direct arrow to $D$, because, by assumption, $E^*$ causally influences $D$ only through its effect on $E$. We observe that under the causal null in which the arrow from $E$ to $D$ is removed, $E$ and $D$ will be associated (d-connected) owing to their common cause $U$ both marginally and within levels of $E^*$. Hence, both $OR_{ED}$ and $OR_{ED|E^*}$ are confounded and have no causal interpretation. Under this causal null, however, $E^*$ and $D$ will be independent, because they have no unmeasured common cause. Hence, we can test
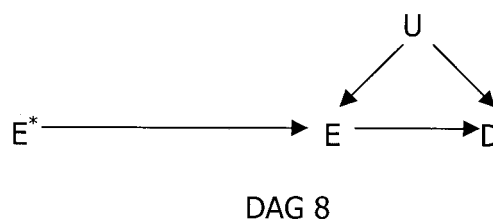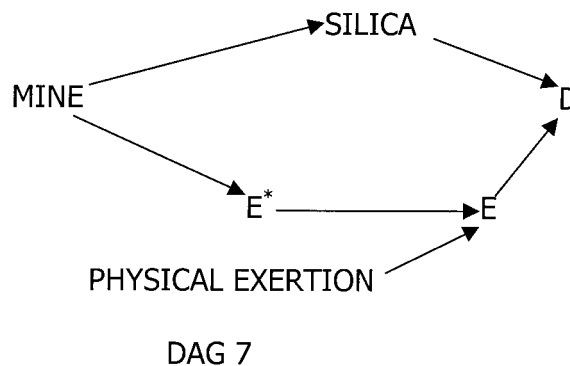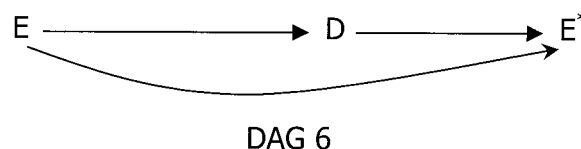


**FIGURE 2.** Directed acyclic graphs (DAG) for the Justifications of Answers section in Thought Experiment 1. D = disease (in DAG 6, D = congenital defect in offspring); E = exposure; E* = a misclassified version of E; U = an unmeasured common cause of E and D, such as underlying health status.

for the absence of an arrow between $E$ and $D$ (that is, lack of causality) by testing whether $E^*$ and $D$ are independent. This test amounts to the standard intent-to-treat analysis of a randomized trial. Thus, even in the presence of nonrandom noncompliance as a result of $U$, an intent-to-treat analysis provides for a valid test of the causal null hypothesis that $E$ does not cause $D$. Because $OR_{E^*D} = 0.5$ in our data, we conclude that we can reject the causal null and that $E$ protects against $D$ in at least some patients. Now, $OR_{E^*D}$ represents the effect of assignment to a low-fat diet on the outcome. Owing to noncompliance, this measure in general will differ from the causal effect $OR_{causal,ED}$ of actually following a low fat diet. Indeed, the magnitude $OR_{causal,ED}$ of the causal

effect of $E$ in the study population is not identified (that is, estimable), and one can only compute the bounds for it. Finally, note that the conditional association $OR_{E*D|E}$ = 0.3 also fails to have a causal interpretation. This conclusion reflects the fact that under the causal null of no arrow from $E$ to $D$, $E*$ and $D$ will be conditionally associated within levels of $E$, because $E$ is a common effect of both $E*$ and $U$, and $U$ is a cause of $D$.

## Thought Experiment 2: Postmenopausal Estrogens and Endometrial Cancer

Consider causal DAG 9 with $D$ being endometrial cancer, $C$ being vaginal bleeding, $A$ being ascertained (that is, diagnosed) endometrial cancer, $E$ being postmenopausal estrogens, and $U$ being an unmeasured common cause of endometrial cancer and vaginal bleeding (Figure 3). For simplicity, we assume that our diagnostic procedures have 100% sensitivity and specificity. So, every woman with $D$ who receives a diagnostic test will be successfully ascertained, as is represented by the arrow from $D$ to $A$. There may, however, be many women with endometrial cancer who have not had a diagnostic procedure and thus remain undiagnosed.

The absence of an arrow from $E$ to $D$ represents the Horowitz and Feinstein[2] null hypothesis that estrogens do not cause cancer. The arrow from $E$ to $C$ indicates that estrogens cause vaginal bleeding. The arrows from $C$ to $A$ indicate that vaginal bleeding leads to endometrial cancer being clinically diagnosed. The arrow from $D$ to $C$ indicates that endometrial cancer can cause vaginal bleeding. The arrows from $U$ to $D$ and $C$ indicate that some unknown underlying uterine abnormality $U$ independently leads to both uterine bleeding and cancer. We will also consider subgraphs of DAG 9 with various arrows removed.
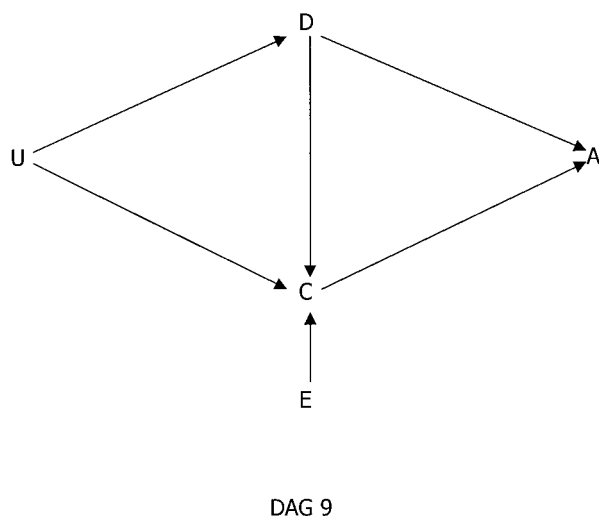
There will be ascertainment bias whenever the arrow from $C$ to $A$ is present, because then, among women with endometrial cancer, those who also have vaginal bleeding are more likely to have their cancer diagnosed.

Furthermore, $D$ and $C$ will be associated (d-connected) in the source population whenever either (1) endometrial cancer causes vaginal bleeding so that the arrow from $D$ to $C$ is present or (2) $U$ is a common cause of cancer and bleeding so that the arrows from $U$ to $D$ and from $U$ to $C$ are present.

Now consider a case-control study in which we find each clinically diagnosed case of endometrial cancer $D$ in a particular locale and select as a control a random age-matched woman yet to be diagnosed with endometrial cancer. If we let $b$ be the number of discordant pairs with the case exposed and $c$ be the number of discordant pairs with the control exposed, $b/c$ is the Mantel-Haenszel odds ratio (MH OR).

The MH OR is biased (that is, converges to a value other than 1) under the null hypothesis of no estrogen effect on endometrial cancer if and only if there is ascertainment bias. To see this, note that, under this design, the MH OR will converge to 1 (that is, be unconfounded) if and only if $A$ (diagnosed cancer) is unassociated with the exposure $E$. But $E$ and $A$ are associated (d-connected) if and only if there is an arrow from $C$ to $A$.

To adjust for vaginal bleeding, we might consider a second bleeding-matched design in which we additionally match controls to cases on the presence of vaginal bleeding in the month before the cases' diagnosis. Under this design, whether or not ascertainment bias is present, the bleeding-matched MH OR is biased away from 1 if and only if endometrial cancer $D$ and vaginal bleeding $C$ are associated (d-connected), owing to an unmeasured common cause $U$ or to $D$ causing $C$ or to both. This result follows by noting that the bleeding-matched MH OR is 1 if and only if $A$ is independent of (d-separated from) $E$ conditional on $C$. But, $A$ is d-separated from $E$ given $C$ if and only if $D$ and $C$ are unassociated. It follows that we have given a graphical proof of the well-known result that one cannot control for ascertainment bias by stratification on determinants of diagnosis if these determinants are themselves associated with disease.[5]

Combining the results, we can conclude that in the presence of both a vaginal bleeding-endometrial cancer association and ascertainment bias, the MH OR and the bleeding-matched MH OR are both biased.

It is now clear why there was a controversy: On biological and clinical grounds, it was believed that endometrial cancer caused vaginal bleeding and that vaginal bleeding led to the ascertainment of undiagnosed cancer. Thus, one could not validly test the Horowitz and Feinstein[2] null hypothesis whether or not one controlled for the determinant of ascertainment bias (that is, vaginal bleeding) in the analysis. We note that Greenland and Neutra,[5] Hutchison and Rothman,[3] and Jick et al[4] reach conclusions identical to ours. Our contribution is to demonstrate how quickly and essentially



DAG 9

**FIGURE 3.**   DAG for Thought Experiment 2. $D$ = endometrial cancer; $A$ = ascertained endometrial cancer; $C$ = vaginal bleeding; $E$ = exogenous estrogens; $U$ = an unmeasured common cause of $D$ and $C$.

automatically one can reach these conclusions by using causal graphs.

## Discussion

If every pair of variables had one or more unmeasured common causes, then all exposure-disease associations would be confounded. I believe that, in an observational study, every two variables have an unmeasured common cause, and thus there is always some uncontrolled confounding. Thus, when, as in our examples, one considers causal graphs in which certain pairs of variables have no unmeasured common causes, this situation should be understood as an approximation. Of course, in an observational study, we can never empirically rule out that such approximations are poor, as there may always be a strong unmeasured common cause of which we were unaware. For example, in the case-control study of our first thought experiment, those without sufficient subject matter expertise would not have had the background needed to recognize the possibility that a subclinical first trimester infection might be a common cause of exposure and the outcome. As epidemiologists, we should always seek highly skeptical subject-matter experts to elaborate the alternative causal theories needed to help keep us from being fooled by noncausal associations.

## References

1. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology 1999;10:37–48.
2. Horowitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. N Engl J Med 1978;299:1089–1094.
3. Hutchison GB, Rothman KJ. Correcting a bias? N Engl J Med 1978;299:1129–1130.
4. Jick H, Watkins RN, Hunter JR, Dinan BJ, Madsen S, Rothman KJ, Walker AM. Replacement estrogens and endometrial cancer. N Engl J Med 1979;300:218–222.
5. Greenland S, Neutra R. An analysis of detection bias and proposed corrections in the study of estrogens and endometrial cancer. J Chron Dis 1981;34:433–438.
6. Pearl J, Verma T. A theory of inferred causation. In: Allen JA, Fikes R, Sandewall E, eds. Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd International Conference. San Mateo, CA: Morgan Kaufmann, 1991:441–452.
7. Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search. New York: Springer Verlag; 1993.
8. Pearl J. Causal diagrams for empirical research. Biometrika 1995;82:669–688.
9. Pearl J, Robins JM. Probabilistic evaluation of sequential plans from causal models with hidden variables. In: Uncertainty in Artificial Intelligence: Proceedings of the 11th Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1995:444–453.
10. Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco: Morgan Kaufmann, 1988.
11. Lauritzen SL, Dawid AP, Larsen BN, Leimar HG. Independence properties of directed Markov fields. Networks 1990;20:491–505.
12. Balke A, Pearl J. Bounds on treatment from studies with imperfect compliance. J Am Stat Assoc 1997;92:1171–1176.