A Critique of the Statistical Evidence Associating Estrogens with Endometrial Cancer

Alvan R. Feinstein and Ralph I. Horwitz

Departments of Medicine and Epidemiology, Yale University School of Medicine, New Haven, Connecticut 06510

Abstract

The epidemiological problem of detection bias occurs when the evidence needed to diagnose a particular disease is sought more intensively in the group of people exposed to a particular agent than in the comparative group, who did not receive the agent. The bias can occur during community surveillance before hospitalization, during the ordering of diagnostic tests for hospitalized patients, or during the interpretation of the tests.

In postmenopausal women with endometrial cancer, detection bias is produced because of the dilation and curettage (or other diagnostic tests) ordered when bleeding occurs as a side effect of estrogen therapy. In estrogen takers, these diagnostic explorations allow the detection of endometrial cancers that might otherwise be silent and undetected in women who do not take estrogen. Since this bias occurs before patients are hospitalized, its removal requires special sampling and analytic techniques that have not been used in most epidemiological case-control studies where high risk ratios were found for the association between estrogens and endometrial cancer. When cases and controls have been selected with the use of a diagnostic procedure as the sampling frame, the odds ratios are much closer to 1; with appropriate stratification for the diagnostic stimulus of bleeding, the odds ratios approximate 1.

Other sources of bias for the estrogen-endometrial cancer association can be the "protopathic" prescription of estrogen for women with unrecognized endometrial cancer, the absence of "double-blind" interviewing techniques, and arbitrary definitions of "exposure to estrogens."

The difficulties in interpreting the relationship of estrogens and endometrial cancer indicate the need for development of rigorous scientific standards in the epidemiological methods used for case-control research.

Introduction

An important distinction in medical science is the difference between a diagnosis and a disease. If someone cites statistics showing that the 19th century killer disease dropsy has been eradicated, a knowledgeable scientist would nod quietly and say, "Nonsense. It is all due to changing standards and criteria for diagnosis. Dropsy still occurs abundantly as a disease, but it has disappeared as a diagnosis because we now use another name." Analo-
must be interpreted; if the interpretation is positive, the patient has finally become a diagnosed case.

When we compare the rate of disease in people receiving a particular pharmaceutical agent, such as estrogen, against that of people who are not receiving that agent, detection bias can arise at several places along this complex pathway. The first source of bias is in community surveillance. Exposure to an ongoing treatment, such as estrogen, may increase a person’s medical surveillance without affecting a silent signal event, such as a lump in the breast, but may thereby allow the lump a greater chance of medical detection. Alternatively, exposure to an agent such as estrogen may convert a “no signal” to an overt signal event, such as uterine bleeding, without affecting the true occurrence of a disease such as endometrial cancer. This overt signal event may then lead to the diagnostic testing and identification of a previously asymptomatic and otherwise undetected cancer.

In both of the 2 instances just cited, the exposed group would receive a greater community surveillance than would the people who were nonexposed. Thus, if estrogens provoke the bleeding that leads to a referral for dilation and curettage, asymptomatic uterine cancer is much more likely to be found in women receiving estrogens than in those who do not receive it. Evidence to support this hypothesis has been obtained in almost every study where the grade of endometrial cancer was suitably examined and correlated. Regardless of the total occurrence rate of the cancer, the proportion of the frequently asymptomatic Grade I tumor has been higher in estrogen users than in nonusers and higher in women with uterine bleeding than in nonbleeders.

A counterpart of the disparity in community surveillance occurs as an additional, second source of detection bias. This source of bias arises not in the community but in the hospital. Regardless of why the patients were admitted to the hospital, a diagnostic examination bias will occur when exploratory or definitive diagnostic procedures are ordered more often in estrogen users than in nonusers. This preferential ordering will occur when a history of uterine bleeding is encountered more often during the routine review of systems in estrogen users than in nonusers.

A third source of detection bias arises when a knowledge of exposure or nonexposure alters the (objectivity of the person who interprets the results of the tests. Thus, the diagnostic decision of a pathologist examining endometrial tissue may be affected by previous knowledge about whether or not the patient is receiving estrogen. This third source of detection bias is sometimes considered in case-control studies, and a separate review (5) may be sought for histological or other diagnostic evidence of the cancer. Unfortunately, such reviews deal only with false-positive diagnoses. They do not deal with the problem of false-negative diagnoses in the controls or with the problem of diagnostic referral bias in both cases and controls.

The term “diagnostic referral bias” can be used to include the impact of both community surveillance bias and diagnostic examination bias. This problem is substantially different from the type of hospitalization bias that was first described by Berkson (2) as a purely passive mathematical phenomenon, in which people with 2 ailments have a higher probability of hospitalization than do people who have only one ailment. Diagnostic referral bias is an active clinical entity in which physicians create different rates of hospitalization and/or diagnostic testing according to the clinical events that occur in the 4 groups of people who are exposed or nonexposed and diseased or nondiseased.

The effects of diagnostic referral bias can readily be demonstrated for the case-control studies in which the alleged association of estrogens and endometrial cancer has been studied. The demonstration needs only some simple algebra, as shown in Table 1. In the community the proportion of exposed people can be indicated as e, with the nonexposed as 1 – e. The target disease develops at the rate p1 to create the diseased cases of the nonexposed group and at the rate p2 to create the diseased cases in the exposed group. The proportions found in the corresponding nondiseased control groups are 1 – p1 and 1 – p2. These 4 groups will have hospital referral rates ranging from h1 in the nonexposed controls up to h2 in the exposed cases. What has been found for these 4 groups in the hospital is shown in the proportions in the column on the far right.

When the odds ratio is calculated for the hospitalized cases and controls, the terms containing e and 1 – e are canceled. If p1 and p2 are very small (as they usually are), the terms in 1 – p1 and 1 – p2 are very close to 1 and can be ignored, so that the odds ratio becomes (p1/p2) x (h2/h1) x (h1). If we let k, indicate h1/h2 as the bias due to exposure on the case group’s hospital referral, and if we let k–, indicate h1/h_, as the analogous bias of exposure on the control group’s referral, the odds ratio is the true risk ratio p1/p2, multiplied by k,/(k–,), which is the ratio of the 2 referral biases.

The accuracy of the results found in a case-control study will depend upon the way these biases in k, and k–, are managed. If the 2 biases are equal, they will cancel each other; k,/(k–,) will equal 1; and the odds ratio will be a true substitute for the risk ratio. If k, is substantially larger than k–, their ratio will exceed 1, and the effect will falsely elevate the odds ratio, so that it provides an inflated value for the true risk ratio. We shall return to these algebraic effects later on.

---

**Table 1**

<table>
<thead>
<tr>
<th>Presence or absence of exposure</th>
<th>Proportion</th>
<th>Presence or absence of disease</th>
<th>Proportion</th>
<th>Proportion referred to hospital</th>
<th>Proportions in hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposed</td>
<td>e</td>
<td>Case</td>
<td>p1</td>
<td>h1</td>
<td>e(1 – p2)h1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>1 – p2</td>
<td>h2</td>
<td>e(1 – p2)h2</td>
</tr>
<tr>
<td>Nonexposed</td>
<td>1 – e</td>
<td>Case</td>
<td>p2</td>
<td>h2</td>
<td>(1 – e)p1h2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Control</td>
<td>1 – p1</td>
<td>h1</td>
<td>(1 – e)(1 – p1)h1</td>
</tr>
</tbody>
</table>
Reduction of Detection Bias

The problems of detection bias have been ignored in most case-control studies and the problems are difficult to deal with, because the bias has already occurred externally before the collected cases and controls enter the groups assembled in a conventional case-control study. To try to equalize those external forces of surveillance and referral, we must alter the conventional method of epidemiological selection. Instead of choosing 2 individual groups (one group of cases and a separate group of controls), we can select both the cases and the controls from a single sampling frame, consisting of people referred for performance of the particular diagnostic test used to identify the target disease. For example, in a study of breast cancer, the sampling frame might be a registry of all patients who have received a biopsy of breast lesions. For endometrial cancer, the sampling frame might be all patients who have received dilation and curettage or hysterectomy. Since the results of the diagnostic test will indicate those people who are potentially eligible to be cases or controls, this kind of sampling frame also has the advantage of avoiding arbitrary decisions by the investigator about the kind of diagnoses to be contained in the control group.

Some additional algebra could be used\(^2\) to show that this kind of sampling frame will reduce diagnostic referral bias but will not eliminate it. To come closer to elimination, we need yet another step, which consists of stratifying patients according to the cogent clinical reason that led to their referral to the hospital for the diagnostic test. For example, in studies of endometrial cancer, the patients sampled from the diagnostic-test registry should be stratified into 2 groups, those who were or were not referred to the hospital for the diagnostic test. For example, in a study of breast cancer, the sampling frame might be all patients who have received dilation and curettage or hysterectomy. Since the results of the diagnostic test will indicate those people who are potentially eligible to be cases or controls, this kind of sampling frame also has the advantage of avoiding arbitrary decisions by the investigator about the kind of diagnoses to be contained in the control group.

Some additional algebra could be used\(^2\) to show that this kind of sampling frame will reduce diagnostic referral bias but will not eliminate it. To come closer to elimination, we need yet another step, which consists of stratifying patients according to the cogent clinical reason that led to their referral to the hospital for the diagnostic test. For example, in studies of endometrial cancer, the patients sampled from the diagnostic-test registry should be stratified into 2 groups, those who were or were not referred to the hospital for uterine bleeding. If a similar odds ratio is found in both groups, the 2 odds ratios are quite likely to reflect the true risk ratio.

Some of these distinctions can be seen by studying the data in Table 2, which shows results that have emerged from various case-control studies of the relationship of estrogens and endometrial cancer. The left side of the table contains the results of 4 studies in which the cases and controls were chosen in the conventional epidemiological manner. The cases all had endometrial cancer. The controls in the 4 studies consisted of women with other gynecological cancers, members of a health plan, members of a retirement community, or members of the same medical community. The odds ratios ranged from 4.9 to 8.0 in those studies (12, 13, 16, 18). On the right side of the table, however, are the results obtained when cases and controls were taken from a diagnostic-test registry. In the study by Gray et al. (6), the cases and controls were chosen from women who had all had hysterectomies. Although the results were not stratified for bleeding or nonbleeding, the odds ratio was lowered to 3.1. In the 2 older studies by Dunn and Bradbury (3) and Pacheco and Kempers (15), the patients were all women who had had dilation and curettage for bleeding. The odds ratios were 1.1 and 0.5. In research done by Horwitz and Feinstein (10), the results confirm what was found in the 2 latter studies.

The differences in these odd ratios can readily be understood when we consider the values of \(k_c\), \(k_r\), and the \(k_c/k_r\) ratio that were discussed earlier. As shown in Table 3, in patients with endometrial cancer, estrogen takers are more likely to bleed and be referred to the hospital than are non-estrogen takers. Thus, the value of \(k_c\) will exceed 1. To counteract this bias, we need a control group in which \(k_c\) will exceed 1. For controls who are chosen from women with other gynecological cancer, estrogen usage will have little or no impact on hospital referral, since hospitalization for these diseases is provoked by many symptoms other than uterine bleeding. In fact, it may be found that among patients with these diseases there may be a deficit of estrogen takers, because the use of estrogen may have been stopped when the first symptoms of those other cancers appeared. Thus, in patients with other gynecological cancers, \(k_r\) is less than or equal to 1, so that \(k_c/k_r\) will be substantially larger than 1, causing even further inflation of the risk ratio. In controls consisting of women chosen at large from the community who had no special uterine condition that is susceptible to estrogen-induced bleeding, estrogen usage will not have an impact on hospital referral. The \(k_c\) value will essentially equal 1, so that the \(k_c/k_r\) ratio will remain elevated, preserving its existing bias.

When the control group is chosen from women with noncancerous uterine disease (via selection from a diagnostic-test registry), estrogen-induced bleeding will in-

Table 2

<table>
<thead>
<tr>
<th>Principal investigator</th>
<th>Control group</th>
<th>Odds ratio</th>
<th>Principal investigator</th>
<th>Control group</th>
<th>Odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smith (10)</td>
<td>Other gynecological cancer</td>
<td>7.5</td>
<td>Gray (6)</td>
<td>Not stratified</td>
<td>3.1</td>
</tr>
<tr>
<td>Ziel (18)</td>
<td>Health plan</td>
<td>7.6</td>
<td>Dunn (3)</td>
<td>Bledgers</td>
<td>1.1</td>
</tr>
<tr>
<td>Mack (12)</td>
<td>Community</td>
<td>8.0</td>
<td>Pacheco (15)</td>
<td>Bledgers</td>
<td>0.5</td>
</tr>
<tr>
<td>McDonald (13)</td>
<td>Community medicine</td>
<td>4.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3

<table>
<thead>
<tr>
<th>Clinical condition</th>
<th>Referral rates for patients treated with estrogen vs. NE</th>
<th>Referral ratio</th>
<th>(k_c/k_r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endometrial cancer</td>
<td>E &gt; NE</td>
<td>(k_c &gt; 1)</td>
<td>(k_c/k_r)</td>
</tr>
<tr>
<td>Other gynecological cancer</td>
<td>E (\leq) NE</td>
<td>(k_c \leq 1)</td>
<td>(k_c/k_r)</td>
</tr>
<tr>
<td>No uterine disease</td>
<td>E (=) NE</td>
<td>(k_c = 1)</td>
<td>(k_c/k_r)</td>
</tr>
<tr>
<td>Other uterine disease</td>
<td>E (&gt;) NE</td>
<td>(k_c &lt; 1)</td>
<td>(k_c/k_r)</td>
</tr>
</tbody>
</table>

\(a\) E, estrogen; NE, no estrogen.

crease the referral of the estrogen takers, thus elevating the value of \( k_e \). This will have a chance to counterbalance the elevated \( k_e \), so that the \( k_e/k_r \) ratio can approximate 1, thus giving the odds ratio an opportunity to yield a result closer to the true value of the risk ratio.

Susceptibility Bias

In addition to this major problem in detection bias, another important potential source of generally ignored error in case-control studies is susceptibility bias. Susceptibility bias can arise in 2 different ways, clinically and protopathically. Clinically, the possibility exists that women who develop the postmenopausal syndrome may also be particularly likely to develop endometrial cancer. If such women also happen to be treated with estrogens for their postmenopausal syndrome, a distinct association will be found between the use of estrogens and the subsequent occurrence of endometrial cancer. The estrogens in this circumstance, of course, would be a prognostic marker or prognostic indicator but not an etiological agent. This possibility cannot be ruled out at the moment, particularly since no satisfactory cohort studies have been conducted to demonstrate the natural course and outcome of women with postmenopausal syndrome.

Protopathic bias creates a different kind of problem in susceptibility. In this instance the women who receive the estrogen are not merely susceptible to endometrial cancer; they already have it at the time at which they get the estrogen. Consider the situation, for example, in which a woman develops postmenopausal bleeding. Because her Papanicolaou smear is negative and her pelvic examination is negative, her doctor may not do a more extensive diagnostic work-up and may instead give her estrogen to suppress the bleeding. After it has remained dormant for a period of time, further bleeding or some other symptoms may occur, at which time the Papanicolaou smear and the pelvic examination may still be negative. A dilation and curettage or some other intraendometrial diagnostic examination procedure may now be done, however, and reveal a cancer. When this revelation occurs, the estrogen therapy may then be falsely associated with the endometrial cancer.

The existence of protopathic bias has already been demonstrated in several circumstances, most strikingly when Janerich et al. (11), in studying the alleged relationship of oral contraceptive pills and benign breast tumors, discovered that women with breast tumors were regularly told to discontinue oral contraceptive pills at the time at which the lump in the breast was discovered. When such women later become the cases in a case-control series, they will have a deficit in the usage of oral contraceptive pills, because the pills were stopped at the protopathic discovery of their breast lumps.

Reduction of Susceptibility Bias

The best way to help remove both clinical and protopathic sources of susceptibility bias would be to determine the reason for the prescription of estrogen in each patient who received it. The data can then be stratified and analyzed appropriately. Nevertheless, a question as to why the agent was prescribed has not been asked in most of the case-control studies about estrogens and endometrial cancer, or, if the question was asked, the answers have not been appropriately analyzed.

Interviewer Bias

At least 2 other biases may be noted in passing. Interviewer bias can occur when the person making the inquiry about usage or nonusage of estrogens is aware of the research hypothesis or of the patient's identity as a "case" or "control." This problem can be avoided with "blind interviewing," a tactic that was not deliberately used in many of the cited case-control studies. Recall bias (or anamnestic bias) occurs when the cases are more inherently stimulated than are the controls to recall previous pharmaceutical exposure. For avoidance of this problem, the interview process should be particularly thorough, and the results should be checked with physicians or pharmacists. This tactic was also not used in most of the cited case-control studies.

Bias Due to Differing Definitions of Exposure

Finally, the last bias relates to the definition of exposure itself. What is meant by "antecedent usage of estrogens"? As defined by McDonald et al. (13), it meant conjugated estrogens for at least 6 months. With that definition, McDonald et al. found an odds ratio of 4.9. On the other hand, Smith et al. (16) defined antecedent usage as exogenous estrogen for more than 6 months. With that definition, McDonald found that the odds ratio in their data dropped to 2.3. A third definition, coming from Ziel and Finkle (18), was that exposure consisted of conjugated estrogen of any duration. With that definition, McDonald's odds ratio dropped to 2.0. Finally, according to the results of Mack et al. (12), exposure consisted of any estrogen, given for any duration. With that definition, McDonald's odds ratio dropped to 0.9, exonerating the estrogens as a possible cause of endometrial cancer. Consequently, for interpretation of the results of the case-control studies, a close look should be taken at what the authors define as "exposure to estrogens." The definition can be jugged to provide almost any odds ratio that is desired.

Conclusion

There is a peculiar double standard in the scientific attitudes of our era. To claim a cause-effect relationship for the benefits of a therapeutic agent, a clinical investigator must comply with rigorous methodological criteria that are intended to avoid imprecision and bias. The base-line condition and outcome events must be clearly delineated, the compared agents must be allocated by randomization, and the subsequent observations must be conducted with double-blind methods. On the other hand, to claim a cause-effect relationship for the alleged hazards of a therapeutic agent, an epidemiological investigator has no scientific obligations or constraints. The work can be retrospective, the cases and controls can be chosen in any way that the investigator wishes, the definition of exposure can be ad-
justed arbitrarily, the data can be acquired without efforts at blinding, and the analysis can be performed with little or no attention to major biases in detection or susceptibility.

The substantial conflicts and contradictions that have occurred in studies of reserpine, estrogens, saccharin, and many other topics (9) indicate that we can no longer remain complacent about the absence of scientific standards for case-control research. Perhaps the ultimate virtue of the current controversy about estrogens and endometrial cancer is that thoughtful scientists will be confronted with the inadequate quality of the conventional methods used for epidemiological research in chronic disease. Because of logistics and feasibility, we may remain unable to conduct randomized trials, but with careful attention to the existing defects, we can begin to remove the complacency of the past and to construct rigorous scientific standards for the future.

References

A Critique of the Statistical Evidence Associating Estrogens with Endometrial Cancer

Alvan R. Feinstein and Ralph I. Horwitz


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/38/11_Part_2/4001

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.