Endocrine, *vol. 24, no. 3, 203–210, August 2004* 0969–711X/04/24:203–210/$25.00

Adverse Neoplastic and Cardiovascular Outcomes of HRT

The Validity of the Evidence

Samuel Shapiro

Emeritus Director, Slone Epidemiology Centre, Boston University, Visiting Professor of Epidemiology, Columbia University and the University of Cape Town

The possibility that HRT may increase the risk of cancer, female cancers in particular, has been a general concern ever since HRT was first advocated in the mid-1950s. For cancer of the uterus that concern was vindicated when it was shown that unopposed estrogens greatly increase the risk. However, because HRT is undoubtedly effective in relieving menopausal symptoms, and because it was thought that HRT may reduce the risk of coronary heart disease, the use of estrogens in combination with progestins soon supplanted unopposed estrogen use (if less so among women who had undergone hysterectomies), in the belief, soon confirmed by epidemiological evidence, that combined therapy probably eliminates any increase in the risk of uterine cancer, or perhaps even reduces it.

Key Words: HRT; cardiovascular outcomes; MWS breast cancer.

Introduction

The possibility that HRT may increase the risk of cancer, female cancers in particular, has been a general concern ever since HRT was first advocated in the mid-1950s. For cancer of the uterus that concern was vindicated when it was shown that unopposed estrogens greatly increase the risk *(18)*. However, because HRT is undoubtedly effective in relieving menopausal symptoms, and because it was thought that HRT may reduce the risk of coronary heart disease *(20)*, the use of estrogens in combination with progestins soon supplanted unopposed estrogen use (if less so among women who had undergone hysterectomies), in the belief, soon confirmed by epidemiological evidence *(9)*, that combined therapy probably eliminates any increase in the risk of uterine cancer, or perhaps even reduces it.

Because breast cancer is a more common and serious disease than uterine cancer, the possibility that HRT may increase the risk of that condition has, if anything, been an even greater concern, and there has been a considerable investment of epidemiological research into that question. Initially, since unopposed estrogens were the most common therapy, the research was concentrated on those compounds, and then, as combination therapy replaced unopposed estrogens, the research shifted to the evaluation of that treatment.

For many years, among a large number of epidemiological studies, the evidence concerning breast cancer risk, initially in relation to estrogen use, and later in relation to estrogen plus progestin use, was conflicting. However, in a meta-analysis [the Collaborative Re-analysis (CR) *(6)*] of 51 studies undertaken in 1997, and representing over 90% of all studies undertaken up to that time, it was concluded that among current users of HRT (over 80% of which was unopposed estrogens) for 5 yr or more, the risk of "having breast cancer diagnosed" was increased 1.35-fold; findings for combination therapy were inconclusive, but were interpreted as being compatible with an increased risk. It was assumed further that the apparently conflicting findings among the studies conducted up to that time mainly reflected statistical instability because of their limited size. And subsequently, when later observational studies of combination therapy became feasible, they were interpreted as suggesting an increased risk in women on that treatment—and perhaps an even higher risk than that conferred by estrogens alone *(11)*.

In observational research, relative risk estimates well below 3.0 are considered small, and making the distinction between bias, confounding, and causation can be difficult, if not impossible. In addition, since combination therapy appears not to be associated with an increased risk of uterine cancer, and since a large body of evidence pointed to potential beneficial effects of unopposed estrogens, as well as combined therapy, on the risks of atherosclerotic outcomes *(20)*, principally coronary heart disease, and on osteoporosis *(23)*—and perhaps on Alzheimer's disease *(21)* and large bowel cancer *(8)*—it was widely believed that the benefits of HRT might outweigh the risks.

Anxiety about breast cancer risk nevertheless persisted. In addition, a growing body of evidence suggested that some of the apparent beneficial effects of HRT, especially on the risk of coronary heart disease, may partly or wholly be accounted for by confounding if women who used HRT tended to be at lower risk (due to factors such as socioeco-

Received June 21, 2004; Revised August 5, 2004; Accepted August 10, 2004.
Author to whom all correspondence and reprint requests should be addressed: Samuel Shapiro, Visiting Professor of Epidemiology, Columbia University and the University of Cape Town. E-mail: ss1832@columbia.edu

nomic status, education, exercise, diet, and so on) than those who did not *(2)*. If this were so, any increase in the risk of breast cancer, no matter how small, might be unacceptable. And at minimum it may be necessary to weigh the overall benefits [(such as those due to a reduced risk of osteoporotic fractures, Alzheimer's disease, or large bowel cancer) against the potential hazards [such as increased risks of breast cancer, or venous thromboembolism *(4)*].

Against this background two further studies were undertaken. The first study was a large-scale randomized controlled trial, the Women's Health Initiative (WHI) *(24)* designed primarily to assess the risks of cardiovascular disease and breast cancer among women treated with conjugated estrogens and medroxyprogesterone acetate—and secondarily to assess the risks of several other outcomes; it was also designed to determine whether, overall, the anticipated benefits exceed the risks, or vice versa. The second study was a massive observational follow-up study, the Million Women Study (MWS) *(13)*, designed specifically to assess the risk of breast cancer among HRT recipients in a population of women undergoing routine screening mammography. In that study it was thought that any statistical uncertainties could for practical purposes be eliminated, and it would also be possible to conduct detailed analyses within subgroups of interest.

Initially, relatively little stir was aroused by the publication of the CR results, probably because it was still widely believed that, overall, HRT was beneficial. However, that situation changed dramatically in 2002 when the WHI was discontinued *(24)*. Because the WHI was represented as a randomized controlled trial, it was claimed that the findings established beyond any reasonable doubt, firstly, that HRT increases the risk of breast cancer while not at the same time decreasing the risk of coronary heart disease, and, secondly, that a global index comprising several outcomes indicated that the overall hazards posed by estrogen plus progestin significantly outweigh the benefits. Among the benefits of HRT identified in the WHI were significantly reduced risks of fractures and of cancer of the colon, but those benefits were outweighed by the combined effects of four outcomes, each of which were significantly associated with increased risks: breast cancer, coronary heart disease, stroke, and pulmonary embolism.

Finally, when the results of the MWS were published a year later *(13)*, it was claimed that for practical purposes there was now no longer any question at all that current HRT use increases the risk of breast cancer, and that estrogens plus progestins increase the risk more than do estrogens alone.

Today the general perception with regard to breast cancer risk is that the associations with HRT must be regarded as causal because in the WHI randomization and "double-blinding" have finally resolved issues of bias, and because of their massive size the CR, and more especially the MWS, have eliminated chance. The latter two studies have also demonstrated that unopposed estrogens increase the risk of breast cancer, but the findings in the MWS indicate that they do so to a lesser degree than does estrogen plus progestin. With regard to cardiovascular risk, it is claimed that the WHI has demonstrated that earlier concerns about confounding were probably correct, and that there is no protection conferred by HRT upon the risk of coronary heart disease: it is even possible that HRT may increase the risk; HRT also appears to increase the risk of stroke and venous thromboembolism.

While the findings from the CR had relatively little impact on medical practice, the publication of the WHI findings, followed by those of the MWS, has had a profound influence. The magic phrases, "randomized controlled trial" and "million women," compounded by statistical arcana that are virtually impenetrable to the ordinary reader, appear to have silenced virtually all criticism because the "truth" has now been established.

Here I present evidence to show that the "truth" has by no means been established, and that there are strong grounds to challenge the validity of the findings concerning breast cancer risk in all three studies, and of the findings concerning cardiovascular disease in the WHI. That evidence has been published in detail elsewhere *(18,19)*, and this article gives a general overview. More recent evidence from the WHI concerning the effects of exposure to conjugated estrogens alone is also considered.

Breast Cancer

Two potential sources of bias are especially relevant to any assessment of breast cancer risk. Information bias is relevant because the hypothesis that HRT may increase the risk of breast cancer has been widely broadcast for many years. Consequently, in case-control studies, breast cancer cases are likely to remember their exposures more fully than controls.

Detection bias is relevant because in any population of women there is always a pool of clinically silent breast cancers: indeed, the rationale for the use of mammography is to detect breast cancer sooner than would otherwise be the case. Thus it is possible that breast cancer can selectively be identified among HRT users by screening. For example, it can be anticipated that HRT users will more frequently and more carefully search for breast lumps than will non-users. Their medical attendants will also do so—and among the latter it is standard practice to examine the breast on each occasion that a woman renews her prescription, while non-users, of course, do not renew prescriptions. It can also be anticipated that users will more frequently attend for mammography—and that mammographers might search more intensely for evidence of neoplasia if they become aware that any particular woman is on HRT:

Table 1
Collaborative Re-analysis. Excess Cases of Breast Cancer Among HRT Users

Duration of current HRT use (yr)	Distribution among controls (%)	Relative risk	Expected excess of cases (%)	Absolute excess (%)
Never	74.5	1.00	74.50	0.00
<1	2.7	0.99	2.67	–0.03
1–4	6.4	1.06	6.78	0.38
5–9	4.0	1.31	5.24	1.24
10–14	2.0	1.24	2.48	0.48
≥15	1.6	1.56	2.50	0.94
			Total excess	3.01
			Average excess per year	<0.20

they may especially do so because HRT increases breast tissue density *(7)*, a suspected risk factor for breast cancer.

Still further, the longer the duration of HRT, the greater the anxiety about breast cancer risk is likely to become, so that both information bias and detection bias become progressively more accentuated with the passage of time. On the null hypothesis, such a bias could readily produce an seeming duration/response effect with progressively higher relative risk estimates over time.

No epidemiological study is perfect, but it may nevertheless be possible to reach a causal inference if several reasonably well-conducted studies identify statistically significant relative risk estimates that are so large that it is unlikely that any imperfections in those studies can fully account for them *(17)*. In the present context, however, small relative risks, virtually all of them less than 2.5, are at issue, and even modest degrees of bias might readily account for them.

What evidence is there to support the likelihood of bias in each of the three studies?

***The Collaborative Re-analysis* (6)**

This study was a meta-analysis of "raw data" from 51 studies conducted before 1997, representing over 90% of the studies conducted up to that time. Among women who were using HRT or were last exposed less than 5 yr previously (defined as current users) there was no evidence of an increased risk if the total duration of use was less than 5 yr; among those currently exposed for ≥5 yr, the relative risk was 1.35, and statistically significant. Among women last exposed ≥5 yr previously there was no increase in the risk, regardless of duration of exposure.

Critique

The great majority of the data were derived from case-control studies, and because the hypothesis was widely known, and feared, there could have been both information bias, as well as detection bias, and in the same direction, in all or most of the studies; in the follow-up studies detection bias could have been present. In addition, especially among currently exposed women, for reasons given above, on the null a duration/response effect could have been anticipated.

Could the bias have been of sufficient magnitude to account for a relative risk of 1.35 among women exposed for ≥5 yr? Consider detection bias. Table 1 gives the percentage distribution of current HRT use among the control women in the collaborative re-analysis according to durations of >1, 1–4, 5–9, 10–14, and ≥15 yr of exposure. If the percentages in each of the time categories are multiplied by their respective relative risk estimates, and if those data are then summed, then on the null the findings could have been accounted for if, over a span of more than 15 yr, bias augmented the detection of otherwise clinically silent breast cancer by 3%: that is, by an average of less than 0.2% per year. Moreover, this degree of detection bias is an overestimate because it takes no account of the likelihood of information bias, which would have resulted in overestimation of the relative risk estimates used in this calculation. Thus a minimal amount of bias—far less than can conceivably be ruled out in an observational study—could readily have accounted for the findings.

There was also clear-cut evidence to indicate detection bias. Table 2 gives relative risk estimates according to time and duration of exposure, and according to whether the breast cancer was early (confined to the breast) or late. Among currently exposed women, in each of the duration categories, the relative risk estimates were higher among women with early breast cancer. Moreover, although, overall, past HRT was not associated with breast cancer, the same pattern of higher relative risk estimates for early tumors was evident within that category as well. That is, there was a consistent pattern for all relative risks to be higher for those cancers which were most susceptible to differential detection.

Contrary to what has been claimed, it is impossible to distinguish between bias and causation as alternative explanations for the findings in the CR.

Table 2
Collaborative Re-analysis.
Relative Risk Estimates for Localized and Widespread Breast Cancer Among HRT Users

Duration of HRT use (yr)	Last use <5 years previously		Last use ≥5 years previously	
	Localized	Widespread	Localized	Widespread
<1	1.09	0.68	1.12	1.01
1–4	1.32	0.90	1.13	1.08
5–9	1.67	1.04	1.23	0.88
≥10	1.42	1.25		

The Women's Health Initiative Randomized Controlled Trial (24)

Elsewhere I have published a detailed critique of this study *(18)*. Briefly, initially 8175 women were randomly assigned to conjugated estrogens and medroxyprogesterone acetate, and 8102 were assigned to placebo; later 331 women with intact uteri, who were initially assigned to conjugated estrogens alone, were unblinded and reassigned to combination therapy (because of concerns about the risk of uterine cancer), giving a total of 8506 estrogen plus progestin recipients.

The trial was discontinued after a mean follow-up of 5.2 yr. The reasons for discontinuation were "…because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits" (p. 321). For breast cancer, the hazard ratio (a hazard ratio is roughly the same as a relative risk estimate) was 1.26, an association that "…almost reached nominal statistical significance" (p. 327). The findings were interpreted as indicating that "…estrogen plus progestin does increase the risk of breast cancer" (p. 330).

Critique

Given the design of this study it was predictable that it would soon cease to be double-blind because women on estrogen plus progestin commonly experience tenderness of the nipple and breast, weight gain, bloating, elimination of hot flashes and sweats, elimination of discomfort during sexual intercourse, and most especially, episodes of uterine bleeding, a phenomenon which in post-menopausal women raises the possibility of uterine cancer and the need to rule it out—including the need to unblind the treatment. Furthermore, many women do not adhere to therapy: large losses to follow-up were predictable, and would afford added opportunity for bias to occur.

Those predictions were confirmed in the data: fully 44.4% of the women on estrogen plus progestin, as against 6.8% of the placebo recipients, were unblinded, mainly because of persistent uterine bleeding, giving a relative risk of 6.5 (44.4/6.8), and a difference in the blinding rates of 37.2% (44%–6.8%). With such a high rate of unblinding among the exposed women, and with such high differential rates of unblinding, bias became a strong likelihood, and perhaps even inevitable. That likelihood was further augmented by large losses to follow-up: among the exposed and non-exposed women these were 42% and 38%, respectively, and the corresponding crossover rates (from exposure to non-exposure, or vice versa) were 6.2% and 10.7%.

Given such high rates detection bias could have arisen in one or more of the following ways:

- More frequent breast examinations by the exposed women and their medical attendants (surveillance).
- Progressively more intensive surveillance with the passage of time.
- Selective attendance for follow-up at the study clinics when exposed women (or their medical attendants) discovered breast lumps, which in placebo recipients could have remained undiscovered.
- Selective attendance for follow-up mammograms.
- Selective detection of breast cancer among exposed women who had mammograms.

Especially given a hazard ratio of 1.26, which to begin with only "almost reached significance," virtually any amount of bias, no matter how small, could have reduced a lower 95% confidence limit of 1.0 to <1.0. Even the hazard ratio point estimate of 1.6 could have been accounted for by only a small amount of bias: the average incidence of breast cancer in the estrogen plus progestin recipients was 3.8 per 1000 per year, and the corresponding rate in the placebo recipients was 3.0 per 1000. That is, if, on average, among 1000 women bias resulted in the detection each year of 0.8 cases that would otherwise have remained clinically silent, it would entirely have accounted for the observed association.

Contrary to what was claimed, the WHI did not establish that combination HRT increases the risk of breast cancer, and again it is impossible to distinguish between causation and bias as alternative explanations of the findings. Although the WHI commenced as a controlled trial, in terms of the potential for detection bias it soon ceased to be one, and it

came to resemble a follow-up observational study, with the limitations intrinsic to that methodology. Moreover, consciousness of being a participant in an experiment, coupled with a high level of consciousness among HRT users of being exposed, could even have made the problem of detection bias more intractable than in a non-experimental follow-up study.

In spite of the above considerations the likelihood of detection bias was not considered in the initial WHI report *(24)*, and only briefly mentioned as being unlikely in a subsequent report based on a mean follow-up of 5.6 yr *(5)*. After that report was published I asked why hazard ratios for the blinded and unblinded women had not been presented *(18)*. Such data might shed some light on the question of detection bias, although it is unlikely that they would fully resolve it But in any event, some 2 yr after the findings were first published, hazard ratios according to blinding status have still not been presented.

The Million Women Study (13)

Elsewhere I have published an editorial commentary on this study *(19)*. Briefly, British women aged 50–64 yr who were scheduled to undergo a routine screening mammography (offered by the National Health Service every 3 yr) were invited to participate in a study of breast cancer. On recruitment the women completed a questionnaire. They were then followed in cancer registries for a mean of 2.6 yr for breast cancer incidence, and in death registries for 4.1 yr for mortality.

The main analyses were confined to past and present HRT use, as recorded at recruitment, among 828,923 postmenopausal women. As compared with women who had never used HRT, the risk of breast cancer was not increased among women last exposed more than 1 yr earlier. Among women who at the time of recruitment were HRT users (current users), the relative risk estimates for estrogen alone, estrogen plus progestin, tibolone, and other or unknown HRT were 1.30, 2.00, 1.45, and 1.44, respectively, and in all instances statistically significant—as was the difference between the estimate of 2.00 for combination therapy vs 1.30 for unopposed estrogen. For both of the latter two exposures, the relative risk estimates were significantly higher for total durations of use for 5 yr or more. For fatal breast cancer the relative risk was 1.22, and significant.

The findings have been interpreted as having confirmed earlier studies, and as now having established beyond any reasonable doubt that HRT increases the risk of breast cancer, with combination therapy conferring a greater risk than does estrogen alone.

Critique

As a conceptual matter it is questionable whether volunteers in a breast cancer screening program can serve as a population in which a valid estimate of breast cancer risk in relation to HRT use can be obtained, for the following reasons:

- It is known that women with breast lumps tend selectively to participate in screening programs *(15)*.
- It is likely that women on HRT also tend selectively to participate because of anxiety about possible breast cancer risk.
- Hence it can be anticipated that women who both have breast lumps, and who use HRT would on average be those most likely to participate.
- As mentioned above, increased mammographic density would lead to a more intense search for evidence to suggest breast cancer; such increased density is provoked by HRT use, and knowledge of such use might render the search even more intense.

The data supported this reasoning: firstly, an invitation to participate in breast cancer screening inevitably creates anxiety, and if that were not enough, the opening paragraph of the questionnaire (available at the study website: http:\\www.millionwomenstudy.org) would inevitably have heightened that anxiety, and have heightened it specifically with regard to the possible carcinogenicity of HRT: "We have a unique opportunity…to learn about the way the different types of [HRT]…affect a woman's health, particularly her breasts" (sic).

Secondly, the incidence of breast cancer at or soon after recruitment was higher (2.8 per 1000 per year) than in the population at large (2.0 per 1000 per year) *(13,18)*—a phenomenon that can only be explained by a selective tendency of women with breast lumps to participate.

Thirdly, the prevalence of HRT use among women who participated in the MWS was 32%, as against 19% among those who did not ($p < 0.0001$) *(1)*—a striking difference that clearly indicated a selective tendency for users to participate.

Fourthly, the mean time interval from recruitment to diagnosis was only 1.2 yr, indicating that a substantial proportion of the breast cancers were already present before the women enrolled in the study. In addition, for a disease that on average takes about 10 yr to become clinically evident, it is biologically implausible that an increased risk can become evident in such a short time interval.

Fifthly, although the time from recruitment to death was not given, the average time from diagnosis to death can be estimated by subtracting the mean time to diagnosis (2.6 yr) from the duration of follow-up for mortality (4.1 yr): i.e., 1.7 yr. Such a short interval can only be explained if a large proportion of the tumors were already present, and advanced, at the time of recruitment.

Sixthly, death certificates are known to be unreliable, and there could have been a selective tendency among women with advanced breast cancer to record that diagnosis as the cause of death among HRT users, and some other terminal illness (e.g., bronchopneumonia) as the cause among non-users.

Seventhly, under causal assumptions it is virtually inconceivable that HRT can cease to have an effect within 1 yr of discontinuation. The most extreme example which illustrated this point was the highest relative risk estimate identified in this study for the current use of estrogen plus progestin for ≥10 yr of 2.31; yet that elevated risk had entirely disappeared within 1 yr of discontinuation. It is difficult to conceive of any explanation other than bias for the disparate findings for current and recently discontinued use.

Eighthly, the significantly elevated relative risk estimate of 1.45 for current tibolone use also pointed to the likelihood of bias. Tibolone blocks the action of estrogen, and it has not been suspected as a possible cause of breast cancer.

Further additional shortcomings in the MWS included failure to record switches in HRT use over time, failure to record crossovers in exposure status after recruitment, and hence, misclassification of current exposure and of non-exposure after recruitment, as well as misclassification of the type and duration of HRT use. When these defects were pointed out in correspondence the authors' response was that "…after a careful review of the correspondents' comments, we considered that none invalidates our main conclusions" *(3)*. They did not give their reasons.

There were also numerous errors in the MWS report, including discordant confidence intervals in the abstract, text, or figures for current HRT use, past HRT use, death, current tibolone use, equine estrogen plus medroxyprogesterone use for less than 5 yr (for which the relative risk estimate was also discordant); faulty arithmetic (Table 1); and incorrect designation of estradiol as ethinyl estradiol. Only two of these errors have subsequently been acknowledged *(14)*. In an important paper destined to have a major impact on medical practice, such a large number of errors is unacceptable, and their existence must raise doubts about the care with which this study was conducted and analyzed by the MWS investigators, and evaluated by the journal editors and its reviewers.

One view that has been advanced is that, while the MWS must be acknowledged to have defects, none of them explain why the relative risk for current estrogen plus progestin use (2.00) should be significantly higher than for unopposed estrogen use (1.30). There is a ready explanation: the MWS only recorded the name, and duration of use, of the most recent HRT used. Over the years anxiety about the possibility that HRT might cause breast cancer has steadily increased, and since combination therapy was the most recent arrival, it is likely that bias, as described above, would have been strongest for that combination. In addition, that bias could have been further reinforced by reports that suggested that combination therapy increases the risk of breast cancer more than do estrogens alone *(11)*.

In my view the multiple defects in the MWS are sufficient to disqualify the findings as evidence to suggest an increased risk of breast cancer in HRT users. And at minimum, it is certainly the case that, based on the MWS findings, it is not possible to distinguish between bias and causation as alternative explanations for the observed associations.

Overall Conclusions

Concerning Breast Cancer Risk

One of the claims made by the authors of the CR, WHI, and MWS is that the findings in the three studies broadly accord with each other. In fact, there was major discordance among the three studies. In the CR, for unopposed estrogens an increased risk of breast cancer remained evident for 5 yr after discontinuation of use. In the MWS such an effect was no longer evident within 1 yr of stopping. And in the WHI, based on 6.8 yr of follow-up, it has recently been reported that for current users of conjugated estrogens alone there was no evidence of an increased risk *(22)*: if anything, the risk may actually have been decreased (hazard ratio, 0.77; 95% confidence interval 0.59–1.01).

The WHI findings for conjugated estrogens have clearly troubled the WHI investigators since there was a lack of consensus about stopping the study. Perhaps of this reason, the absence of an association with breast cancer was dismissed in an accompanying editorial as "…best interpreted, for now, as due to chance" *(10)*. Yet in statistical terms the upper 95% confidence limit of 1.01 constituted evidence to suggest that there was virtually no increase in the risk that could be attributed to unopposed estrogen. In addition, whereas 44.4% and 6.8% of the combination HRT and placebo recipients, respectively, were unblinded, the corresponding proportions for estrogen alone were remarkably low, at 1.9% and 1.5%. One reason for the virtual absence of unblinding in both comparison groups would have been because the estrogen-only arm of the WHI was confined to hysterectomized women, among whom there would have been no postmenopausal bleeding. And the effect of the low unblinding rate would have been to substantially reduce any likelihood of detection bias.

For combination therapy, the findings among the studies were again discordant. In the WHI there was no increase in the risk of breast cancer among HRT recipients during the first 2 yr of follow-up. In the MWS there was an immediate increase. In the CR the data on combination therapy, although interpreted as being consistent with the WHI findings, were limited, and non-significant. In addition, in the latter study it is likely that most women who were current combination HRT users, and had also used HRT in the past, had previously used estrogens alone.

For any number of reasons minor sources of discordance can be expected in a range of epidemiological studies, but when they are multiple, and major—as is the case here—they must raise further doubts about the validity of the evidence, and whether the associations documented to date have indeed established that HRT, or any specific form of HRT, increases the risk of breast cancer: perhaps it does, perhaps it does not, but the epidemiological microscope quite simply lacks the power to resolve that question.

Finally, since breast cancer is 100 times more common in women than in men, it is clear the female hormones must indeed play a critical role in the etiology of breast cancer. If that role is to be understood, however, it becomes crucially important to study hormone levels, and their determinants, within the breast itself. Yet few such studies have been done, probably because they are difficult. Until they are done it is doubtful whether the question of whether HRT increases the risk of breast cancer by 1.5- or 2-fold can be resolved.

Cardiovascular Outcomes

Of the studies reviewed here only the WHI had information on cardiovascular outcomes, and three of those outcomes, coronary heart disease, stroke, and pulmonary embolism were the ones included in the global index which, together with breast cancer, accounted for the claim that, on balance, the net effect of combination HRT was to confer overall harm. In the WHI estrogen/progestin arm the hazard ratios for coronary heart disease, stroke and pulmonary embolism were 1.29, 1.41, and 2.13, respectively, and each of them was statistically significant. In the later publication of the WHI 6.8-yr follow-up results for conjugated estrogens only *(22)*, the corresponding estimates were 0.91, 1.39, and 1.34, and only the estimate for stroke was significant. The study was terminated because of the increased risk of stroke.

Critique

For the evaluation of combination therapy the investigators once again failed to consider the possibility that the findings could be accounted for by detection bias. While clear-cut or severe myocardial infarction, stroke, or pulmonary would indeed usually be diagnosed, there is ample opportunity for differential detection of "silent," mild, or atypical myocardial infarction (e.g., "indigestion"), stroke (e.g., mild hemianesthesia with recovery), or pulmonary embolism (e.g., cough; syncope), with selectively greater detection occurring among HRT recipients.

Actions taken by the WHI investigators also made it inevitable that there would be detection bias: about 3.5 yr after the study commenced, and again a year later, all participants were warned of a possible increase in the risks of myocardial infarction, stroke, and pulmonary embolism, in combination HRT recipients. I have stated elsewhere *(18)*: "In any follow-up study, it is doubtful whether findings for outcomes that are susceptible to detection bias can be interpreted after the participants, who are aware that they are receiving the drug at issue, are also explicitly and repeatedly warned that their exposures may increase their risk of sustaining the outcomes at issue" (p. 307). Such blatant bias, in my view, nullified the interpretability of any data collected after the warnings were issued.

How great would detection bias have to have been in order to account for the results? If the incidence rates in the women exposed to placebo are subtracted from the rates in the women exposed to estrogen plus progestin, bias could have accounted for the associations if it augmented the detection of coronary heart disease, stroke, and pulmonary embolism by 0.7, 0.8, and 0.8 per 1000 per year, respectively. For such small differences it is again impossible to distinguish between bias and causation as alternative explanations for the observed associations. Indeed, the data in the WHI do not exclude the possibility that the bias could have been so strong that it could actually have concealed some degree of protection against the risk of coronary heart disease, stroke, or pulmonary embolism. For the latter outcome, however, that likelihood is the weakest since the hazard ratio was 2.13, and since multiple studies have documented increased risks of venous thromboembolism among HRT users *(4)*.

In the later report on the use of unopposed conjugated estrogens, the findings were markedly at odds with those reported for combination therapy, except for the increased risk of stroke. One partial explanation may be greater statistical instability in that study: while 8506 and 8102 women, respectively, received combination therapy or placebo, the corresponding numbers for estrogen only and placebo were 5310 and 5429. However, if statistical variability does not account for the discrepant results, the findings for conjugated estrogens must be considered more valid than those for combination therapy because of the reduced likelihood of detection bias due to unblinding.

The study of estrogen only was discontinued on ethical grounds after an average follow-up of 6.8 yr because of the increased risk of stroke. It is clear that the decision to do so was a difficult one, since there was lack of consensus about the advisability of stopping the trial. Given the evidence, therefore, all that is at issue with regard to possible adversity is the relative risk of 1.39 for stroke. That finding was of limited interpretability since at the time of termination fully 53.8% of the women had stopped taking the assigned medication, and crossover rates were 5.7% and 9.1%, respectively, in the women on active treatment and placebo. Once again it is impossible to discriminate between bias and causation as alternative explanations for the findings.

Overall Conclusions

Contrary to what has been claimed, and widely believed, based on the three studies reviewed here, it has not been established that combination HRT increases the risk of breast cancer, coronary heart disease, or stroke. Such therapy may or may not do so, but with the epidemiological methods at our disposal, and the small relative risks at issue, that question cannot be resolved. Not even the possibility of a reduced risk in HRT recipients has been ruled out. With regard to pulmonary embolism, the relative risk was again small, but higher than for the other outcomes assessed in the WHI; in addition that association is perhaps more plausible because it has been documented in other studies.

Also contrary to what has been claimed, there is no good evidence to suggest that unopposed estrogens increase the risk of breast cancer. Indeed, the existing evidence is even contradictory. The MWS findings have been interpreted as clearly indicating that estrogens increase the risk, if to a lesser degree than does combination therapy; yet in the WHI data there is no evidence of an increased risk.

And still further, the claim that the WHI has established that the overall adverse effects of combination HRT exceed the benefits is not tenable, since the findings for all of the adverse outcomes forming the basis for that claim are suspect.

It remains finally to ask why, their defects notwithstanding, the findings from the three studies, especially those concerning breast cancer, have been accepted so widely as having established adversity beyond any doubt? One reason may be that there is a widespread failure to appreciate that, when randomized controlled trials are conducted on the population scale, they may cease either to be randomized or to be "blind." When such studies last many years, high unblinding rates, losses to follow-up, and crossovers may effectively nullify the advantages of randomization and blinding and result in confounding and bias. That is, randomized controlled trials of the type considered here take on the characteristics of observational studies, and they must be interpreted accordingly, with full recognition of the limitations intrinsic to such research.

A further reason for the wide acceptance of questionable results, illustrated by the CR and the MWS, is failure to appreciate that massive numbers and robust statistical power do not ipso facto translate to validity. Given the presence of bias, the only effect of massive numbers is to increase statistical significance, and to set narrower confidence limits around the magnitude of associations that are biased.

Perhaps the most important lesson to be drawn from these studies is that the major strength of epidemiology is in the documentation of large risks. When risks are small, any purported causal associations should be interpreted with caution. It may be that such associations, if "real," may be of public health importance, and it is understandable that we should wish to find out. However, public health importance does not translate to validity, and for small relative risk estimates it may simply not be possible to discriminate between causation and bias.

References

1. Banks, E., Beral, V., Cameron, R., et al. (2001). *Breast Cancer Research*, http://breast-cancer-research.com/content/4/1/R1.
2. Barrett-Connor, E. (1991). *Ann. Int. Med.* **115,** 455–456.
3. Beral, V., Banks, E., Reeves, G., and Bull, D. (2002). *Lancet* **362,** 1330–1331.
4. Castellague, J., Perez-Gutthann, S., and Garcia Rodriguez, L. A. (1998). *Drug Safety* **18,** 117–123.
5. Chlebowski, R. T., Hendrix, S. L., Langer, R. D., et al. (2003). *JAMA* **289,** 3243–3253.
6. Collaborative Group on Hormonal Factors in Breast Cancer. (1997). *Lancet* **350,** 1047–1959.
7. Greendale, G. A., Reboussin, B. A., Sie, A., et al. (1999). *Ann. Int. Med.* **362,** 419–427.
8. Grodstein, F., Newcomb, P. A., and Stampfer, M. J. (1999). *Am. J. Med.* **106,** 574–582.
9. Hill, D. A., Weiss, N. S., Beresford, S. A.A ., et al. (2000). *Am. J. Obstset. Gynecol.* **183,** 1456–1461.
10. Hulley, S. B. and Grady, D. (2004). *JAMA* **291,** 1769–1771.
11. Magnusson, C., Baron, J. A., Correia, N., Bergstrom. R., Adami, H.-O., and Persson, I. (1999). *Int. J. Cancer* **81,** 339–344.
12. Manson, J. E., Hsia, J., Johnson, K. C. et al. (2003). *JAMA* **289,** 3243–3253.
13. Million Women Study Collaborators. (2003). *Lancet* **362,** 419–427.
14. Million Women Study Collaborators. (2003). *Lancet* **362,** 1160.
15. Morrison, A. S., Brisson, J., and Khalid, N. (1988). *J. Natl. Cancer Inst.* **80,** 1540–1547.
16. Shapiro, S., Slone, D., Kaufman, D. W. et al. (1980). *N. Engl. J. Med.* **303,** 485–489.
17. Shapiro, S. (2000). *Am. J. Epidemiol.* **151,** 939–945.
18. Shapiro, S. (2003. *Climacteric* **6,** 302–310.
19. Shapiro, S. (2004). *Climacteric* **7,** 3–7.
20. Stampfer, M. and Colditz, G. (1991). *Prev. Med.* **20,** 47–63.
21. Tang, M. X., Jacobs, D., Stern, Y., et al. (1996). *Lancet* **348,** 429–432.
22. The Women's Health Initiative Steering Committee. (2004). *JAMA* **291,** 1701–1712.
23. Weiss, N. S., Ure, C. L., Ballard, J. H., et al. (1980). *N. Engl. J. Med.* **303,** 1195–1198.
24. Writing Group for the Women's Health Initiative Investigators. (2003). *JAMA* **228,** 321–333.