Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies

Part 3. The Women's Health Initiative: unopposed estrogen

Samuel Shapiro, Richard D T Farmer, Alfred O Mueck, Helen Seaman, John C Stevenson

Abstract
Background Studies from the Women's Health Initiative have reported an increased risk of breast cancer in users of estrogen plus progestogen. Among users of estrogen alone an increased risk was not observed.
Objective To evaluate the evidence for unopposed estrogen.
Methods In a related article (Part 2) the authors apply generally accepted causal criteria to the findings for estrogen plus progestogen. Here (Part 3) the authors apply the criteria to the findings for unopposed estrogen, as reported in a clinical trial, and in combined data from the trial and an observational study.
Results In the clinical trial, after 7.1 years of follow-up the relative risk (RR) of invasive breast cancer for women assigned to estrogen was 0.77 in an ‘intention-to-treat’ analysis (95% CI 0.59–1.01) and 0.67 (95% CI 0.47–0.97) in an ‘as treated’ analysis; after 10.7 years the risk reduction persisted. Time order was correctly specified; detection bias was minimal; in the ‘as treated’ analysis confounding was unlikely; duration-response and internal consistency could be evaluated only to a limited extent because of scanty data; the findings were discordant with increased risks observed in the Collaborative Reanalysis and the Million Women Study; biological plausibility could not be assessed. In the combined analysis, among women who had previously used estrogen soon after the menopause there was no clear evidence of either a reduction or an increase in the risk of breast cancer among women assigned to estrogen during the trial, or among women who were using estrogen in the observational study when follow-up commenced. The combined analysis did not satisfy the criteria of time order, bias, confounding, statistical stability and strength of association, duration-response, and internal consistency; biological plausibility could not be assessed.
Conclusions The evidence from the clinical trial suggests that unopposed estrogen does not increase the risk of breast cancer, and may even reduce it. The latter possibility, however, is based on statistically borderline evidence.

Background
In Part 1 of this series of articles we have evaluated the effect of hormone replacement therapy (HRT) on the risk of breast cancer, as reported in the Collaborative Reanalysis (CR), and in Part 2 the effects of estrogen plus progestogen (E+P), as reported in the Women’s Health Initiative (WHI) clinical trial and observational study. We concluded that the studies did not accord with generally accepted epidemiological principles of causation. Here, in Part 3 we apply causal principles to WHI reports on the effect of estrogen therapy (ET) without an added progestogen on the risk of breast cancer.

The Women’s Health Initiative clinical trial: estrogen vs placebo

The WHI trial commenced in 1993, and 5310 and 5429 women, respectively, were randomly assigned to conjugated estrogen, 0.625 mg, or a placebo. The assignment was ‘double-blind’. Originally hysterectomised and non-hysterectomised women...
were included, but because another trial reported an increased risk of endometrial hyperplasia in estrogen users,21 women who were not hysterectomised were ‘unblinded’ and re-allocated to E+P.15 There was a continuing increase in the risk of stroke (not considered here),15 and the trial was terminated after an average of 6.8 years of follow-up, apparently on safety grounds, but not on the recommendation of the Data and Safety Monitoring Board.

First report13
In the first report the risk of several outcomes in relation to ET exposure was evaluated. Here consideration is confined to the risk of breast cancer.

At the time the trial was terminated 1.9% and 1.5% of the ET and placebo recipients, respectively, had been ‘unblinded’ (our calculation). Discontinuation rates were virtually identical, and the overall rate was 53.8%. Among the ET and placebo recipients 5.7% and 9.1%, respectively, were prescribed HRT by their own doctors.

In an ‘intention-to-treat’ (ITT) analysis the hazard ratio (HR) for invasive breast cancer was 0.77 (95% CI 0.59–1.01), and “this comparison narrowly missed statistical significance (p = 0.06)”. The authors commented that “the trend toward a reduction in breast cancer incidence was unanticipated and … opposite to that observed in the WHI [E+P] trial … [as well as] … contrary to the preponderance of observational study results, including those from the … [MWS]”.

Second report14
This report was focused specifically on breast cancer. The average duration of follow-up was 7.1 years. In an ITT analysis the HRs for ET recipients were as follows: all breast cancers, 0.82 (95% CI 0.65–1.04); invasive breast cancer, 0.80 (95% CI 0.62–1.04); in situ breast cancer, 0.86 (95% CI 0.51–1.46). In an ‘as treated’ analysis the HR for invasive cancer was 0.67 (95% CI 0.47–0.97; p = 0.03). There was no significant evidence of a duration effect (trend p = 0.29). Invasive cancers were larger in the ET than in the placebo recipients: 1.8 vs 1.2 cm (p = 0.03), and localised disease was less common among the former (HR 0.69; 95% CI 0.51–0.95). The respective proportions of abnormal mammograms that necessitated further investigation in the ET and placebo recipients were 36.2% and 28.1% (p<0.001).

The authors stressed that the findings in subgroup analyses needed to be interpreted cautiously, and they alluded to the discordance with increased risks observed for ET users observed in some earlier observational studies2 and in the MWS.18 They concluded that “treatment with [ET] alone does not increase breast cancer incidence in postmenopausal women with hysterectomy”.

Third report15
Several outcomes were evaluated in this report, and again consideration is confined to the findings for breast cancer. Following termination of the trial after an average follow-up of 7.1 years (intervention phase) the women continued to be followed in a post-intervention phase, part of which extended beyond the termination date specified in the study protocol. For the interval beyond that date 77.9% of the surviving participants consented to be followed. Overall, the mean duration of follow-up was 10.7 years.

In ITT analyses the respective HRs in the intervention phase, the post-intervention phase and overall, were 0.79 (95% CI 0.61–1.02), 0.76 (95% CI 0.61–1.09) and 0.77 (95% CI 0.62–0.95). When the data were censored 6 months after becoming non-adherent to treatment (i.e. and ‘as treated’ analysis) the overall HR was 0.68 (95% CI 0.49–0.95). The risk reduction was consistently evident when the data were stratified by decade of age.

The authors concluded that with more prolonged follow-up the “decreased risk of breast cancer persisted”.

Evaluation
Below we evaluate whether the evidence in the clinical trial accorded with generally accepted principles of causality.12–14 The principles are inter-related, and when appropriate we cross-refer.

Time order
At baseline the mammograms of all participants were free of cancer, and the criterion of time order was satisfied.

Information bias
This was a prospective study and information bias was unlikely.

Detection bias
Without any question the ET trial was less biased than the E+P trial, in which the respective ‘unblinding’ rates among the E+P and placebo recipients were 44.4% and 6.7%.4 By contrast, in the ET trial the rates were 1.9% and 1.5%.15 The major reason for these striking differences, of course, was that the ET-exposed and non-exposed women were hysterectomised, vaginal bleeding did not occur, and ‘unblinding’ was seldom necessary – although why the ‘unblinding’ rate among the placebo recipients in the ET trial (1.5%) was lower than in the E+P trial (6.8%) is not clear.

Some minimal bias may perhaps have occurred among women who suspected that they were receiving ET because they developed enlarged or tender breasts, and the finding that abnormal mammograms necessitating further investigation were more common in the ET recipients supports that possibility.16 Alternatively, the more common need for investigation in the ET recipients may have been due to increased breast tissue density22 – which may also explain the larger size of the invasive cancers, as well as the less common occurrence of localised disease among the ET recipients.

To the extent that detection bias may have been present, its effect would have been to underestimate...
the magnitude of the observed reduction in the risk of breast cancer among ET recipients. In effect, in respect of ‘unblinding’ the ET study remained a controlled trial.

Confounding
In respect of confounding the ET study did not remain a controlled trial, but became an observational study. Among the 53.8% of participants who stopped their allocated treatments, the reasons for stopping could have confounded the findings, and additional confounding could have occurred after stopping. In addition, confounding could have occurred among the ET and placebo recipients prescribed HRT by their own doctors. Had the discontinuation rate approximated, say, 10%, ITT analysis could conceivably have reduced confounding to some extent. However, since more than half the women stopped their treatments, an ITT analysis of what were essentially observational data made no sense.

For these reasons the ‘as treated’ analysis was the most valid analysis. In that analysis there was a 33% reduction in the risk of invasive breast cancer among the ET recipients \( (p = 0.03) \), and the reduction was still evident after 10.7 years of follow-up (HR 0.68; 95% CI 0.49–0.95). That reduction, however, must be cautiously interpreted, since it was statistically borderline and of low magnitude, and uncontrolled confounding could have accounted for it (see: Statistical stability and strength of association).

The authors stated that the “results [for invasive breast cancer] were not altered by adjusting for the small differences in the number of first-degree relatives with breast cancer or history of benign breast disease”, Why those two factors, but not 11 additional factors listed in their Table 2, such as age at first birth or age at hysterectomy, were the ones allowed for was not explained. In any event, however, as would be expected following randomisation, the distributions of all the potential confounders were similar in the two comparison groups, and it is unlikely that there was significant uncontrolled confounding in the ‘as treated’ analysis.

Statistical stability and strength of association
The lowest documented HR was 0.67 (‘as treated’ analysis; invasive cancer), the upper 95% confidence limit was 0.97, and the \( p \) value was 0.03. That is, the association was only of borderline statistical significance, it was identified in a subgroup analysis, and it should be interpreted cautiously. In addition, the 1.49-fold risk reduction (the inverse of the HR estimate of 0.67) was “small”, and it could possibly have been accounted for by minimal bias or confounding (see above). If present, such bias or confounding could have persisted after termination of the clinical trial, and possibly have explained the statistically significant risk reduction after 10.7 years of follow-up.

Duration-response
There was no significant duration trend. However, in the ‘as treated’ analysis the risk reduction for invasive breast cancer commenced after about 2 years of follow-up, and it became more marked between Years 2 and 7 (Figure 2 in Reference 16). The trend was not commented on, perhaps because it was not significant \( (p = 0.09) \).

Internal consistency
Statistical power within some subgroups was limited, but to the extent that consistency could be evaluated, the findings were broadly consistent within relevant strata [e.g. age, body mass index (BMI), history of benign breast disease, family history of breast cancer].

External consistency
The findings were discordant with the increased risk of breast cancer among ET users observed in the CR and in the MWS. The MWS investigators suggested that the discordance may have occurred because the American participants in the WHI trial were more obese than the British participants in the MWS. That suggestion was not supported by the data. In the WHI trial the distribution of BMIs in the ET-exposed and non-exposed women were similar, and the HRs were <1.0 in non-obese (BMI<25), moderately obese (BMI 25–29.9) and severely obese women (BMI≥30). A more plausible explanation for the discrepancy was the absence of detection bias in the ET clinical trial, and its presence in the CR and MWS.

Biological plausibility
Some of the experimental evidence is compatible with the hypothesis that estrogen alone may accelerate the onset of clinically detectible breast cancer, while other evidence suggests that it may have the opposite effect. It is also possible that different estrogens may have different effects, and at least 10 different estrogenic compounds, in varying concentrations, are present in conjugated equine estrogens. With regard to potential carcinogenicity, two main mechanisms have been proposed, the first being proliferative effects of estrogens on pre-existing estrogen-sensitive cancer cells. The second possible mechanism may be excessive metabolism of estrogens to highly active compounds having strong proliferative effects, even at low concentrations. Such metabolites could also be genotoxic, resulting in new cancer cells. A qualification to both mechanisms, however, is that until a clone reaches the size of about 10^9 malignant cells (a tumour of about 1 cm in diameter) breast cancer is seldom clinically detectible. Based on what is known about the doubling times of the most aggressively multiplying cells that process would take at least 10 years. With regard to a possible reduction in the risk of breast cancer, over the course of a decade or longer other mechanisms could operate by destroying cancer.
cells before clinically detectable breast cancer develops; it has been demonstrated that estrogens have anti-proliferative and pro-apoptotic effects. The latter mechanisms have even been invoked as a rationale for the treatment of breast cancer.

There is still a further paradox. It has been shown that estrogens can be metabolised not only to potentially genotoxic metabolites, but also to carcinoprotective metabolites, such as 2-methoxy-estradiol. If the predominant overall effect is for estrogens to up-regulate those mechanisms that destroy proliferating cells before they develop into clinically detectable breast cancer, the net effect could be a reduction in the risk. Alternatively, however, depending on which mechanisms predominate, the effect could be no risk reduction or an increased risk.

Combined data from the WHI clinical trial and observational study

First report
In this report the clinical trial data were restricted to a “sub cohort” of women whose date of onset of the menopause was known (ET 4493; placebo 4596), and the observational data comprised a “sub cohort” of 4493 ET users and 8101 non-users with the same restrictions. Allowance was made for confounding in the observational data.

During follow-up the incidence rates of breast cancer in the ET-exposed and non-exposed women were higher in the observational study than in the clinical trial, both among women who had and had not previously used HRT. After “control for prior use of [HRT] and for confounding factors, … HR estimates [for ET-exposed women] were higher from the observational study compared with the clinical trial by 43% (p = 0.12). However, after additional control for [the elapsed time from onset of the menopause to first use of ET] the HRs agreed closely between the two cohorts (p = 0.82). For women who [began HRT] use soon after menopause, combined analysis of the clinical trial and observational study data [did] not provide clear evidence of either an overall reduction or an increase in breast cancer risk with [ET]."

Evaluation
Because of failure to estimate breast cancer risk according to reasons for non-eligibility or refusal to participate in the WHI clinical trial, the validity of the observational data cannot be fully assessed. To the extent feasible, below we apply causal criteria to the evidence from the combined analysis.

Time order
In the observational study the women were not screened for the presence of breast cancer at the time of recruitment. ET users aware of as yet undiagnosed breast lumps could selectively have consented to be followed because they were worried, but were unwilling to participate in an experiment (see: Detection bias).

Detection bias
In the observational study, all ET users and non-users were aware of their exposure status, whereas in the clinical trial over 98% of the women remained “blinded.” Thus, in the observational study detection bias was present, whereas in the clinical trial it was virtually absent. In the observational study bias would have been especially marked among ET users who had declined to participate in an experiment, or who were ineligible, but who nevertheless agreed to be followed. That bias would have been further reinforced at recruitment, when the women were informed that one objective of the WHI study was to evaluate the risk of breast cancer in HRT users. Bias would have been still further reinforced when the trial was terminated, and the women were informed of an increased risk of breast cancer in HRT users in writing, and when the increased risk was also given extensive publicity. The higher incidence rates of breast cancer in the observational study than in the clinical trial, both among women who had and had not previously used HRT, was quantitative evidence to support the likelihood of detection bias.

The WHI investigators argued that since the two studies were drawn from the same populations, and over essentially the same time periods, it was legitimate to combine them. They also argued that since there was good overall agreement between the two studies after allowing for the time lapse from the menopause to the time of first use of ET, and for duration of use among adherent women, the evidence suggested an “absence of important bias due to a woman’s knowledge of her hormone therapy exposure”.

Two populations, one of which comprised women who consented to participate in a ‘double-blind’ randomised controlled trial, and the other which comprised women who refused to participate in the trial, or were ineligible, cannot be regarded as ‘the same’. And contrary to what was claimed, good agreement was not shown between the clinical trial and observational study: virtually all subgroup comparisons in the combined analysis were based on sparse data, the 95% CIs were wide, and the findings, including the findings among women who had previously used HRT soon after the menopause, were compatible with substantial disagreement (see: Statistical stability and strength of association). Moreover, if for the sake of the argument the sparsity of the data is set aside for the moment, it would have taken relatively little bias – much less than could confidently have been excluded – to account for the higher incidence rates of breast cancer, both among ET-exposed and non-exposed women, in the observational data.

Confounding
Adjustment was made for confounding in the observational data, but not in the clinical trial data.

Statistical stability and strength of association
In subgroup comparisons, in Tables 2, 4, 5 and 6 of the report 32 HRs were estimated, virtually all of them
based on small numbers, and in all but one of them 95% CIs included 1.00. The single exception was a HR of 0.58 (95% CI 0.36–0.93) for ET exposure that commenced >5 years after the menopause. Among the remaining 31 HRs the lowest was 0.63 and the highest was 1.63. The associations might readily have been due to chance. In addition, as pointed out above, such low-magnitude associations could also have been due to bias or confounding.

**Duration-response**

Duration of ET use was not evaluated in this study.

**Internal consistency**

There were insufficient data to evaluate consistency within relevant strata, such as BMI, or a history of benign breast disease.

**External consistency**

In this study the evidence to suggest a reduced risk, increased risk, or no effect of ET was ambiguous, and the findings were discordant with the clinical trial evidence, which suggested no increase, and possibly a decrease, in the risk of breast cancer.16

**Biological plausibility**

For the reasons given above in the evaluation of the clinical trial findings, biological plausibility cannot be assessed.

**Conclusions**

In the clinical trial bias was minimal, and in the ‘as treated’ analysis major confounding was unlikely. Apart from the criterion of biological plausibility (which could not be assessed) the trial otherwise satisfied all but one of the criteria of causality (duration-response). By contrast, the combined analysis of the clinical trial and observational data failed to satisfy the criteria of time order, bias, statistical stability and strength of association, duration-response, internal consistency, and external consistency.

The clinical trial findings, although limited in some respects because of sparse data, are the best evidence produced to date, and they suggest that ET without an added progestogen does not increase the risk of breast cancer. That evidence is statistically robust. ET may even reduce the risk, but the evidence to support that possibility is statistically fragile. A possible reduction in the risk of breast cancer must be regarded as tentative, and in need of confirmation – and a further controlled trial may be needed, since any observational study is likely to be biased.

Whether or not unopposed ET reduces the risk of breast cancer, the evidence in the clinical trial suggests that ET does not increase the risk. That evidence has implications for the validity of the CR,2 the WHI,4–11 and the MWS.19 Since those studies were biased13,37 it is likely that they overestimated the risk of breast cancer in E+P-exposed women.

Finally, as a cautionary note, evidence from a single study can never be regarded as conclusive, and it remains possible that unopposed ET increases the risk of breast cancer. The best evidence, however, suggests that it does not.

**Competing interests**

Samuel Shapiro, Alfred Mueck and John Stevenson presently consult, and in the past have consulted, with manufacturers of products discussed in this article. Richard Farmer has consulted with manufacturers in the past.

**Provenance and peer review**

Not commissioned; externally peer reviewed.

**References**


Shapiro et al.