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

Part 2. The Women's Health Initiative: estrogen plus progestogen

Samuel Shapiro,[1] Richard D T Farmer,[2] Alfred O Mueck,[3] Helen Seaman,[4] John C Stevenson[5]

[1]Visiting Professor of Epidemiology, Department of Epidemiology, University of Cape Town, Cape Town, South Africa
[2]Emeritus Professor of Epidemiology, Department of Epidemiology, University of Surrey, Guildford, UK
[3]Professor of Clinical Pharmacology and Experimental Endocrinology, Department of Endocrinology, University Women's Hospital, Tübingen, Germany
[4]Freelance Medical Writer, Aldershot, UK
[5]Consultant Physician and Reader in Metabolic Medicine, National Heart and Lung Institute, Imperial College, London and Royal Brompton Hospital, London, UK

Correspondence to
Professor Samuel Shapiro, Department of Public Health and Family Medicine, University of Cape Town Medical School, Anzio Road, Observatory, Cape Town, South Africa; samshap@mweb.co.za

Abstract

Background Based principally on findings in three studies, the Collaborative Reanalysis (CR), the Women's Health Initiative (WHI), and the Million Women Study (MWS), it is claimed that combined hormone replacement therapy (HRT) with estrogen plus progestogen is now an established cause of breast cancer. For unopposed estrogen therapy the evidence in the three studies is conflicting: the CR and MWS have reported increased risks in estrogen users, while the WHI has not. The authors have previously reviewed the findings in the CR (Part 1).
Objective To evaluate the evidence for causality in the WHI studies.
Methods Using generally accepted causal criteria, in this paper (Part 2) the authors evaluate the findings in the WHI for estrogen plus progestogen; in a related paper (Part 3) the authors evaluate the findings for unopposed estrogen. An evaluation of the MWS (Part 4), and of trends in breast cancer incidence following publication of the WHI findings in 2002 (Part 5) will follow.
Results For estrogen plus progestogen the findings did not adequately satisfy the criteria of bias, confounding, statistical stability and strength of association, duration-response, internal consistency, external consistency or biological plausibility.
Conclusion HRT with estrogen plus progestogen may or may not increase the risk of breast cancer, but the WHI did not establish that it does.

Background

Based on reports from the Collaborative Reanalysis (CR),[1] the Women's Health Initiative (WHI) clinical trial,[2] and the Million Women Study (MWS)[3] published, respectively, in 1997, 2002 and 2003, it is now widely believed that hormone replacement therapy (HRT) causes breast cancer. More specifically, all three studies have reported an increased risk for the use of estrogen plus progestogen (E+P), and the CR and MWS have also reported an increased risk for the use of estrogen therapy (ET). However, in a WHI clinical trial of ET versus placebo, published in 2004,[4] the risk was not increased.

Following publication of the initial WHI report in 2002[2] there was a decline in the use of HRT, and it was claimed that there was a corresponding decline in the incidence of breast cancer.[5]

In Part 1 of this series we evaluated the CR report,[6] and concluded that it did not accord with generally accepted epidemiological principles of causation.[7–9] Here, in Part 2 we apply the principles to the WHI evidence implicating E+P, firstly as reported in the clinical trial,[2 10–13] and then as reported in the clinical trial data combined with data from a WHI observational study that commenced at the same time.[14–16] In a related article (Part 3) we evaluate the WHI studies of ET.[17] In future articles we will evaluate the MWS findings (Part 4), and the purported secular decline in the incidence of breast cancer following the decline in the use of HRT (Part 5).

The Women's Health Initiative

In 1993, several studies were initiated in 40 centres in the USA under the rubric of the WHI.[18] In two clinical trials and in one cohort study the benefits and risks associated with the use of E+P or ET were evaluated, and one objective was to assess the risk of breast cancer.

Article

In the E+P trial menopausal women were randomly assigned to conjugated equine estrogen, 0.625 mg per day, plus medroxyprogesterone acetate, 2.5 mg per day, or a placebo.[2] In the ET trial the assignments were to conjugated equine estrogen, 0.625 mg per day, or a placebo (ET).[2 4] The assignments were 'double-blind'. Initially, in both trials women were included whether or not their uterus had been removed. However, because another study reported an increased risk of endometrial hyperplasia,[19] "the WHI protocol was changed to randomise women with a uterus to only [E+P] or placebo in equal proportions. The 331 women previously randomised to unopposed [ET] were unblinded and reassigned to [E+P]".[2] In the observational study women originally approached for inclusion in the two trials, but not included because they were ineligible or because they declined, were followed.[18]

In the clinical trials, unless otherwise stated, intention-to-treat (ITT) analysis was used.

Clinical trial: estrogen plus progestogen vs placebo

First report[2]

Menopausal women aged 50–79 years were randomly assigned to E+P (8506 women) or a placebo (8102 women). The allocation was 'double-blind', except for the 331 (3.9%) women initially assigned to ET and reassigned to E+P. The women were interviewed every 6 months, and they attended study clinics for annual breast examinations and mammography. The trial was terminated in 2002 after an average of 5.2 years of follow-up because of a statistically significant adverse 'global index' comprising several outcomes (not considered here) and an increased risk of invasive breast cancer.

When follow-up ended 40.5% of the E+P recipients had had their treatments 'unblinded', primarily because of persistent vaginal bleeding, and with the additional reassigned ET recipients (3.9%) the total 'unblinding' rate was 44.4%; 6.8% of the placebo recipients were 'unblinded'. Discontinuation rates were 42% and 38%, respectively, and 6.2% and 10.7% of the two groups were prescribed HRT by their own doctors.

For E+P exposure the hazard ratio (HR) of invasive breast cancer was 1.26 (95% CI 1.00–1.59), a finding that "almost reached nominal statistical significance" while "the weighted test statistic used for monitoring [$z = -3.19$] was highly significant". The HRs during 1–≥6 years of follow-up were 0.62, 0.83, 1.16, 1.73, 2.64 and 1.12, respectively (trend score, $z = 2.56$). The estimated absolute risk attributable to the use of E+P was 8 per 10 000 woman-years.

The authors concluded that "the WHI [was] the first randomised controlled trial to confirm that combined [E+P] *does* [our emphasis] increase the risk of incident breast cancer".

Second report[10]

In an updated analysis specifically focused on breast cancer the mean [standard deviation (SD)] duration of follow-up was 5.6 (1.3) years, and 349 invasive and 84 *in situ* cases were analysed. For all breast cancers the HR was 1.24 (95% CI 1.02–1.50; $p<0.001$); for invasive cancer it was 1.24 (95% CI 1.01–1.54; $p = 0.003$); and for *in situ* cancer it was 1.18 (95% CI 0.77–1.82; $p = 0.09$). Among the E+P recipients invasive cancers were larger than among the placebo recipients [mean (SD) 1.7 (1.1) vs 1.5 (1.5) cm; $p = 0.04$], more commonly node-positive (25.9% vs 15.8%; $p = 0.03$), and had more commonly spread to the regional tissues or metastasised (25.4% vs 16.0%; $p = 0.04$). The *in situ* cancers were also larger (1.6 vs 1.1 cm; $p = 0.33$).

The authors acknowledged that failure to adhere to the assigned treatments was a limitation in their study, but asserted that "the discontinuation of study hormones … [was] … likely to dilute the estimate of effects of [E+P]". They also acknowledged that "because vaginal bleeding led to a high prevalence of *de facto* unblinding, some potential for detection bias [existed]", but asserted that any such bias was likely to be small. They concluded that E+P "increases the risk of incident breast cancers, which are diagnosed at a more advanced stage compared with placebo use".

Third report[11]

Again based on 5.6 years of follow-up, the investigators evaluated whether a history of having used HRT before the trial began influenced the subsequent risk of invasive breast cancer among women assigned during the trial to E+P. Overall, the HR was 1.24 (95% CI 1.02–1.50; $p = 0.003$). After adjustment for age, race, body mass index (BMI), physical activity, smoking, alcohol use, parity, age at first birth, use of oral contraceptives, family history of breast cancer, family history of fractures, mammography use and presence of moderate to severe vasomotor symptoms, the HR was 1.20 (95% CI 0.94–1.53; weighted $p = 0.025$). Among women who had never used HRT previously the HR for those assigned in the trial to E+P was 1.02 (95% CI 0.77–1.36), and among those who had used HRT previously it was 1.96 (95% CI 1.17–3.27), a statistically significant difference ($p = 0.027$). However, in the former group the HR increased significantly during follow-up, from 0.48 in Year 1 to 1.24 in Year 6 (trend $p = 0.02$), while in the latter group the corresponding HRs were 0.90–1.99, and there was no significant trend (trend $p = 0.10$).

The investigators concluded that the absence of an effect among women who had not previously used HRT, and who were assigned during the trial to E+P, "should not be interpreted as overall breast safety [sic] given the statistically significant test for increasing risk with time since randomization". They suggested that "durations only slightly longer than those in the WHI trial are associated with increased breast cancer risk".

Fourth report[12]

Risks of breast cancer for the use of E+P before termination of the study ('clinical trial phase', 1993–2002), after termination ('post-intervention' phase, 2002–2005) and overall (1993–2005) were compared. In the clinical trial phase the HR was 1.26 (95% CI 1.02–1.55), in the post-intervention phase it was 1.27 (95% CI 0.91–1.78). The combined HR was 1.27 (95% CI 1.06–1.51). The authors acknowledged that in the post-interventional phase "health care-seeking behavior and cancer screening practices could have differed".

Fifth report[14]

All 8506 women assigned at recruitment to E+P and 8102 women assigned to placebo were followed through three phases: from 15 November 1993 to 7 July 2002, when the trial was terminated ('intervention phase'); from 8 July 2002 to 31 March 2005, the original termination date specified in the study protocol ('post-intervention phase'); and from 1 April 2005 to 14 August 2009 ('extension phase'). In the extension phase consent to participate was obtained a second time ('reconsent'), and 6545 and 6243 women originally assigned, respectively, to E+P and placebo (78.9% and 77.1% – our calculations) were followed. The mean (SD) duration of follow-up was 11.0 (2.7) years. "Analyses for deaths due to breast cancer among women who did not reconsent were censored on 31 December 2005 … because mortality may be incomplete at more recent times [sic]".

The overall HR for invasive breast cancer in women originally assigned to E+P was 1.25 (95% CI 1.07–1.46), and the HR was consistently elevated in strata of age, BMI, Gail risk score,[20] previous E+P use and duration of use, and time since menopause. There were 25 and 12 deaths directly attributed to breast cancer in the E+P and placebo groups (HR 1.96; 95% CI 1.00–4.04; p = 0.049). Breast cancers were more commonly node-positive in the E+P (23.7%) than in the placebo group (HR 1.78; 95% CI 1.23–2.58; p = 0.03).

The investigators concluded that "breast cancer mortality also appears to be increased with combined use of [E+P]", and that exposure increases the risk of node-positive tumours.

Evaluation

Below we evaluate whether the evidence in the clinical trial accorded with generally accepted principles of causality.[7–9] 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

E+P causes vaginal bleeding, and it was predictable that 'unblinding' would occur more commonly in the E+P than in the placebo recipients, as proved to be the case: 40.5% were 'unblinded' mainly for that reason, and with the addition of the 'unblinded' ET users re-allocated to E+P (3.9%) the total 'unblinding' rate was 44.4%, and 6.5-fold greater than for placebo (6.8%).[2] In addition, since E+P causes breast enlargement and tenderness, additional women who ostensibly remained 'unblinded' could correctly have realised that they were on E+P.

When postmenopausal bleeding occurs it is obligatory to rule out uterine cancer, and in order to do so, to determine whether HRT was used. And 'unblinded' E+P-exposed women would have been told that cancer of the uterus must be ruled out. 'Unblinded' E+P-exposed women would also have become anxious about breast cancer, since at recruitment all participants were informed that one of the study objectives was to assess the risk of that outcome. Diagnostic procedures among E+P-exposed women who bled, some of them uncomfortable or painful (e.g. endometrial biopsy) would have reinforced their anxiety. 'Unblinded' women previously allocated to ET and switched to E+P would also have become anxious. That anxiety would further have been reinforced by the extensive publicity given for many years to the possibility that HRT may cause breast cancer.

Since almost half the E+P recipients were 'unblinded', and since additional women could have realised that they were exposed, there could have been a greater tendency for such women than for placebo recipients to repeatedly examine their breasts, for their medical attendants to do the same, for the WHI personnel to do so when they conducted annual breast examinations, and for their mammograms to be scrutinised more intensively. Some 5% of postmenopausal women have occult ('clinically silent') breast cancer,[21] and there was thus ample scope for the selective detection of such cancers in E+P users.

In one of the WHI reports the authors acknowledged that detection bias could have occurred after the trial was terminated, but they failed to recognise that it could also have occurred beforehand.[12] In another report the investigators mentioned detection bias as a possibility,[10] but asserted that "the amount of bias, if any, [was] likely to be small, based on several factors. First, the WHI achieved very high compliance with annual mammography, which was nearly identical between study groups throughout follow-up. Furthermore, the readings and response to mammographic findings were managed by the women's own physicians, independent of WHI and with no access to study reports, thereby minimising the opportunity for reported bleeding to influence these findings". Those assertions were indefensible, for the following reasons.

First, the statement that there was "very high compliance with annual mammography" was incorrect. The

"high compliance" applied only to women who did not stop their assigned treatments. Those who did stop also stopped receiving annual study mammograms, and the respective mammography rates in the E+P and placebo groups declined from 86% and 90% at 1 year of follow-up, to 48% and 41% at ≥6 years (our calculations: Table 5, Reference 10). 'Unblinded' E+P users who stopped their treatments because they were worried about breast cancer could thereafter have sought mammograms on their own initiative more commonly than placebo recipients who stopped.

Second, doctors consulted by women with postmenopausal bleeding would have demanded to know whether they were receiving E+P, and the doctors would have told their patients.

Third, 'unblinded' women could have told the mammographers that they had been given E+P, and that they were worried. The sensitivity of mammography is limited,[22] and among HRT users about 30% of breast cancers actually present go undiagnosed.[23] Since increased density is a predictor of increased breast cancer risk,[22 23] the search for a tumour could have been most intensive in women who both had dense breast tissue, and who also received E+P.

Fourth, large tumours and tumours that have spread are more readily detectible than small tumours – which could readily explain the larger tumour sizes, and their spread beyond the breast, observed in the E+P recipients.[10]

To a limited extent detection bias might have been reduced in an 'as treated' analysis (see: Confounding) confined to women who remained 'blinded'. Such an analysis has not been published. Instead, in correspondence the investigators "analysed the post 1-year trial data by separately estimating HRs … according to whether or not [the women] experienced persistent bleeding throughout the first year of randomization. [HRs] were elevated ($p<0.05$) both among women with and among women without persistent vaginal bleeding".[24] The data forming the basis for that statement have not been published, but in any case the relevant analysis was not presented: what should have been assessed was whether the HRs were increased in an 'as treated' analysis among women who were not 'unblinded', both during the first year of follow-up, as well as thereafter.

How much bias would it have taken to account for overall HRs for incident breast cancer that ranged from 1.20 to 1.27?[2 10–13] If bias augmented the detection of otherwise occult breast cancer by 0.08% per year (the estimated increase in the incidence was 8 per 10 000 woman-years)[2] it would have nullified the association.

Breast cancer death rates were analysed in the fifth report,[13] and after a mean (SD) of 11 (2.7) years of follow-up the HR was 1.96. However, an analysis of breast cancer deaths 8 or more years after assignment to E+P ended was questionable. In addition, the HR was only of borderline significance ($p = 0.049$) (see: Statistical stability and strength of association), and it

is also likely that it was biased. For women who did not 'reconsent' for follow-up in the extension phase (E+P 15.7%; placebo 10.2%; our calculations: Figure 1 in the report), breast cancer deaths were censored in December 2005, whereas among women who did 'reconsent', deaths continued to be ascertained until August 2009. E+P recipients who 'reconsented' could selectively have done so if they were aware of as yet undiagnosed breast lumps – which could have explained the higher breast cancer death rate, as well as the higher incidence of node-positive tumours among the 'reconsenters'.

In terms of detection bias the WHI clinical trial ceased to be 'double-blind', and in effect it became an observational study.

Confounding

A further reason why the clinical trial became an observational study was because confounding was not adequately controlled: 42% and 38% of the E+P and placebo recipients did not adhere to their treatments, and 10.7% of the latter switched to HRT.[2] Women who stopped or switched could have done so for reasons related to breast cancer risk. For example, 'unblinded' E+P recipients with a family history who stopped could more frequently have had mammograms on their own initiative than those without a family history.

No controlled trial is perfect, and full adherence to treatment is exceptional. For this reason it is conventional to use ITT analysis, in which all participants are assumed to have adhered throughout. It is argued that confounding by the underlying reason for stopping is thereby minimised, that although doing so results in non-differential misclassification of exposure, any observed elevation of the HR is 'conservative', and that in the absence of misclassification the 'true' HR would, if anything, be higher.

Based on this reasoning Chlebowski et al. claimed that "discontinuation of study hormones … [was] … likely to dilute the estimate of effects of [E+P]".[10] However, for ITT analysis to be valid, the assumption of non-differential misclassification must be tenable, and in the WHI trial it was not. The discontinuation rates were exceptionally high, and confounding could commonly have arisen either when treatment was stopped, or thereafter. To this consideration it should be added that the ITT analysis would not have reduced confounding, if present, among the placebo recipients who switched to HRT. Since the trial became an observational study, the analysis should have been confined to women who adhered to treatment ('as treated' analysis) as is standard practice in observational research.

In one report in which confounding was controlled, albeit in an ITT analysis, the HR at 5.6 years of follow-up was 1.20, as against 1.24 when confounding was not controlled.[11] However, the possibility that the adjusted HR might have been more markedly reduced in an 'as treated' analysis was not excluded. Moreover,

factors such as physical activity, smoking, alcohol use, family history of fractures, and moderate to severe vasomotor symptoms can only be imprecisely measured and controlled, and residual confounding could have been present. And if that consideration were insufficient, factors such as age at menopause, or type of menopause, were not controlled at all.

Statistical stability and strength of association

Table 1 gives the HR estimates for E+P versus placebo in the five WHI reports. For the same durations of follow-up the 95% CIs varied: in two reports, after an average follow-up of 5.2 years both HRs were 1.26, but the 95% CIs were 1.00–1.59[2] and 1.02–1.55[12]; in two reports, after 5.6 years both HRs were 1.24, but the 95% CIs were 1.01–1.54[10] and 1.02–1.50.[11] The differences were minor, but the CIs should have been identical. This variation was unexplained, but it was not trivial, since the initial HR estimate of 1.26 (95% CI 1.00–1.59) was interpreted as causal even though it only "almost reached nominal statistical significance",[2] and only the "weighted test statistic used for monitoring" was significant.

In the different reports the overall HRs for incident breast cancer were 1.27 or less,[2 10–13] and since the clinical trial became an observational study, bias and confounding could readily have accounted for such small risk elevations.[25] In addition, in the one ITT analysis in which confounding was controlled the HR was 1.20, and the lower 95% CI was 0.94.[11] That is, for the most precise HR estimate, the association was not nominally significant, and it could have been due to chance. Moreover, had an 'as treated' analysis with control for

confounding been performed, the association would have been even less statistically robust.

In contrast to the marginal 95% CIs, the p values for the overall HR estimates were significant. However, in causal research it is a truism that if a finding is only 'significant' using one statistical method, but only 'almost significant' using another method, there are insufficient data.

Duration-response

During the first 2 years of follow-up the HRs were below 1.0, at 5 years the estimate was 2.64, and then it declined to 1.12 at ≥6 years.[2] Thus there was no consistent monotonic trend for the HR to increase from 1.0 at baseline with increasing duration of follow-up.

Among women who had not used HRT previously, the HR for those assigned to E+P was not significantly increased, but there was an ostensible duration effect: the HR increased from 0.48 at Year 1 to 1.24 at Year 6 (trend $p = 0.02$).[11] The investigators asserted that the findings for women not previously exposed to HRT "should not be interpreted as overall breast safety [sic] given the statistically significant test for increasing risk with time since randomization". Or put another way, they speculated that, had follow-up lasted longer, an overall increase would have been observed.

That speculation was not tenable for two reasons: first, the significant duration trend was dependent not only on *increased* HRs after 3 or more years of follow-up, but also on *reduced* HRs during the first 2 years of follow-up. Second, among women who had previously used HRT there was no significant duration trend ($p = 0.10$). Why there should have been a significant trend among women who had not previously used HRT, but not among those who had, was not explained (see: Internal consistency).

Internal consistency

As noted above the data were inconsistent according to the prior receipt of HRT. In addition, as also noted above, HRs stratified according to whether the women were or were not 'unblinded' have not been presented.

External consistency

The findings in the WHI[1 10–13] and MWS[3] were inconsistent: in the clinical trial the HR was decreased during the first 2 years of follow-up, and it only increased thereafter; in the MWS the HR was already significantly elevated after 1 month of follow-up, and during the first 2 years it increased progressively with increasing duration of follow-up.

Biological plausibility

E+P may enhance the proliferation of benign cells and thus increase the likelihood of errors in DNA replication, and via mutation to new cancer cells.[26–28] However, it has not been shown that E+P directly

Table 1 Hazard ratio estimates for invasive breast cancer in the Women's Health Initiative (WHI) clinical trial: estrogen plus progestogen versus placebo

Report	HR	95% CI	P	Mean (SD) duration of follow-up (years)
WHI Writing Group[2]	1.26	1.00–1.59	z = –3.19*	5.2 (1.3)
Chlebowski *et al*.[10]	1.24	1.01–1.54	0.003*	5.6 (1.3)
Anderson *et al*.[11]	1.24	1.02–1.50	0.003*	5.6 (1.3)
	1.20†	0.94–1.53†	0.025†	5.6 (1.3)
Heiss *et al*.[12]				
Intervention phase‡	1.26	1.02–1.55	–	5.2 (1.3)
Post-intervention phase§	1.27	0.91–1.78	–	–
Overall¶	1.27	1.06–1.51	–	–
Chlebowski *et al*.[13]	1.25	1.07–1.46		11.0 (2.7)

*Log rank test statistic.
†Adjusted for confounding (see text).
‡1993–2002.
§2002–2005.
¶1993–2005.
CI, confidence interval; HR, hazard ratio; SD, standard deviation.

damages DNA, leading to mutations (initiation). There is evidence, however, that E+P enhances the proliferation of tumour cells, once initiated (promotion). The hypothesis in the WHI studies, therefore, was not that E+P initiates the cellular changes leading to breast cancer, but that it promotes its onset. Under a promotional hypothesis, before a tumour can be mammographically detected it has to comprise about one billion cells, and it has been estimated that this process takes at least 10 years. Since among women who had not used HRT before randomisation the average duration of E+P use was <5 years, a gradient of increasing risk with increasing duration of follow-up, from 0.48 after 1 year to 1.24 after 6 years (trend $p = 0.02$)[11] was not plausible (see also above).

Combined data from the WHI clinical trial and observational study

In three reports the clinical trial data were combined with data from the WHI observational study[14–16] in which 16 121 HRT users and 25 328 non-users who "were ineligible for, or not interested in, the clinical trials"[14] were followed. The principal findings were as described below.

First report[14]

In the clinical trial data combined with a subset of the observational data the HRs for women who commenced E+P use soon after the menopause the HRs were 1.64 (95% CI 1.00–2.68) after 5 years of use and 2.19 (95% CI 1.56–3.08) after 10 years. The authors concluded that "women who initiate use soon after menopause and continue for many years appear to be at particularly high risk".

Second report[15]

In an 'as treated' analysis of the clinical trial data, after an initial decline in breast cancer risk during the first 2 years of follow-up, there was a trend of increasing risk while the trial continued, followed by a trend of decreasing risk after it ended. The difference between the two trends was significant ($p = 0.005$). In the observational data, following publication of the clinical trial findings[2] there was a year-to-year decline in HRT use. In 2002 and 2003, respectively, 122 and 68 cases of breast cancer were diagnosed, "a 43% reduction". The authors stated that the findings were "unrelated to changes in frequency of mammography". They concluded that stopping the use of [E+P] may lead to rapid regression of preclinical cancer, and that doing so may be the predominant factor accounting for the decline in breast cancer incidence.

Third report[16]

In the combined data covering the period 1993–2004, among E+P users the "[HRs] for breast cancer and total cancer were comparatively higher ($p<0.05$) among women who initiated hormone therapy soon after menopause".

Evaluation

The WHI investigators stated that the combined analyses were valid because "the clinical trial and observational study subjects were drawn from the same populations, over the same time period, with much commonality in data collection, protocol, and procedures".[24] That statement was incorrect: one population comprising women who consent to be randomised and 'blinded', and who are deemed eligible, and another population comprising women who decline or are deemed ineligible, and who are not 'blinded', cannot be regarded as 'the same'.

In the observational study the risk of breast cancer according to reasons for non-eligibility or refusal to participate in the clinical trial was not estimated. Thus the validity of observational data cannot be fully assessed. To the extent feasible, below we apply causal criteria to the evidence from the combined analyses.

Time order

In the observational study E+P users aware of as yet undiagnosed breast lumps could selectively have enrolled for follow-up (see: Detection bias).

Detection bias

As strong as was the likelihood of bias in the clinical trial, that likelihood was stronger in the observational study. At recruitment the women were informed that HRT may increase the risk of breast cancer, and users who refused to participate in an experiment, or who were ineligible to participate, but who nevertheless consented to be followed, would have been more anxious than anyone else. Within weeks of publication of the E+P clinical trial findings[2] all participants in the observational study were informed of the findings in writing. Inevitably any pre-existing anxiety among HRT users would have been reinforced.

Chlebowski *et al.* stated that in the observational study the annual frequency of mammography was lower among non-users of HRT than among users (their Table 3).[15] That statement was incorrect: what their table in fact showed was that *lower percentages of the diagnosed breast cancer cases were mammographically detected* each year among non-users of HRT than among users, and that the differences were consistent over time ($p<0.01$). Those findings were quantitative evidence to suggest detection bias. Overall mammography rates were not compared, and it is likely that HRT users more commonly had mammograms than did non-users.

The authors stated that "mammography use [in the clinical trial] was similar in the [HRT] and placebo groups throughout the trial, including the years immediately before and after the intervention ended" (their Table 2).[15] Again that statement was incorrect: what their table in fact showed was that *among the E+P and placebo recipients similar percentages of diagnosed breast cancer cases were detected with WHI study mammograms during each year of follow-up*. Mammography rates among women who discontinued their assigned treatments were not compared, and it is likely that they were higher in women originally assigned to E+P.

Confounding

Prentice et al.[14] stated that "standard breast cancer risk factors were included in the … observational study analyses". No information on what those factors were was provided. Thus the adequacy with which confounding was controlled cannot be evaluated. In the observational data Chlebowski et al.[15] allowed for age, race, BMI, education, smoking, alcohol consumption, health status, physical activity, and family history. However, factors such as age at first birth or parity, age at menarche and menopause, or history of benign breast disease were not controlled.

Statistical stability and strength of association

One reason for conducting the combined analyses was in order to have sufficient data to assess breast cancer risk in subgroups, and it was claimed that among recently menopausal women long-duration HRT use appeared to place women at particularly high risk of breast cancer.[14 16] However, all 95% CIs in the subgroup analyses overlapped, and none of the differences in the HR estimates were statistically significant. In addition, since the likelihood of detection bias was greater in the observational study than in the clinical trial, higher HRs were to be expected in the combined analyses.

To these considerations it must be added that estimation of statistical significance in combined data from two studies, one of which was biased, and the other of which was more biased (see above), was not valid.

Biological plausibility

Chlebowski et al.[15] speculated that their evidence "suggests that withdrawal of [E+P] leads to regression of preclinical cancers". There is no pathological evidence to support that speculation. In addition, in pathological terms it is inconceivable that a 44% drop in the incidence of breast cancer in a single year (not 43% as stated), from 122 to 68 cases, can be attributed to the withdrawal of HRT.

Conclusions

We conclude that, in effect, the WHI clinical trial became an observational study. Time order was correctly specified and there was no information bias, but the trial failed to satisfy the criteria of detection bias, confounding, statistical stability and strength of association, duration-response, internal consistency, external consistency and biological plausibility. Contrary to what was claimed, the WHI was not "the first randomized controlled trial to confirm that [E+P] does [our emphasis] increase the risk of incident breast cancer".

The combined evidence from the clinical trial and the observational study also failed to satisfy the causal criteria. Nor was it possible to fully assess the combined data because much of the requisite information was not provided. In the observational data there was also quantitative evidence of detection bias, and it is likely that that bias was more marked than in the clinical trial.

It remains possible that the use of estrogen plus a progestogen may increase the risk of breast cancer.

Our overall conclusion, however, is that the WHI studies have not demonstrated that it does.

References

1 **Collaborative Group on Hormonal Factors in Breast Cancer**. Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52 705 women with breast cancer and 108 411 women without breast cancer. *Lancet* 1997;**350**:1047–1059.

2 **Writing Group for the Women's Health Initiative Investigators**. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. Principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;**288**:321–333.

3 **Million Women Study Collaborators**. Breast cancer and hormone replacement therapy in the Million Women Study. *Lancet* 2003;**362**:419–427.

4 **The Women's Health Initiative Steering Committee**. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA* 2004;**291**:1701–1712.

5 **Ravdin PM**, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* 2007;**356**:1670–1674.

6 **Shapiro S**, Farmer RDT, Seaman H, et al. Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies. Part 1. The Collaborative Reanalysis. *J Fam Plann Reprod Health Care* 2011;**37**:103–109.

7 **Hill AB**. The environment and disease: association or causation? *Proc Royal Soc Med* 1965;**58**:295–300.

8 **Susser M**. *Causal Thinking in the Health Sciences*. New York, NY: Oxford University Press, 1979.

9 **Department of Health, Education and Welfare**. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service*. Washington, DC: U.S. Government Printing Office, 1964.

10 **Chlebowski RT**, Hendrix SL, Langer RD, et al. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. *JAMA* 2003;**289**:3243–3253.

11 **Anderson GL**, Chlebowski RT, Rossouw JE, et al. Prior hormone therapy and breast cancer risk in the Women's Health Initiative randomized trial of estrogen plus progestin. *Maturitas* 2006;**55**:103–115.

12 **Heiss G**, Wallace R, Anderson GL, et al. Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. *JAMA* 2008;**299**:1036–1045.

13 **Chlebowski RT**, Anderson GL, Gass M, et al. Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. *JAMA* 2010;**304**:1684–1692.

14 **Prentice RL**, Chlebowski RT, Stefanick ML, et al. Estrogen plus progestin therapy and breast cancer in recently postmenopausal women. *Am J Epidemiol* 2008;**167**:1207–1216.

Article

15 **Chlebowski RT**, Kuller LH, Prentice RL, *et al*. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med* 2009;**360**:573–587.

16 **Prentice RL**, Manson JE, Langer RD, *et al*. Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause. *Am J Epidemiol* 2009;**170**:12–23.

17 **Shapiro S**, Farmer RDT, Mueck AO, *et al*. Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies. Part 3: unopposed estrogen. *J Fam Plann Reprod Health Care* 2011;37:(in press).

18 **Hays J**, Hunt JR, Hubbell FA, *et al*. The Women's Health Initiative recruitment methods and results. *Ann Epidemiol* 2003;**13**:S18–S77.

19 **The Writing Group for the PEPI Trial**. Effects of hormone replacement therapy on endometrial histology in postmenopausal women. The Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA* 1996;**275**:370–375.

20 **Gail MH**, Brinton LA, Byar DP, *et al*. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* 1989;**81**:1879–1886.

21 **Welch HG**, Black WC. Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? *Ann Intern Med* 1997;**127**:1023–1028.

22 **Greendale GA**, Reboussin BA, Sie A, *et al*. Effects of estrogen and estrogen-progestin on mammographic parenchymal density. Postmenopausal Estrogen/Progestin Interventions (PEPI) Investigators. *Ann Intern Med* 1999;**130**:262–269.

23 **Rosenberg RD**, Hunt WC, Williamson MR, *et al*. Effects of age, breast density, ethnicity, and estrogen replacement therapy on screening mammographic sensitivity and cancer stage at diagnosis: review of 183,134 screening mammograms in Albuquerque, New Mexico. *Radiology* 1998;**209**:511–518.

24 **Prentice RL**, Chlebowski RT, Rossouw JE, *et al*. Four authors reply [Letter to the Editor]. *Am J Epidemiol* 2008;**168**:1213–1214.

25 **Shapiro S**. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol* 2000;**151**:939–945.

26 **Dietel M**, Lewis MA, Shapiro S. Hormone replacement therapy: pathobiological aspects of hormone-sensitive cancers in women relevant to epidemiological studies on HRT: a mini-review. *Hum Reprod* 2005;**20**:2052–2060.

27 **Mueck AO**, Seeger H, Shapiro S. Risk of breast cancer during hormone replacement therapy: mechanisms. *Horm Biol Clin Invest* 2010;**3**:329–339.

28 **Santen RJ**, Allred DC, Ardoin SP, *et al*. Postmenopausal hormone therapy: an Endocrine Society scientific statement. *J Clin Endocrinol Metab* 2010;**95**:s1–s66.

JFPRHC

Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies: Part 2. The Women's Health Initiative: estrogen plus progestogen

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

Updated information and services can be found at:

http://jfprhc.bmj.com/content/37/3/165

These include:

References	This article cites 25 articles, 7 of which you can access for free at: **http://jfprhc.bmj.com/content/37/3/165**#BIBL
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes