



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

Part 4. The Million Women Study

Samuel Shapiro,¹ Richard D T Farmer,² John C Stevenson,³ Henry G Burger,⁴ Alfred O Mueck⁵

¹Visiting Professor of Epidemiology, Department of Epidemiology, University of Cape Town, Cape Town, South Africa

²Emeritus Professor of Epidemiology, Department of Epidemiology, University of Surrey, Guildford, UK

³Consultant Physician and Reader in Metabolic Medicine, National Heart and Lung Institute, Imperial College, London and Royal Brompton Hospital, London, UK

⁴Emeritus Director, Prince Henry's Institute of Medical Research, Monash Medical Centre, Clayton, and Consultant Endocrinologist, Jean Hailes Medical Centre for Women, Clayton, Victoria, Australia
⁵Professor of Clinical Pharmacology and Experimental Endocrinology, Department of Endocrinology, University Women's Hospital, Tübingen, Germany

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

Received 28 September 2011

Accepted 2 November 2011

Published Online First

16 January 2012

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 hormone replacement therapy (HRT) with estrogen plus progestogen (E+P) is now an established cause of breast cancer; the CR and MWS investigators claim that unopposed estrogen therapy (ET) also increases the risk, but to a lesser degree than does E+P. The authors have previously reviewed the findings in the CR and WHI (Parts 1–3).

Objective To evaluate the evidence for causality in the MWS.

Methods Using generally accepted causal criteria, in this article (Part 4) the authors evaluate the findings in the MWS for E+P and for ET.

Results Despite the massive size of the MWS the findings for E+P and for ET did not adequately satisfy the criteria of time order, information bias, detection bias, confounding, statistical stability and strength of association, duration-response, internal consistency, external consistency or biological plausibility. Had detection bias resulted in the identification in women aged 50–55 years of 0.3 additional cases of breast cancer in ET users per 1000 per year, or 1.2 in E+P users, it would have nullified the apparent risks reported.

Conclusion HRT may or may not increase the risk of breast cancer, but the MWS did not establish that it does.

Background

In Parts 1–3 of this series of articles we have applied generally accepted epidemiological principles of causality^{1–4} to studies of the risk of breast cancer in users of hormone replacement therapy (HRT), as reported from the collaborative reanalysis (CR)⁵ (Part 1⁶), and the Women's Health Initiative (WHI)^{7–18} (Parts 2¹⁹ and 3²⁰). In

Part 1 we concluded that the CR findings for HRT [both unopposed estrogen therapy (ET) and estrogen plus progestogen (E+P)] did not establish causality. In Part 2 we concluded that the WHI findings for E+P did not establish causality. By contrast, in Part 3 we concluded that valid WHI findings suggested that ET does not increase the risk of breast cancer, and may even decrease it; the latter possibility, however, was statistically borderline.

In 2003, a year after the WHI findings for E+P were published,⁷ the Million Women Study (MWS) investigators reported an increased risk of breast cancer in HRT users,²¹ and based on the combined evidence from the CR, the WHI and the MWS it is now widely believed that E+P is an established and major cause of the disease. The MWS investigators (but not the WHI investigators)^{15–18} claim that ET also increases the risk, although to a lesser degree than does E+P.^{21–24}

Here, in Part 4 we apply causal principles to the evidence from the MWS.^{21–24} In the MWS the estimated levels of risk associated with the use of HRT were greater than in the CR or the WHI, and in view of the impact the study had on regulatory authorities, and on the public perception of safety, it is especially important to evaluate its validity.

The Million Women Study^{21–24}

In the UK all women aged 50–64 years are invited to undergo screening mammography at 3-year intervals.²¹ From May 1966 to December 2001 the MWS investigators sent letters and questionnaires²⁵ to women invited to attend. Follow-up questionnaires²⁶ were sent 2–3 years after recruitment. The women were followed

for breast cancer incidence and mortality in National Health Service Central Registries.

Below, except where otherwise stated, all 95% confidence intervals (CIs) around the relative risk (RR) estimates excluded 1.0, and for convenience they are omitted.

First report²¹ (2003)

Among 828 923 postmenopausal women followed for an average of 2.6 years the RRs of invasive breast cancer for current and past users of HRT were 1.66 and 1.01 (95% CI, 0.94–1.09). Among women currently using HRT at baseline the RRs for users of various types of HRT were as follows: ET, 1.30; E+P, 2.00; tibolone, 1.45; other or unknown HRT, 1.44. The difference between E+P vs ET was significant ($p < 0.0001$).

For current ET use at baseline the RRs for <5 and ≥ 5 years' total duration were 1.21 and 1.34, and for E+P use, 1.70 and 2.21. For ET use the RRs for total durations of <1 , 1–4, 5–9 and ≥ 10 years of use were 0.81 (95% CI, 0.55–1.20), 1.25, 1.32 and 1.37; for E+P use they were 1.45, 1.74, 2.17 and 2.31.

Among women who last used HRT ≤ 1 year previously the RR was 1.14; for exposures that ended 2– ≥ 10 years previously the RRs approximated unity. The average time to diagnosis was 1.2 years, and within 1.7 years of diagnosis the RR of fatal breast cancer was 1.22.

The investigators estimated that the “use of HRT by UK women aged 50–64 years ... resulted in an extra 20 000 incident breast cancers, combined [E+P] accounting for 15 000” of them. They also estimated that HRT would “result in five to six extra cancers per 1000 women with 5 years' use and 15–19 ... per 1000 with 10 years' use”. They concluded that “current use of HRT is associated with an increased risk of incident and fatal breast cancer” ... [which is] ... “substantially greater for [E+P] combinations than for other types of HRT”.

Second report²² (2004)

Among users of HRT at baseline the RRs at 0.1 ('screen-detected'), 0.7, 1.5, 2.5 and 3.4 years of follow-up were 1.37, 2.66, 2.16, 1.66 and 1.70, respectively. The average durations of use ranged from 6.1 to 6.9 years. The RRs were higher for E+P than for ET users, and maximal at 0.7 years (ET, 1.72; E+P, 3.31).

For women aged 50–55 years who used HRT for 5 years the estimated absolute risks attributable to ET and E+P use were 1.5 and 6.0 per 1000.

Third report²³ (2006)

Among 1 031 224 postmenopausal women followed over 3.6 million woman-years (WY) for the incidence of invasive and *in situ* breast cancer “the mean time ... from ... last contact to the end of follow-up was 2.7 years [SD (standard deviation) 1.1]”. “At the time of the analysis follow-up information was available for

the first two-thirds of the study population”, and there were “392 341 (38%) women for whom follow-up information [was] included in [the] analysis”.

Among current users of HRT the respective RRs of *in situ* and invasive breast cancer were 1.55 and 1.74. The RRs were higher for invasive mixed ductal-lobular or tubular tumours (2.13 and 2.66) than for ductal tumours (1.63); the RRs were also higher among E+P than among ET users, but for each type of cancer the RRs did not increase significantly with increasing duration of use. For ductal and lobular tumours the RRs declined with increasing body mass index (BMI) (trend $p < 0.0001$).

The investigators concluded that “the risks of invasive lobular and tubular cancers associated with current use of both [ET and E+P] are higher than for invasive ductal cancer” and higher for E+P users than for ET users.

Fourth report²⁴ (2011)

Among 1 129 025 postmenopausal women followed until “the end of 2002 ... two thirds of the participants had been mailed the second questionnaire and the response was 65%”. During 4.05 million WY of follow-up 15 759 invasive and *in situ* breast cancers were diagnosed.

The RRs for current users of HRT, ET, E+P, tibolone and other and unknown HRT were 1.68, 1.38, 1.96, 1.38 and 1.55, respectively, and the estimates were statistically heterogeneous ($p < 0.001$). In the first 2 years after HRT ceased the RR was 1.16, after which the RRs approximated unity. For durations of use of <5 and ≥ 5 years the respective RRs among ET users were 1.24 and 1.44; among E+P users they were 1.62 and 2.19.

For both ET and E+P users the RRs were lower for breast cancers diagnosed in the first 4 months after recruitment than subsequently [ET, 1.19 and 1.50 ($p < 0.001$); E+P, 1.41 and 2.32 ($p < 0.001$)]. For ET users the RRs of ‘screen-detected’ and ‘non-screen-detected’ cancers were 1.16 and 1.59 ($p < 0.001$); for E+P the corresponding estimates were 1.64 and 2.81 ($p < 0.001$). Those comparisons “should [have included] virtually all breast cancers found at screening soon after the baseline questionnaire was completed”.

For current ET users whose use began <5 and ≥ 5 years after the menopause, the RRs were 1.43 and 1.05 ($p < 0.001$); for E+P users the estimates were 2.04 and 1.53 ($p < 0.001$). “The proportionate increase in risks of breast cancer associated with use of hormone therapy was greater among lean women than among obese women”, but within BMI strata (≥ 25 kg/m² and < 25 kg/m²) the HRT-associated RRs remained higher for those whose use commenced <5 years after the menopause.

For both ET and E+P users the RRs declined with increasing tumour grade (Grades I–III): ET, 1.27, 1.16, 0.87 ($p < 0.001$); E+P, 2.42, 1.67, 1.03 ($p < 0.001$).

For estrogen receptor (ER)-positive vs ER-negative status the RRs for ET users were 1.76 and 1.29 ($p=0.005$); for E+P users the estimates were 3.10 and 1.37 ($p<0.001$). For node-positive vs node-negative tumours among ET users the RRs were 1.19 and 1.09 ($p=0.3$); among E+P users they were 2.00 and 1.66 ($p=0.009$).

The investigators concluded that “risks were substantially greater among users of [E+P] than estrogen only formulations and if hormonal therapy started at or around the time of menopause than later”.

Evaluation of the MWS

Below we evaluate whether the evidence in the MWS accorded with generally accepted principles of causality.¹⁻⁴ The principles are inter-related, and when appropriate we cross-refer.

Time order

If allowance is made for the time from the diagnosis of breast cancer to its recording in a registry, virtually all the cases identified at 0.1 years of follow-up (HRT: RR, 1.37)²² or at 4 months (ET: RR, 1.19; E+P: RR, 1.41),²⁴ were already present when the women were recruited (see: Detection bias) and time order was violated. In a properly designed cohort study breast cancers already present at baseline should have been excluded.

Time order was further violated in respect of the timing and duration of HRT use. In the third report²³ follow-up information on HRT use was unavailable for 62% of the women. In the fourth report,²⁴ by December 2002 the follow-up questionnaire had been received by about 66% of the women, among whom the response rate was 65%. Hence follow-up information on HRT use [and on menopausal status (see: Detection bias) and on confounders (see: Confounding)] was missing for about 57% [$1 - (\pm 0.66 \times 0.65) \times 100$] of the women. Following publication of the WHI findings⁷ there was a rapid and marked decline in the use of HRT.²⁷ For that reason, as well as for other reasons (e.g. HRT-induced breakthrough bleeding),⁷ since 66% of ever-users of HRT at baseline were current users [our calculation: derived from Figure 1 (current use) and Figure 2 (past use) in Reference 21], a substantial proportion could have become past users by the end of 2002.

How unreliable were the data? Recruitment commenced in 1996 and follow-up ended in December 2002.²⁴ For about 50% of the women the time from last contact to diagnosis was >1.2 years,²¹ and to the end of follow-up >2.7 years.²³ For women enrolled in 1996 that interval could have been as much as 6 years. Thus it is likely that much of what was defined in the analysis as current HRT use became past use during follow-up. In addition, the duration data were incorrect (see: Duration-response), as were the data on menopausal status and confounding (see: Detection bias and confounding).

Information bias

Information bias in a cohort study is unusual, but it can occur, and in the MWS it was likely. At recruitment HRT users already aware of as yet undiagnosed breast lumps, or of suspect mammographic changes identified before recruitment (see: Detection bias), could have tended to overestimate the total duration of use. Had women who already had breast cancer at baseline been excluded, that bias could largely have been avoided (see: Time order).

A defect in the study design may also have facilitated the occurrence of information bias. Ethinylestradiol (EE), listed as one of 34 memory-prompts in the questionnaire²⁵ as an HRT preparation, is a synthetic estrogen present exclusively in oral contraceptives. Women who were aware of breast lumps at recruitment, or who had suspect mammographic changes (see: Time order and detection bias), could erroneously have identified EE as HRT. Soon after publication of the MWS report²¹ the authors stated in an erratum that what was meant by ‘ethinylestradiol’ was ‘estradiol’.²⁸ Yet the error was not corrected in the second questionnaire,²⁶ administered 2–3 years after the first questionnaire.²⁵

Detection bias

The design of a study of the risk of breast cancer in relation to the use of HRT in which the women were recruited from a screening programme guaranteed that it would be biased. By definition, women who decided to have mammograms were alerted to the possibility of breast cancer, as has also been acknowledged in an earlier study based on mammographic screening,²⁹ and concern that HRT may cause the disease has been widespread, and has increased over time. The MWS invitation was explicit in the first questionnaire:²⁵ “We have a unique opportunity ... to learn about the way different types of HRT ... [affect] a woman’s health, particularly her breasts”. That wording ensured that HRT users already aware of breast lumps, or of suspected breast cancer, would selectively participate (see: Time order).

There was quantitative evidence of detection bias. First, HRT users were selectively enrolled: 32% of the women who participated and 19% or those who did not were HRT users.³⁰ Second, the data suggested that women already aware of breast lumps, or of suspected breast cancer, tended selectively to participate (see: Time order): whereas the incidence of breast cancer in the MWS population was 2.8 per 1000 WY,³¹ in the population at large it was 2.0 per 1000 WY.²¹ Third, the baseline RRs of 1.37²² or 1.41²⁴ (‘screen-detected’ breast cancer) indicated that women who both used HRT and who were also aware of breast lumps, or of suspect lesions, or of suggestive precancerous changes identified in earlier mammograms, were the most likely to participate. Fourth, the average time from recruitment to breast cancer diagnosis was 1.2 years,²¹ and 1.7 years thereafter the RR of fatal breast cancer was

1.22. An increased risk of fatal cancer among HRT users within 2.9 (1.2+1.7) years of recruitment was not plausible (see: Biological plausibility), and it could have been due to the selective enrolment of HRT users with pre-existing suspected or diagnosed breast cancer. Fifth, the RRs declined with increasing BMI,²⁴ a known risk factor for breast cancer in postmenopausal women (see: Biological plausibility), and the larger the breasts, the less likely was it that otherwise occult breast cancer would selectively have been detected among HRT users by mammography.

Detection bias could also have occurred during follow-up, as previously described in our critique of the CR.⁶ Briefly, HRT users are advised to have regular breast examinations and mammograms, and in the MWS users more frequently underwent mammography than did non-users;³⁰ when mammograms are performed HRT use is routinely recorded, and about 30% of breast cancers actually present go undetected;³² about 5% of postmenopausal women have 'clinically silent' breast cancer;³³ and HRT diminishes the sensitivity of mammography.³² The mammograms of HRT users could have been more intensively scrutinised than those of non-users, especially if they were radiologically dense, and otherwise occult breast cancer could selectively have been detected among the users.

For both ET users and E+P users the RRs were lower during the first 4 months of follow-up than subsequently.²⁴ The investigators stated that "it has been suggested that part of the increased hormone therapy-associated risk ... observed in this study may have resulted from the selective recruitment of hormone therapy users who already had symptoms of breast cancer. If that had happened there would have been a greater hormone therapy-associated excess of breast cancer soon after recruitment than subsequently. However, the opposite was found". They argued that these findings "largely [reflected] the lower hormone therapy-associated risks observed for screen-detected breast cancers than for non-screen-detected breast cancers". That claim ignored the likelihood that during follow-up HRT users could more commonly have had repeat mammograms than non-users (see: Duration-response), and because of a further defect in the study design that possibility could not be assessed: information on repeat mammograms was not solicited in the second questionnaire²⁶ (see: Confounding).

There was further evidence to suggest detection bias. The RRs were consistently lower for ET users than for E+P users.^{21–24} Unopposed ET causes uterine cancer, and ET is preferentially prescribed to hysterectomised women, among whom vaginal bleeding does not occur. By contrast, E+P is preferentially prescribed to women with a uterus, among whom breakthrough bleeding is common;⁷ and bleeding makes it mandatory to rule out endometrial cancer. HRT users alerted to the risk of that cancer would have become worried about breast cancer as well, and have sought to rule it

out. Hence, it was to be expected that detection bias would be greater for E+P users than for ET users.

The RRs for invasive lobular and tubular tumours were higher than for ductal tumours.²³ Lobular and tubular tumours are more highly differentiated, smaller, and more slow-growing than ductal tumours,^{34,35} and the detection of lobular tumours by mammography is also more difficult.³⁶ More intensive scrutiny of mammograms of HRT users than of non-users could have resulted in the selective detection of lobular tumours, especially in radiologically dense mammograms, that might otherwise have gone undetected.

The RR for *in situ* breast cancer was 1.55.²³ *In situ* tumours are seldom clinically detectable, usually they are identified by mammography, and the investigators acknowledged that detection bias was likely (see: Detection bias). Yet in the fourth report²⁴ *in situ* and invasive breast cancers were considered together. In that report the RRs were higher if HRT had commenced within 5 years of the menopause than subsequently. However, since the data for *in situ* breast cancer were biased, the combination of *in situ* and invasive breast cancer was also biased. In addition, most of the women who were premenopausal at recruitment would have reached the menopause during follow-up, among the 57% of women not followed that information was missing, and there was substantial misclassification of menopausal status, and of the time since menopause (see: Time order).

The RRs declined with increasing tumour grade, and were higher for ER-positive than for ER-negative tumours, and for node-positive than for node-negative tumours.²⁴ As shown in Table 1 (our calculations: derived from Figures 1 and 3 in Reference 24) unknown values for tumour grade, ER status and nodal status among current users of ET and E+P, and among never-users of HRT, ranged from 49.5% to 74.1%. Such high rates cast doubt on the validity of the evidence. In addition, the declining RRs with increasing tumour grade could have been biased if more common use of mammography by HRT users than by non-users resulted in the selective detection of low-grade tumours; an association with ER-positivity could have occurred if breast cancers in HRT users were more commonly tested, and if ER-positive, more commonly documented in the registries; and the higher RRs for node-positive than for node-negative tumours could readily have been due to detection bias.

How much bias would it have taken to account for the findings? In the first report²¹ the investigators estimated that among women aged 50–64 years the use of HRT would result in "five to six extra cancers per 1000 women with 5 years' use and 15–19 ... per 1000 with 10 years' use". Thus if detection bias resulted in the identification of 1–1.2 (5–6/5) otherwise occult cases each year among 1000 women exposed for 5 years, or 1.5–1.9 (15–19/10) cases each year among women exposed for 10 years, that bias would have nullified

Table 1 Breast cancer in the Million Women Study: percentages of unknown values for tumour grade, estrogen receptor status and nodal status among current users of estrogen therapy, current users of estrogen plus progestogen and never-users of hormone replacement therapy

	ET (n=2131)*	E+P (n=4292)*	No HRT (n=6127)*
Known tumour grade (n) [†]	978	1893	3096
Percentage unknown (%)	54.1	55.9	49.5
Known ER status (n) [†]	753	1684	1585
Percentage unknown (%)	64.7	60.8	74.1
Known nodal status (n) [†]	898	1781	2848
Percentage unknown (%)	57.9	58.5	53.5

*Derived from Figure 1 in Reference 24.

[†]Derived from Figure 3 in Reference 24.

E+P, estrogen plus progestogen; ER, estrogen receptor; ET, estrogen therapy; HRT, hormone replacement therapy.

the findings. In the second report,²² among women aged 50–55 years the respective absolute risks for ET or E+P use for 5 years were estimated to be 1.5 and 6.0 per 1000. That is, if detection bias resulted in the identification of 0.3 (1.5/5) additional cases in ET users each year, or 1.2 (6.0/5) additional cases in E+P users, that bias would have nullified the findings. Absolute risks ranging from 0.3 to 1.9 per 1000 women per year could plausibly have been due to detection bias.

Confounding

Confounding was incompletely controlled. In the first,²¹ second²² and third²³ reports the factors allowed for included age, time since menopause, parity, age at first birth, family history, BMI, region and socio-economic status. In the fourth report²⁴ age at menopause and alcohol consumption were also allowed for. During follow-up factors such as menopausal status, time since menopause, age at menopause and BMI changed, and for about 57–62% of the women the information was missing (see: Time order). In addition, information on the receipt of a mammogram during follow-up was not solicited in the second questionnaire²⁶ (see: Detection bias).

Statistical stability and strength of association

In our critique⁶ of the CR⁵ we alluded to the relationship between the statistical stability and strength of any given association: if a RR is ‘large’ (say >5.0), a 95% CI that excludes 1.0 (i.e. a ‘statistically significant’ association) can be documented in a relatively small study. But if a RR is ‘small’ (say <2.0), usually it can only be documented in a massive study. The difficulty however, is that “if a massive study is sufficiently massive, any deviation of the RR from 1.0, no matter how small, becomes ‘significant’”; but it may be impossible to discriminate among bias, confounding and causation as alternative explanations. By contrast, “in a well-conducted study, when a RR is large,

it may be reasonable to judge that it might perhaps be reduced, but not be obliterated, even if it were possible to entirely eliminate all sources of bias and confounding. But if an association is small it may be impossible to judge. In the latter circumstance ‘statistical significance’ may not equate with causality: given a massive amount of data, all that may be accomplished is to rule out chance as one possible explanation, but not bias or confounding”.

In the four reports the highest overall RR for HRT users was 1.74,²³ and RRs in excess of 2.0 were identified only in subgroups. For ET users the overall RR was 1.30, and <2.00 in all subgroups. Such small RRs could have been due to bias or confounding. For E+P users the overall RR of 2.00 was again small,²¹ but significantly higher than the estimate of 1.30 for ET ($p < 0.0001$). Or put another way, the RR for E+P versus ET was 1.54 (2.00/1.30). Such a small association could readily have been biased or confounded (see: Detection bias and confounding), illustrating how in a massive study, virtually any deviation of RR from 1.0, no matter how small, can yield a p value of <0.0001.

Dose/duration-response

Under a promotional hypothesis it might reasonably be expected that the use of HRT would confer a greater risk of breast cancer, the higher the dose or the longer the duration of use (see: Biological plausibility).

Dose-response

Dose-response was not analysed.

Duration-response

In the first report,²¹ for women who were using HRT at baseline (defined in the MWS as current users) the total duration of use of all episodes of use, current plus past, was analysed. That analysis was incorrect. Since the RR approximated unity within 2 years of stopping,²⁴ the duration of past use was irrelevant, and only the *duration of the current episode of use* should have been analysed. In addition, the analysis of duration of use, as represented at baseline, misrepresented the actual duration of use, since follow-up information was missing for 62% of the women (see: Time order).

A further defect in the study design made it impossible to analyse the duration of current HRT use among women who used more than one product. In the baseline questionnaire²⁵ five relevant questions were asked: “32. Have you ever used [HRT]?”; “35. For about how many years *in total* have you used HRT?”; “36. Are you now using HRT?”; “37. What is the name of the most recent HRT you have used?”; and “38. For how many years did you use the most *recent* type of HRT?”.

Based on questions 32 and 35, among *current* HRT users at baseline who used more than one product the *total* duration of *ever-use* could be analysed, but based on questions 36, 37 and 38 the duration of *current* HRT use could not be. To illustrate, consider a *current* HRT user who at baseline had used E+P for 9 years,

and following hysterectomy, ET for 1 year: the *current* 1-year duration of ET use would have been recorded, but the *current* 10-year duration of HRT use (E+P, 9 years + ET, 1 year) would not have been.

In the second report²² duration data were not given. In the third report²³ “there was no significant difference in the trends in [RR] with duration of use of either type of hormone therapy [ET or E+P] for ductal, tubular or lobular cancer”. In the fourth report²⁴ the RRs for ≥ 5 years of use of ET and of E+P at baseline were higher than for < 5 years of use, and higher for E+P than ET users. Those differences could have been due to detection bias; trends according to durations of < 1 , 1–4, 5–9 and ≥ 10 years were not presented; ‘total duration’ of use again referred to all episodes of use, not to current use; and the duration data were again misclassified, because of follow-up information on HRT use was missing for 57% of the women.

Finally, the RRs for increasing duration of follow-up were inconsistent with a duration-response effect (see: Internal consistency). Among women who used HRT, ET or E+P at baseline the RRs were highest at 0.7 years of follow-up, after which they declined.²² Yet under causal assumptions, the longer the duration of follow-up, the higher should the RRs have been. A plausible explanation of these inconsistent findings is that violation of time order and detection bias could have been greatest during the first year of follow-up.

Internal consistency

As described above, the RRs according to duration of follow-up were inconsistent (see: Duration-response).

External consistency

For ET users the MWS findings were inconsistent with those of the WHI clinical trial^{15 16} in which the evidence suggested that unopposed ET does not increase the risk of breast cancer. In the MWS there was quantitative evidence of bias, whereas in the WHI trial women were randomly assigned, ‘double-blind’, to ET or placebo, all participants were hysterectomised, vaginal bleeding did not occur, ‘unblinding’ was seldom necessary, the ‘unblinding’ rate was $< 2.0\%$, and there was little or no bias.²⁰

For E+P users the MWS findings were inconsistent with those of the CR. In the MWS the RRs approximated unity within 2 years of stopping HRT;²⁴ in the CR the RR only declined to unity 5 years after stopping.⁵

Biological plausibility

Elsewhere we have considered relevant pathological and experimental evidence for and against the possibility that HRT may cause breast cancer.^{6 19 20} Briefly, the hypothesis is not that HRT causes genetic mutation (initiation), but that estrogens, and probably progestogens as well, accelerate the proliferation of otherwise slowly growing malignant cells (promotion). Possible mechanisms are the proliferative

effects of estrogens on estrogen-sensitive cells,³⁷ or the excessive metabolism of estrogens to highly active compounds³⁸ with strong proliferative as well as possibly genotoxic effects. However, estrogens also have antiproliferative and pro-apoptotic effects,³⁹ which could possibly reduce the risk of breast cancer. In addition, estrogens can be metabolised not only to potentially genotoxic metabolites, but also to carcino-protective metabolites, such as 2-methoxy-estradiol.⁴⁰

In short, some mechanisms could possibly increase the risk of breast cancer in HRT users, and other mechanisms could decrease it. However, under a promotional hypothesis, for the most aggressively multiplying cells it is generally accepted that on average it takes at least 10 years to attain a tumour diameter of about 1 cm, which is about the smallest lesion that can be diagnosed clinically.³⁸ In the MWS the average *total duration of HRT use* at baseline was 6.1–6.9 years,²² and the *duration of current use* would have been appreciably less. Since the RR approximated unity within 2 years of discontinuing HRT use, among *current* users of HRT the *duration of past use* cannot have had any effect. It is implausible that the *current use* of HRT at baseline for less than 6 years could have increased the risk of breast cancer. It is also implausible that cancer cells, *once* already promoted, and *once already* invasive, could have ‘unpromoted’ within 2 years of stopping HRT.²⁴

Obesity is a risk factor for breast cancer in postmenopausal women, perhaps because of increased endogenous estrogen secretion,³⁷ and the RRs declined with increasing BMI.²⁴ Under a causal hypothesis, however, although obesity itself increases the risk of breast cancer, the RRs among HRT users should have been higher than among non-users within strata of BMI, and the decline in the RR was explicable by the diminished sensitivity of mammographic screening with increasing BMI (see: Detection bias).

Conclusions

The name ‘Million Women Study’ implies an authority beyond criticism or refutation. Many commentators, and the investigators, have repeatedly stressed that it was the largest study of HRT and breast cancer ever conducted. Yet the validity of any study is dependent on the quality of its design, execution, analysis and interpretation. Size alone does not guarantee that the findings are reliable. The MWS was an observational study, and it had the attendant problems and uncertainties intrinsic to such studies. If the evidence was unreliable, the only effect of its massive size would have been to confer spurious statistical authority to doubtful findings.

Here we conclude that the evidence in the MWS was indeed unreliable. There were defects in the study design, and the findings did not adequately satisfy the principles of causation. In terms of time order, information bias, detection bias, confounding, statistical

stability and strength of association, dose/duration-response, internal consistency, external consistency and biological plausibility the study was defective.

HRT may or may not increase the risk of breast cancer, but the MWS did not establish that it does.

Acknowledgement The authors thank Helen Seaman for providing editorial assistance.

Competing interests Samuel Shapiro, John Stevenson, Henry Burger, and Alfred Mueck 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

- Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295–300.
- Susser M. *Causal Thinking in the Health Sciences*. New York, NY: Oxford University Press, 1979.
- US Dept of Health, Education and Welfare: Public Health Service. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the Public Health Service* (Public Health Service Publication No. 11030). Washington, DC: US Government Printing Office, 1964.
- Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991;133:635–648.
- 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.
- Shapiro S, Farmer RD, 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- The Women's Health Initiative Steering Committee. Effect of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA* 2004;291:1701–1712.
- Stefanick ML, Anderson GL, Margolis KL, *et al.* Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA* 2006;295:1647–1657.
- LaCroix AZ, Chlebowski RT, Manson JE, *et al.* Health outcomes after stopping conjugated equine estrogens among postmenopausal women with prior hysterectomy: a randomized controlled trial. *JAMA* 2011;305:1305–1314.
- Prentice RL, Chlebowski RT, Stefanick ML, *et al.* Conjugated equine estrogens and breast cancer risk in the Women's Health Initiative clinical trial and observational study. *Am J Epidemiol* 2008;167:1407–1415.
- Shapiro S, Farmer RDT, Mueck AO, *et al.* 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. *J Fam Plann Reprod Health Care* 2011;37:165–172.
- Shapiro S, Farmer RDT, Seaman H, *et al.* Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies. Part 3. The Women's Health Initiative: unopposed estrogen. *J Fam Plann Reprod Health Care* 2011;37:225–230.
- Million Women Study Collaborators. Breast cancer and hormone replacement therapy in the Million Women Study. *Lancet* 2003;362:419–427.
- Beral V, Banks E, Reeves G, *et al.* The effect of hormone therapy on breast and other cancers. In: Critchley H, Gebbie A, Beral V (eds). *Menopause and Hormone Replacement*. London: RCOG Press, 2004;136–150.
- Reeves GK, Beral V, Green J, *et al.* Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *Lancet Oncol* 2006;7:910–918.
- Beral V, Reeves G, Bull D, *et al.* Breast cancer risk in relation to the interval between menopause and starting hormone therapy. *J Natl Cancer Inst* 2011;103:1–10.
- The Million Women Study: A National Survey of Women Invited for Breast Screening. First questionnaire. <http://www.millionwomenstudy.org/files/mws-web1.pdf> [accessed 14 September 2011].
- The Million Women Study: Confidential National Study of Women's Health. Second questionnaire. <http://www.millionwomenstudy.org/files/mws-web2.pdf> [accessed 14 September 2011].
- Gompel A, Plu-Bureau G. Is the decrease in breast cancer incidence related to a decrease in postmenopausal hormone therapy? *Ann N Y Acad Sci* 2010;1205:268–276.
- Million Women Study Collaborators. Errata. *Lancet* 2003;362:1160.
- Morrison AS, Brisson J, Khalid N. Breast cancer incidence and mortality in the breast cancer detection demonstration project. *J Natl Cancer Inst* 1988;80:1540–1547.
- Banks E, Beral V, Cameron R, *et al.* Comparison of various characteristics of women who do and do not attend for breast cancer screening. *Breast Cancer Res* 2002;4:R1.
- Shapiro S. The Million Women Study: potential biases do not allow uncritical acceptance of the data. *Climacteric* 2004;7:3–7.

- 32 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.
- 33 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.
- 34 Li CI, Moe RE, Daling JR. Risk of mortality by histologic type of breast cancer among women aged 50 to 79 years. *Arch Intern Med* 2003;163:2149–2153.
- 35 Porter PL, El-Bastawissi AY, Mandelson MT, *et al.* Breast tumor characteristics as predictors of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 1999;91:2020–2028.
- 36 Krecke KN, Gisvold JJ. Invasive lobular carcinoma of the breast: mammographic findings and extent of disease at diagnosis in 184 patients. *AJR Am J Roentgenol* 1993;161:957–960.
- 37 International Agency for Research on Cancer (IARC). *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans (Volume 72): Hormonal Contraception and Postmenopausal Hormone Therapy*. Lyon, France: IARC, 1999;474–530.
- 38 Seeger H, Wallwiener D, Kraemer E, *et al.* Comparison of possible carcinogenic estradiol metabolites: effects on proliferation, apoptosis and metastasis of human breast cancer cells. *Maturitas* 2006;54:72–77.
- 39 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.
- 40 Seeger MH. 2-Methoxyestradiol – biology and mechanism of action. *Steroids* 2010;75:625–631.