

Risks and benefits of hormone therapy: has medical dogma now been overturned?

S. Shapiro, T. J. de Villiers*, A. Pines[†], D. W. Sturdee[‡], R. J. Baber^{**}, N. Panay^{††}, J. C. Stevenson^{‡‡}, A. O Mueck^{***} and H. G. Burger^{†††}

Department of Public Health and Family Medicine, University of Cape Town Medical School, Cape Town, South Africa; *Department of Obstetrics and Gynecology, Stellenbosch University, Cape Town, South Africa; [†]Department of Medicine, Tel Aviv University, Tel Aviv, Israel; [‡]Department of Obstetrics and Gynecology, Solihull Hospital, Solihull, UK; ^{**}Department of Women's and Children's Health, The University of Sydney, Australia; ^{††}Department of Obstetrics and Gynecology, Queen Charlotte's and Chelsea and Westminster Hospitals, London, UK; ^{‡‡}National Heart and Lung Institute, Imperial College, Royal Brompton Hospital, London, UK; ^{***}Women's Hospital, University of Tübingen, Germany; ^{†††}MIMR-PHI Medical Research Institute, Monash University, Melbourne, Australia

Key words: HORMONE THERAPY, RISK, BENEFIT

ABSTRACT

Background In an integrated overview of the benefits and risks of menopausal hormone therapy (HT), the Women's Health Initiative (WHI) investigators have claimed that their 'findings ... do not support use of this therapy for chronic disease prevention'. In an accompanying editorial, it was claimed that 'the WHI overturned medical dogma regarding menopausal [HT]'.
Objectives To evaluate those claims.
Methods Epidemiological criteria of causation were applied to the evidence.
Results A 'global index' purporting to summarize the overall benefit versus the risk of HT was not valid, and it was biased. For coronary heart disease, an increased risk in users of estrogen plus progestogen (E + P), previously reported by the WHI, was not confirmed. The WHI study did not establish that E + P increases the risk of breast cancer; the findings suggest that unopposed estrogen therapy (ET) does not increase the risk, and may even reduce it. The findings for stroke and pulmonary embolism were compatible with an increased risk, and among E + P users there were credible reductions in the risk of colorectal and endometrial cancer. For E + P and ET users, there were credible reductions in the risk of hip fracture. Under 'worst case' and 'best case' assumptions, the changes in the incidence of the outcomes attributable to HT were minor.
Conclusions Over-interpretation and misrepresentation of the WHI findings have damaged the health and well-being of menopausal women by convincing them and their health professionals that the risks of HT outweigh the benefits.

INTRODUCTION

The Women's Health Initiative (WHI) investigators¹ have published an integrated overview of data regarding menopausal hormone therapy (HT) derived from two randomized, 'double-blind', controlled trials that commenced in 1993 (intervention phase), and from follow-up data collected after the trials were terminated (post-intervention phase). In one trial, women with an intact uterus were assigned to conjugated

equine estrogen (0.625 mg/day) plus medroxyprogesterone acetate (2.5 mg/day) ('estrogen plus progestogen', E + P) or placebo, and the trial was terminated after a median of 5.6 years of follow-up; in the second trial, hysterectomized women were assigned to conjugated equine estrogen (0.625 mg/day) without an added progestogen ('estrogen therapy', ET) or placebo, and the trial was terminated after 7.2 years of follow-up. In both trials, cumulative follow-up in the combined intervention and post-intervention phases lasted 13 years.

Correspondence: Professor S. Shapiro, Department of Public Health and Family Medicine, University of Cape Town Medical School, Observatory, Cape Town, South Africa; E-mail: samshap@mweb.co.za

The investigators concluded that ‘findings ... in the two WHI [HT] trials do not support use of this therapy for chronic disease prevention, although it is appropriate for symptom management in some women’. In an accompanying editorial², the findings were described as ‘a victory for women and their health’, and it was claimed that ‘the WHI overturned medical dogma regarding menopausal [HT]’. In this commentary, we evaluate those claims.

SUMMARY OF THE WHI FINDINGS

The primary endpoints specified in the WHI study were coronary heart disease (CHD) and invasive breast cancer. Hazard ratios (HRs) for those outcomes, together with other selected endpoints (stroke, pulmonary embolism, colorectal cancer, endometrial cancer, hip fracture, and all-cause mortality) were combined to yield what was designated as a ‘global index’, ‘summarizing the balance of risks and benefits’³. A large number of secondary endpoints, selected secondary endpoints, and self-reported endpoints were also studied. Here we focus on the results in the intervention phase and in the total cumulative follow-up data for the primary endpoints of CHD and breast cancer, and on the other components of the global index.

HRs for the primary endpoints of CHD and breast cancer, and for the global index are given in Table 1. For E+P users, the 95% confidence intervals (CIs) around the HRs excluded 1.00 for breast cancer (intervention phase, 1.24; cumulative follow-up, 1.28) and the global index (intervention

phase, 1.12). For ET users, the HR for breast cancer was 0.79 (cumulative follow-up).

HRs for constituents of the global index (other than the primary endpoints) for which the 95% CIs around the HRs excluded 1.00 are given in Table 2. For E+P users, in the intervention phase the HRs were significantly elevated for stroke (HR 1.37) and pulmonary embolism (1.98), and reduced for colorectal cancer (0.62) and hip fracture (0.67); in the cumulative data, the HRs were reduced for endometrial cancer (0.67) and hip fracture (0.81). For ET users, in the intervention phase the HR was increased for stroke (1.35) and reduced for hip fracture (0.67).

Estimated absolute risks (incidence rates in the non-exposed women subtracted from the rates in the exposed women) are given in Table 3. For E+P users, in the intervention phase the absolute risks ranged from –6 per 10 000 women per year (colorectal cancer; hip fracture) to 20 per 10 000 women per year (global index); in the cumulative follow-up data, the range was –5 (hip fracture) to 12 (global index). For ET users, in the intervention phase the range was –7 (breast cancer) to 11 (stroke); in the cumulative data, it was –2 (hip fracture) to 5 (stroke).

EVALUATION

Below we apply relevant epidemiological criteria of causation (or prevention)⁴⁻⁷ to the WHI evidence. The criteria are inter-related and, when appropriate, we cross-refer. But first, we consider the global index, which on conceptual grounds must be rejected in its entirety.

Table 1 Women’s Health Initiative: hazard ratios for intervention phase and cumulative follow-up of 13 years

<i>Outcome</i>	<i>Exposed/controls</i>	<i>Hazard ratio</i>	<i>95% confidence interval</i>
<i>E+P/placebo</i>	8506/8102		
Intervention phase			
coronary heart disease	196/159	1.18	0.95–1.45
breast cancer	206/155	1.24	1.01–1.53
global index	876/736	1.12	1.02–1.24
Cumulative follow-up			
coronary heart disease	487/430	1.09	0.96–1.24
breast cancer	434/323	1.28	1.11–1.48
global index	2198/1997	1.06	1.00–1.13
<i>ET/placebo</i>	5310/5429		
Intervention phase			
coronary heart disease	204/222	0.94	0.78–1.14
breast cancer	104/135	0.79	0.61–1.08
global index	753/755	1.03	0.93–1.13
Cumulative follow-up			
coronary heart disease	363/393	0.94	0.82–1.09
breast cancer	168/216	0.79	0.65–0.97
global index	1374/1397	1.02	0.94–1.09

E+P, estrogen plus progestogen therapy; ET, estrogen therapy

Table 2 Significantly elevated/reduced hazard ratios for outcomes other than coronary heart disease and breast cancer included in the global index

<i>Outcome</i>	<i>Exposed/controls</i>	<i>Hazard ratio</i>	<i>95% confidence interval</i>
<i>E+ P/placebo</i>	8506/8102		
Intervention phase			
stroke	159/109	1.37	1.07–1.76
pulmonary embolism	87/41	1.98	1.36–2.87
colorectal cancer	50/75	0.62	0.43–0.89
hip fracture	53/75	0.67	0.47–0.95
Cumulative follow-up			
endometrial cancer	68/96	0.67	0.47–0.91
hip fracture	212/270	0.81	0.68–0.97
<i>ET/placebo</i>	5310/5429		
Intervention phase			
stroke	169/130	1.35	1.07–1.70
hip fracture	48/74	0.67	0.46–0.90

E+ P, estrogen plus progestogen therapy; ET, estrogen therapy

The global index

In statistical terms, there is no reason why HRs estimated for an array of outcomes cannot be synthesized, with appropriate weighting, to derive a summary HR estimate. What is at issue, however, is whether in epidemiological or clinical terms that estimate, designated as a 'global index', can be considered an interpretable indicator of overall benefit or risk.

CHD, breast cancer, stroke, pulmonary embolism, colorectal cancer, endometrial cancer, and hip fracture are dissimilar clinical and pathological entities, with different age distributions, geographic distributions, ethnic distributions, genetic determinants, environmental determinants, dosage and duration effects, and causes other than HT (some known, and some unknown). And as for all-cause mortality, there are as many determinants of increased or decreased risk as there are causes of death, or more.

It is absurd to propose that the HRs associated with such a widely heterogeneous array of outcomes and their determinants can be synthesized to produce a single summary HR

that indicates whether the overall benefit of HT outweighs the risk, or vice versa. And, to the degree that it remains uncertain whether E+ P or ET does or does not increase or decrease the risk of individual components of the global index, their inclusion renders it all the more uninterpretable as a risk indicator. In addition, an incidence rate of an 'index' has no discernible meaning. A consideration of benefit and risk would be different, for example, for women with or without osteopenia, hypertension, hyperlipidemia, a family history of colorectal cancer, benign breast disease, or a previous history of venous thromboembolism.

The global index was also biased. An index was not specified a priori before the study commenced in 1993, and the protocol was first modified to include one in December 1996^{8,9}, 3 years after follow-up commenced. Formal monitoring began in 1997, and the index was again modified in June 1997. In April 1998, after the first analysis of the data had been conducted, it was once again modified, and vertebral fractures were removed⁸. The effect of the removal was to bias the index: in the intervention phase, for vertebral fracture the

Table 3 Absolute risks per 10 000 women per year

	<i>Estrogen plus progestogen therapy</i>		<i>Estrogen therapy</i>	
	<i>Intervention phase</i>	<i>Cumulative follow-up</i>	<i>Intervention phase</i>	<i>Cumulative follow-up</i>
Coronary heart disease	6	3	-3	-4
Breast cancer	9	9	-7	-7
Stroke	9	5	11	5
Pulmonary embolism	9	4	4	4
Colorectal cancer	6	-6	2	-3
Endometrial cancer	1	-3	NA	NA
Hip fracture	-6	-5	-6	-2
All-cause mortality	-1	-1	3	-1
Global index	20	12	4	1

NA, not applicable

respective HRs among E+P and ET users were 0.68 (95% CI 0.48–0.96) and 0.64 (95% CI 0.44–0.93). In the cumulative data, HRs for vertebral fracture were not given. The bias resulted in overestimation of the HR.

We turn next to the application of relevant causal principles to the WHI evidence.

Time order

At entry the women were screened for CHD and breast cancer, and, in respect of those endpoints the criterion of time order was satisfied.

Information bias

This was a prospective study, there was no misclassification of the exposures, and there was no information bias.

Detection bias

At recruitment, the women were informed of a hypothesized reduced risk of CHD among HT users. In the intervention phase, ‘unblinding’ was 6.9-fold more common among the E+P recipients than among the placebo recipients (44.4%/6.4%: see below). Awareness that this outcome was of interest could have resulted in overestimation of the magnitude of the HRs for CHD among E+P users.

Among the E+P users, biased estimation of an elevated HR for breast cancer was inevitable since the women were informed of a hypothesized increase in the risk when recruited³, and, during the clinical trial, in the exposed and non-exposed women the respective ‘unblinding’ rates, mainly because of vaginal bleeding, were 44.4% and 6.4%. Because of the differential ‘unblinding’ rates, the HRs for CHD (see above), stroke and pulmonary embolism could also have been overestimated. In the cumulative data, following discontinuation of the use of E+P, after 13 years of follow-up, 7.2 years of which were ‘unblinded’, among 100% of the participants knowledge of exposure status could further have biased the data.

By contrast, it is unlikely that the reduced HR estimates among E+P users for colorectal cancer, endometrial cancer, and hip fracture would also have been biased. If anything, detection bias would have tended to result in overestimation of the HRs, not underestimation, and at recruitment the women were not alerted to possible risk reductions.

Among ET users, it is unlikely that the increased risk of stroke would have been due to detection bias. In the intervention phase, ‘unblinding’ occurred in <2% of the women¹⁰ because they were hysterectomized and bleeding did not occur. Nevertheless, the increased risk must be interpreted with caution (see: confounding; strength of association; statistical stability).

With regard to the reduced risks among ET users of breast cancer and hip fracture, detection bias, if present, would have resulted in overestimation of the risk, not underestimation. Again, however, the risk reductions must be cautiously interpreted (see: confounding; strength of association; statistical stability).

Confounding

For the E+P users, uncontrolled confounding could readily have explained the increased risks of CHD, breast cancer, stroke, and pulmonary embolism. Since the respective ‘unblinding’ rates in the E+P and placebo recipients were 44.4% and 6.4%³, it is likely that the differential knowledge of exposure status would have given rise not only to detection bias, but also to confounding. High rates of non-adherence to treatment (E+P, 42%; placebo, 38%³) augmented that likelihood; during follow-up, 10% of the placebo recipients switched to HT, which would further have augmented it: the reasons for stopping or switching (e.g. hyperlipidemia, benign breast disease, osteopenia, hypertension) could independently have been associated with one or more of the outcomes (e.g. CHD, breast cancer, hip fracture, stroke).

To reduce confounding due to non-adherence, ‘intention-to-treat’ analyses in which the HRs were estimated according to the initially assigned treatment, regardless of whether or not it was stopped, were used¹⁰. In trials in which non-adherence rates are low (say, $\leq 10\%$), at the cost of somewhat attenuating the HRs, ‘intention-to-treat’ analyses can reduce confounding. In the E+P trial, however, the non-adherence rates were exceptionally high, as was the ‘unblinding’ rate (see: detection bias). In effect, both in the intervention and post-intervention phases the trial became an observational study. For that reason, ‘as treated’ analyses, with full control for individual confounders, should have been performed. In addition, to minimize both detection bias and confounding, ‘as treated’ analyses in women who remained ‘blinded’ during the intervention phase should also have been performed (see: detection bias)¹¹. And still further, ‘intention-to-treat’ analysis would not have reduced confounding among women who switched their treatments.

In the post-intervention phase, almost 20% of the women did not consent for follow-up, and the reasons for refusal (e.g. chest pain, a suspicious breast lump, hypertension) could still further have confounded the comparisons. Moreover, following publication of the initial WHI findings in 2002³, in the United States some 66% of menopausal women stopped using HT¹²; in the WHI report, the magnitude of the decline in the use of E+P was not mentioned, but it would have been substantially greater than in the intervention phase. Thus, in the analysis of the cumulative data, the use of ‘intention-to-treat’ methods was indefensible.

By contrast, in the ET trial, <2% of the participants were ‘unblinded’¹⁰, and confounding (and detection bias) was less likely than in the E+P trial. Nevertheless, since 53.8% of the women stopped their allocated treatments, and 9.1% of

the placebo recipients switched to HT¹⁰, some degree of confounding was possible. Again, the use of ‘intention-to-treat’ methods was not valid, and the increased risk of stroke could partly or wholly have been accounted for by confounding (see: strength of association).

Among the ET users, the risk reductions for breast cancer and hip fracture could also have been due to confounding. However, it is less likely that the reduced risk of breast cancer was confounded: in an earlier WHI report¹³, based on 10.7 years of follow-up, in an ‘as treated’ analysis of the combined intervention and post-intervention data, the HR for breast cancer was 0.68 (95% CI 0.49–0.95).

Strength of association

For the E+P and the ET recipients, all significantly elevated HR estimates were <2.0 (Tables 1 and 2), and, with the single exception of pulmonary embolism (E+P: HR, 1.98; intervention phase), they were <1.5. For such low estimates, particularly HRs of <1.5, it is impossible to discriminate among bias, confounding and causation as alternative explanations¹⁴. In observational studies, causation can only confidently be accepted if it is unlikely that bias or confounding can fully account for a substantially increased risk (say, a relative risk of ≥ 3.0). For example, among heavy smokers the risk of lung cancer is increased some 30-fold or more¹⁵.

Among the E+P users, the inverse values of the reduced HR estimates were 1.61 (1.00/0.62) for colorectal cancer, 1.49 (1.00/0.67) for endometrial cancer, and 1.49 (1.00/0.67) for hip fracture, and the same considerations apply. And among the ET users, the respective inverse values for breast cancer and hip fracture were 1.27 (1.00/0.79) and 1.49 (1.00/0.67), and again the same considerations apply.

Statistical stability

Among the E+P recipients, for pulmonary embolism the lower 95% confidence limit around the HR of 1.98 was 1.36 (intervention phase), and the estimate was statistically stable. For the remaining significantly elevated HRs, however, the lower 95% confidence limits were 1.01 (breast cancer; intervention phase), 1.11 (breast cancer; cumulative follow-up), 1.07 (stroke; intervention phase), and 1.07 (stroke; cumulative follow-up) (Tables 1 and 2). Thus, with the possible exception of pulmonary embolism, the significantly elevated HRs were statistically unstable, and it would have taken minimal detection bias or confounding to reduce the lower confidence limits to <1.00.

With regard to the significantly reduced HRs among the E+P recipients, the upper 95% confidence limits were 0.89 (colorectal cancer; intervention phase), 0.91 (endometrial cancer; cumulative follow-up), 0.95 (hip fracture; intervention phase), and 0.97 (hip fracture; cumulative follow-up). That is, the HR estimates for those outcomes were again statistically unstable, and it would have taken relatively little detection

bias or confounding to increase the upper confidence limits to >1.00.

Among the ET recipients, in the intervention phase, the lower 95% confidence limit around the significantly elevated HR for stroke was 1.07, and the upper limit for hip fracture was 0.90. Again, it would have taken minimal detection bias or confounding to nullify the statistical significance of those associations. However, for breast cancer, since the respective upper confidence limits around the HR of 0.79 were 1.08 (intervention phase) and 0.97 (cumulative follow-up) (Table 1), the evidence to suggest no increase in the risk was statistically robust.

Duration–response effects

In the WHI report¹, duration of use of E+P or ET was not mentioned, but in the intervention phases the discontinuation rates ranged from 38%³ to 53.8%¹⁰, and the average durations of use would have been substantially less than 5.6 years (E+P) or 7.2 years (ET). In addition, since most of the women stopped using HT soon after the E+P trial ended, in the cumulative data the average durations of use would have been substantially less than 13 years. For those reasons, duration–response effects cannot be evaluated.

Internal consistency

In published WHI reports, the results for the primary outcomes have been inconsistent. For CHD, among the E+P recipients, in the initial report (2002)³ the HR was 1.29 (95% CI 1.02–1.63) after a mean of 5.2 years of follow-up; yet, after the same duration of follow-up, when what were represented as the ‘final results’ were published (2003)¹⁶, the HR was 1.24 (95% CI 1.00–1.54). And after 5.6 years¹ of follow-up (intervention phase), the HR was 1.18 (95% CI 0.95–1.45) (Table 1). That is, over an interval of an additional 0.4 years (5.6–5.2) of follow-up, inconsistent but ‘statistically significant’ HRs of 1.29 and 1.24 changed to a ‘non-significant’ value of 1.18, and after 13 years of follow-up to 1.09 (0.96–1.24) (Table 1).

The WHI investigators stated that ‘some HRs differ slightly from those previously reported due to the more complete outcome ascertainment in the present report’¹. However, the shift from statistical significance^{3,16} to non-significance¹ was not ‘slight’: in the initial WHI publication³, it was stated that the HR of 1.29 ‘[reached] nominal statistical significance’, and that ‘a sufficient number of CHD events had occurred by 5.2 years of follow-up to suggest that continuation to the planned end would have been unlikely to yield a favorable result’. The initially reported increased risk of CHD among E+P recipients, not subsequently confirmed, had a major public health impact.

For breast cancer, in a previous WHI report¹⁷, ‘the increased risk of breast cancer associated with the use of [E+P] declined markedly soon after discontinuation of combined [HT]’. Yet

a decline was not evident after 13 years of follow-up¹ when the majority of the women would have discontinued the use of HT years previously: the respective HRs in the intervention phase and in the cumulative follow-up data were 1.24 and 1.28 (Table 1).

The designation of breast cancer as an endpoint also changed in the course of the study. In the original protocol¹⁸, it was listed as a secondary outcome. Among the secondary outcomes, breast cancer was then designated as the primary adverse outcome⁸. Then, in the WHI report¹, together with CHD, breast cancer was designated as a primary endpoint. The shifts in the designation from a secondary outcome, to a primary adverse outcome, and then to a primary endpoint were not explained. Had breast cancer remained a secondary outcome as originally specified, the adjusted lower 95% confidence limit around the initially reported HR of 1.26 would have been 0.83³, and the association could have been due to chance (see: statistical stability).

External consistency

In the Collaborative Reanalysis¹⁹, among current users of HT an increased relative risk of breast cancer declined to baseline within 5 years of stopping; in the Million Women Study²⁰, the increased risk among HT users declined within 2 years. Yet, in the WHI study, the HR for E+P users remained elevated after 13 years of follow-up, when the majority of the women had stopped using HT years earlier.

In the Collaborative Reanalysis¹⁹ and Million Women Study²⁰, the relative risks for breast cancer were elevated for ET users; in the WHI¹, the HRs were reduced.

With regard to CHD, the preponderance of the evidence from observational studies^{21,22} suggests that HT reduces the risk, perhaps most markedly in recently menopausal women. Yet, in the WHI study, a risk reduction was not evident (see: detection bias).

Biological plausibility

HT has beneficial effects on many metabolic and vascular risk factors²², and a reduced risk of CHD, although biologically plausible, was not observed in the WHI study, possibly because of detection bias.

With regard to the risk of breast cancer among E+P users, some mechanisms may increase it, and other mechanisms may decrease it^{23,24}. However, based on the doubling times for the multiplication of malignant cells, it takes at least 10 years for breast cancer to become clinically evident, and among E+P users the increased risk in the intervention phase was biologically implausible. For the risk reduction among ET users, since some mechanisms may reduce the risk, while other mechanisms may increase it^{23,24}, biological plausibility cannot be assessed.

With regard to stroke and venous thromboembolism, among women of child-bearing age, the use of oral contraceptives

increases the risks²⁵, and it is plausible that HT may have the same effects in menopausal women.

With regard to hip fracture, HT has beneficial effects on menopausal bone loss²⁶, and a risk reduction is plausible.

CONCLUSIONS

With regard to CHD, earlier observational and experimental studies^{21,22} raised the possibility that HT may reduce the risk. The original protocol-specified aim in the WHI study¹⁸ was to evaluate that possibility and, surprisingly, in the initial WHI report³ it was instead claimed that E+P increases it. Based on the total evidence, however, the initial claim has not been confirmed, and the WHI study has not established that E+P increases the risk of CHD; nor has it ruled out the possibility that HT may even reduce the risk.

The increased risk of breast cancer observed among E+P users did not satisfy the criteria of detection bias, confounding, strength of association, statistical stability, internal consistency, external consistency, and biological plausibility. E+P may or may not increase the risk of breast cancer, but the WHI findings did not establish that it does.

For ET users, the data were sufficiently stable statistically to rule out an increased risk of breast cancer. It is even possible that ET reduces the risk, but the evidence to support that possibility was statistically unstable, and it needs to be independently confirmed.

With regard to the increased risk of stroke among E+P and ET users, evidence from studies of oral contraceptive use²⁵ supports the possibility of some increase in the risk, some limitations to the causal criteria in the WHI study notwithstanding. However, the increased risk needs to be independently confirmed. For pulmonary embolism, the association was statistically robust and consistent with other evidence²⁵.

With regard to the decreased risks of colorectal and endometrial cancer in E+P users, the findings are credible²⁷, although statistically borderline. Again, the risk reductions need to be independently confirmed.

With regard to hip fracture, the decreased risks are again credible²⁶. In addition, in the intervention phase, the risks of vertebral fracture (E+P: HR 0.68, 95% CI 0.48–0.96; ET: HR 0.64, 95% CI 0.44–0.93), and of all fractures (E+P: HR 0.76, 95% CI 0.69–0.83; ET: HR 0.72, 95% CI 0.64–0.80) were also reduced¹, adding further to the credibility of the evidence. Again, however, the risk reductions need to be independently confirmed.

If, despite the limitations of the WHI study, causality (or protection) is assumed, and the ‘worst case’ assumption is made that all the elevated HRs were causal, and the ‘best case’ assumption is made that all reduced HRs were protective, what are the public health implications? For the primary endpoints, and the endpoints included in the global index, the absolute risks ranged from –7 per 10 000 women per year (breast cancer in ET users) to 11 per 10 000 women per year (stroke in ET users) (Table 3). Or put another way, at most the changes in the incidence of CHD, breast cancer,

stroke, pulmonary embolism, colorectal cancer, hip fracture, or all-cause mortality attributable to the use of HT would have been minor.

For the global index, the absolute risks, ranging from 1 to 20 per 10 000 women per year (Table 3), must be disregarded since, in epidemiological and clinical terms, a synthesized summary estimate of benefit and risk is uninterpretable. That index was a *post hoc* invention of the WHI investigators, and the selection of the included components, one of which was subsequently removed, was also biased.

Possible benefit versus risk needs to be separately assessed for each particular population and outcome under study, and women with menopausal symptoms, osteoporosis, or a family history of colorectal cancer benefit from the use of E+P; E+P users are also at reduced risk of endometrial cancer. Hysterectomized women benefit from the use of ET, and the WHI findings now raise the possibility that ET may reduce the risk of breast cancer. In addition, the preponderance of the epidemiological evidence suggests that recently menopausal women who use HT are at reduced risk of CHD^{21,22}.

Finally, hormones other than conjugated equine estrogen and medroxyprogesterone acetate were not evaluated in the WHI study, as has been acknowledged by the investigators. The products currently in use have different biological effects and are prescribed in low doses²⁸. For these reasons, further studies are needed.

We conclude that over-interpretation and misrepresentation of the findings in the WHI study have resulted in major damage to the health and well-being of menopausal women. The WHI was not ‘a victory for women and their health’², and the claim that the findings ‘do not support the use of this therapy for chronic disease prevention’¹ is not defensible. Nor can the pejorative editorial statement that ‘the WHI overturned medical dogma regarding menopausal [HT]’² be defended.

Conflict of interest SS, TJD, AP, DWS, DEA, RJB, NP, JCS, AOM and HGB presently consult, or in the past have consulted, with manufacturers of products discussed in this article. All authors participated in the preparation of this article.

Source of funding Nil.

References

- Manson JE, Chlebowski RT, Stefanick ML, *et al.* Menopausal hormone therapy and health outcomes during the interventions and postintervention and extended poststopping phases of the Women’s Health Initiative randomized trials. *JAMA* 2013;310:1353–68
- Nabel EG. The Women’s Health Initiative – a victory for women and their health. *JAMA* 2013;310:1349–50
- 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;288:321–33
- 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: US Government Printing Office, 1964
- Hill AB. The environment and disease: association or causation? *J Proc R Soc Med* 1965;58:295–300
- Susser M. Causal thinking in the Health Sciences. New York: Oxford University Press, 1979
- Susser M. What is a cause and how do we know one? *Am J Epidemiol* 1991;133:635–48
- Freedman L, Anderson G, Kipnis V, *et al.* Guidelines for the statistical monitoring of the Women’s Health Initiative clinical trial: April 1, 1998. WHI website. Protocol for clinical trial and observational components. cleo.whi.org/study/doc/consents/whi%20protocol.pdf
- Anderson GL. WHI results withstand the test of time. *Br Med J Rapid Response BMJ Fri*, 2010-93-12 13:47
- 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;291:1701–12
- Shapiro S, Farmer RDT, Mueck AG, Seaman H, Stevenson JC. Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies. 2. The Women’s Health Initiative: estrogen plus progestogen. *J Fam Plann Reprod Health Care* 2011;37:165–72
- Hersh AL, Stefanick ML, Stafford RL. National use of menopausal hormone therapy: annual trends and response to recent evidence. *JAMA* 2004;291:47–53
- LaCroix AZ, Chlebowski Rt, Manson JT, *et al.* Health outcomes after stopping conjugated equine estrogen among postmenopausal women with prior hysterectomy. *JAMA* 2011; 305:1305–14
- Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol* 2000;151:939–45
- IARC Monographs on Carcinogenic Risk to Humans 2004. *Tobacco Smoke and Involuntary Smoking*. Volume 83. Lyon: International Agency for Research on Cancer (IARC), 2004
- Manson JE, Hsia J, Johnson KC, *et al.* Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 2003; 349:523–34
- Chlebowski RT, Kuller LH, Prentice RL, *et al.* Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med* 2009;360:573–87
- Study protocol and policies. WHI Manuals 1998;1:1–64 [sect 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;350:1047–59
- Million Women Study collaborators. Breast cancer and hormone replacement therapy in the Million Women Study. *Lancet* 2003; 362:419–27
- Stevenson JC, Hodis HN, Pickar JH, Lobo RA. Coronary heart disease and menopause management. The swinging pendulum of HRT. *Atherosclerosis* 2009;207:336–40
- Maclaren K, Stevenson JC. Primary prevention of cardiovascular disease with HRT. *Women’s Health* 2012;8:63–74

23. 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;20:2052–60
24. Mueck AO, Seeger H, Shapiro S. Risk of breast cancer during hormone therapy: mechanisms. *Horm Biol Clin Invest* 2010;3:329–39
25. Petitti DB. Combination estrogen-progestin oral contraceptives. *N Engl J Med* 2003;349:1443–50
26. De Villiers T, Stevenson JC. The WHI: the effect of hormone replacement therapy on fracture prevention. *Climacteric* 2012;15:263–6
27. Barrett-Connor E. Hormone replacement therapy – risks and benefits. A review. *Am J Epidemiol* 2001;30:423–6
28. Panay N, Ylikorkola O, Archer DF, et al. Ultra-low dose estradiol and norethisterone acetate: effective menopause symptom relief. *Climacteric* 2007;10:120–31