



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

Part 1. The Collaborative Reanalysis

Samuel Shapiro,¹ Richard D T Farmer,² Helen Seaman,^{2,3}
John C Stevenson,⁴ Alfred O Mueck⁵

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

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

³Freelance Medical Writer, Aldershot, UK

⁴National Heart and Lung Institute, Imperial College and Royal Brompton Hospital, London, UK

⁵Department of Endocrinology, University Women's Hospital of Tuebingen, Tuebingen, 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 18 August 2010

Accepted 28 December 2010

Abstract

Background Concern that hormone replacement therapy (HRT) may cause breast cancer has existed since the time it was introduced, and based on evidence in three studies, the Collaborative Reanalysis (CR), the Women's Health Initiative (WHI) and the Million Women Study (MWS), it is claimed that causality is now established.

Objective To evaluate the evidence for causality in the three studies.

Methods Using generally accepted causal criteria, in this paper the authors begin with an evaluation of the CR. Analogous evaluations of the WHI and MWS will follow.

Results The findings in the CR did not adequately satisfy the criteria of time order, bias, confounding, statistical stability and strength of association, dose/duration-response, internal consistency, external consistency or biological plausibility.

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

Background

The publication in 1997 of a meta-analysis [the Collaborative Reanalysis (CR)] of 51 studies of the risk of breast cancer in relation to the use of hormone replacement therapy (HRT)¹ marked a watershed in the public perception. Before that date it was generally thought that HRT may increase the risk, but the link was uncertain and unproven;² now it was claimed that the synthesised evidence across the studies established that HRT “increases the risk of having breast cancer diagnosed”. Then, in 2002 it was claimed that any lingering doubts about causality should be dispelled by the findings in a

randomised controlled trial, the Women's Health Initiative (WHI).³ And based on additional findings in the Million Women Study,⁴ published a year later, it is now widely believed that HRT is a substantial and significant cause of breast cancer.

In this series of articles we evaluate whether the three studies accorded with generally accepted principles of causality.^{5–8} We begin with the CR.

The Collaborative Reanalysis¹

Data were pooled from 51 studies (mostly case-control studies), representing >90% of all studies published before 1997. Cohort studies were included by re-casting them as nested case-control studies, with four randomly selected controls matched to each case. Data on individual women were provided by the original investigators so that the “analyses could, as far as possible, use similar definitions across studies”.¹ There were 52 705 premenopausal and postmenopausal women with breast cancer and 108 411 controls. The analysis of HRT use was confined to postmenopausal women [17 949 cases (34%); 35 916 controls (33%)].

In analyses adjusted for confounding aggregated relative risks (RRs) and their 95% or 99% confidence intervals (CIs) (or standard errors or two-tailed *p* values) were presented. When more than two groups were compared floated estimates were presented, in which the RRs were unchanged, and any two groups could be compared “even if neither [was] the baseline group”.

The RR for ever-use versus never-use of HRT (stratified by study, age at diagnosis, body mass index (BMI), age at first

birth, parity and time since menopause) was 1.14 ($2p = 0.00001$). In the pooled hospital-based and population-based case-control studies the RRs were 1.27 and 1.15, respectively, both statistically significant; in the cohort studies the RR was 1.09, and not significant. For durations of ever-use of <1, 1–4, 5–9, 10–14 and ≥ 15 years the RRs were 1.09, 1.05, 1.19 and 1.58, respectively (trend $p = 0.003$).

Among cases using HRT when diagnosed (current users), the RR was 1.21 ($2p = 0.00002$) and following cessation of use it declined to 1.10 after 1–4 years, and to 1.01 after 5–9 years. Overall, among women who were current HRT users, or who had stopped <5 years previously, and who had used HRT for ≥ 5 years, the RR was 1.35 (95% CI 1.21–1.49). In that category, for durations of use of <1, 1–4, 5–9, 10–14 and ≥ 15 years the RRs were 0.99, 1.08, 1.31, 1.24 and 1.56, respectively. In the latter analysis “the risk increased by a factor of 1.023 ([standard error] 0.060) – i.e. by 2.3% [0.6% (sic)] – for each year of use ($2p = 0.0002$)”.

“Cancers diagnosed in women who had ever used HRT tended to be less advanced clinically than those diagnosed in never-users”; and among women who were current users or who had stopped <5 years previously, only the risk of localised cancer was increased, while the risk of widespread disease was not. However, for widespread breast cancer “there was a significant increase in the [RR] . . . with increasing duration of use (trend test, $2p = 0.007$) . . . largely because women who began using HRT in the 5 years before their cancer was diagnosed had a low [RR] of spread (sic) disease (RR 0.59; $2p = 0.001$)”.

The investigators concluded that “the risk of having breast cancer diagnosed is increased in women using HRT and increases with increasing duration of use”, but this excess risk “. . . is reduced after cessation of use of HRT and has largely, if not wholly, disappeared after about 5 years”. Based on “incidence rates. . . typical for women in North America or Europe. . .” they estimated that among women aged 50–70 years “. . . use [of HRT] for about 13 years would result in one extra cancer being diagnosed in every 100 users”.

Evaluation of the CR

The validity of meta-analysis as a tool in causal research has been debated.^{9–13} It is therefore important to consider whether the conclusions in the CR report were justified, and whether they accorded with generally accepted principles of causality.^{5–8} The principles are interrelated, and below, when appropriate, we cross-refer.

Time order

As is obvious, HRT cannot cause breast cancer if first used only after its onset. But what is meant by the term ‘onset’ is not straightforward. Broadly, carcinogenesis commences after damage to (e.g. X-rays), or spontaneous mutations in, cellular genes (initiation). Following

initiation, on average it takes at least 5–10 years before clinical breast cancer develops (promotion).^{14 15} The hypothesis, therefore, is not whether HRT initiates carcinogenesis, but whether its use accelerates the multiplication and malignant transformation of cells already initiated. [Under a promotional hypothesis it is perhaps also possible that HRT may accelerate the growth of already existing breast cancer (see Biological plausibility).]

Since the evolution of cancer is a continuous process, determining the date of onset becomes arbitrary, and what has generally been done has been simply to specify the date of diagnosis as an index date. Sometimes breast cancer can remain ‘clinically silent’ unless actively searched for, and otherwise occult cases can be detected by examining the breast, or failing that, by mammography.¹⁶ If not actively searched for, slow-growing tumours may go undetected, sometimes for years. In autopsies of postmenopausal women who have died of unrelated causes, ‘clinically silent’ breast cancer has been found in about 5%,^{17–19} and among women who already have breast cancer, undetected cancer is commonly found in the seemingly normal breast.²⁰ Nor does the difficulty end there: on average, the more advanced the tumour is, the longer it has been present, but for how long is unknown, rendering determination of the actual date of onset even more uncertain.

As there can be no certainty about the date of onset of clinical breast cancer, limited reassurance that time order has not been violated can only be gained by placing greatest reliance on exposures that can reasonably be assumed to have commenced well before the index date. If, for example, HRT use began, say, ≥ 10 years earlier, it may be reasonable to assume that the exposure came first. Note, however, that the measured duration of HRT use, and the measured intervals since the commencement or termination of use, are often uncertain. Moreover, if current HRT use has only been brief – say, a year or two – it is impossible to be sure whether the cancer or the exposure came first.

In the CR the requirement that HRT use should unambiguously have commenced before the index date was not met. The earliest reported median year of diagnosis was 1974²¹ and the latest was 1992.²² Over that interval mammography rates increased,¹⁹ and they would have given rise to progressively earlier diagnosis. The rates also differed by ethnic group, being less common among black women in the USA, where the majority of the studies were performed. In addition, in many studies there was no information on tumour size or stage. Hence, in the different studies the index dates shifted. In addition, women aware of as yet undiagnosed breast lumps could selectively have participated in the studies (see Detection bias).

Based on these considerations, among current HRT users there could be no reassurance that short-duration use commenced before the onset of clinically

or mammographically detectable breast cancer (see Detection bias); the duration data were imprecise; and the specified intervals after stopping HRT use were also imprecise. It follows that the estimated duration-dependent incidence rates attributable to HRT use were also uncertain (see Dose/duration-response), as was the variation in the RRs at varying intervals after stopping HRT use.

Bias

Anxiety about the possibility that HRT may cause breast cancer has existed since its introduction, and has increased over time. That anxiety could have given rise to information and detection bias, and it could have done so across multiple studies.

Information bias

Most of the CR data were derived from interview- or questionnaire-based case-control studies, and it is likely that the cases would have been at pains to recall all episodes of HRT use. By contrast, the controls could have under-reported their actual use, especially if it was short-term, or if it had stopped years earlier. They may also have under-reported the duration of use.

The investigators acknowledged the possibility of information bias in the case-control studies, but claimed that it was unlikely because the results were similar in the cohort studies. In fact, the results were dissimilar. In the case-control studies, in which information bias was likely, for ever-use of HRT the RRs were 1.27 (hospital controls) and 1.15 (population controls), and both estimates were statistically significant; in the cohort studies, in which bias was unlikely, the RR was 1.09, and not significant. There was also quantitative evidence to support the likelihood of bias in the case-control studies: in the hospital-based data, in which the RR was highest, ever-use of HRT by the controls was the lowest (12.2% – our calculation: Figure 3 in the CR report); in the population-based studies the RR was lower, and ever-use among the controls was higher (32.8%).

How much information bias would it have taken to account for the findings? The rates of ever-use of HRT in the cases and controls were 30.5% and 34.4%, respectively (our calculation: Figure 3 in the CR report), a difference of 3.9%. That difference implies that the overall association could have been accounted for by under-reporting of ever-use of HRT among the controls of the order of 3.9%. [Although ever-use of HRT was lower in the cases than in the controls (crude RR, 0.84 – our calculation) the confounder-adjusted RR was 1.14 (see Confounding).]

Detection bias

When HRT is prescribed, women are advised to have regular breast examinations and mammograms, and there is quantitative evidence that HRT users undergo mammography more frequently than do non-users.²³ Thus it is likely that detection bias resulted in a selective

tendency to more commonly diagnose otherwise occult breast cancer among HRT users. In addition, the longer the duration of use, the more often would screening have taken place, and the increase in the RR with increasing duration of HRT use could also have been due to detection bias (see Dose/duration-response).

Detection bias may also have influenced the radiologists who interpreted the mammograms. Combined estrogen/progestogen compounds increase the radiological density of breast tissue,^{24–26} and when a mammogram is performed it is standard practice to record HRT use. Hence, if a woman is a user, and if in addition the breast tissue is dense, cancer is likely to be more intensively searched for than among women who are not HRT users, and who have normal breast density. Moreover, there is ample scope for differential detection: in the presence of increased breast density, some 30% of breast cancers go undiagnosed on mammography.²⁵

The authors claimed that detection bias due to mammography did not influence their results. Yet most of the included studies had no data on mammography, and when they did, no distinction was made between mammograms used for screening, as against their use in diagnostic work-ups of women with already suspected or diagnosed breast cancer.

There was evidence of detection bias in the CR data: “cancers diagnosed in women who had ever used HRT tended to be less advanced clinically than those diagnosed in never-users”, and among current HRT users, only the risk of localised breast cancer was increased. The authors acknowledged that detection bias due to “more frequent mammographic or other examinations” could have accounted for those findings in the case-control studies, but asserted that this was unlikely because “the results were similar in prospective studies, where no such bias could have occurred”. That assertion was incorrect: in both the case-control and prospective studies it is likely that mammographic screening would have been more common among HRT users than among non-users, and that the mammograms could also have been more intensively scrutinised.

The decline in the risk of breast cancer with increasing BMI was further evidence of detection bias (see Internal consistency): on average, the more obese a woman is, the larger are her breasts, and the less likely is it that otherwise occult breast cancer will be detected.

Among HRT users whose cancer had spread beyond the breast, the authors stated that the absence of an increase in the risk was “largely because women who began using HRT in the 5 years before their cancer was diagnosed had a low [RR] of spread (sic) disease (RR 0.59; $2p = 0.001$)”, but the risk nevertheless increased with increasing duration of use (trend test, $2p = 0.007$). Under a causal hypothesis HRT cannot at the same time increase the risk of localised disease, but decrease the risk of widespread disease

(see Internal consistency). And how the use of HRT that commenced in the 5 years before diagnosis can have brought about a statistically significant 1.69-fold reduction (1.00/0.59) in the risk of advanced breast cancer was not explained. In addition, for advanced disease the duration-response effect was identified *post hoc* in a subgroup analysis: although it was statistically significant it could nevertheless have been due to chance, or possibly, to repeated screening (see Dose/duration-response).

It is likely that there was still a further source of detection bias. In both the case-control and cohort studies, women who were already aware that they had breast lumps, and who were also users of HRT, could selectively have tended to enrol in the studies, and again that tendency has been documented with quantitative evidence.^{3,23} It is also likely that in the cohort studies women who became aware of breast lumps would less commonly have been lost to follow-up.

How much detection bias would it have taken to account for the findings? The investigators estimated that among women between the ages of 50 and 70 years the use of HRT “. . . for about 13 years would result in one extra cancer being diagnosed in every 100 users”. Hence, if detection bias augmented the number of diagnosed cases of breast cancer by about 0.08 per 100 users per year, that bias would have accounted for the association (1/13 = 0.08).

Table 1 Distribution of potentially confounding factors among ever-users of hormone replacement therapy in the Collaborative Reanalysis

Factor	Cases (n = 5482)*	Controls (n = 12 348)*
	Known [†] (% unknown)	Known [†] (% unknown)
Controlled [‡]		
Age at diagnosis	3692 (32.7)	8076 (34.6)
Parity	3679 (32.9)	8038 (34.9)
Age at first birth	3018 (44.9)	6860 (44.4)
Time since menopause [§]	–	–
Body mass index	3334 (39.2)	6860 (44.4)
Other [¶]		
Family history	3470 (36.7)	7719 (37.5)
Ethnic group	3041 (44.5)	7128 (42.3)
Education	3600 (34.3)	7899 (36.0)
COC use in past 10 years	3344 (39.0)	7267 (41.1)
Alcohol	2692 (50.9)	4899 (60.3)
Smoking history	3167 (42.2)	6133 (50.3)
Type of menopause	3692 (32.7)	8076 (34.6)

*Denominators from Figure 3 in the CR report.

[†]Numerators from Figure 6 in the CR report.

[‡]Potentially confounding factors controlled by stratification.

[§]Data not provided in the CR report.

[¶]Other potentially confounding factors.

COC, combined oral contraceptive.

Confounding

The authors stated that “data on individual women were sought so that analyses could, as far as possible, use similar definitions across studies”. What is meant by “as far as possible” is open to different interpretations, and in the different studies some variables (e.g. family history of breast cancer, history of hysterectomy, alcohol consumption) were recorded differently. There can be no reassurance that the definitions were indeed similar.

Not only were the variables differently recorded in the different studies, commonly they were not recorded at all. Unknown values for potentially confounding factors were not mentioned in the CR report, but for ever-users of HRT they can be derived from Figures 3 and 6 (Table 1: our calculations). Among the factors controlled in the CR analysis information was missing for 32.7–44.9% of the exposed cases and controls. Among other potential confounders listed in the table, the corresponding range was 32.7–60.3%.

As mentioned above, based on respective rates of ever-use of HRT of 30.5% and 34.4% among the cases and controls, the crude RR was 0.84, while the adjusted RR was 1.14. The shift in the RR was not explained, but presumably it was due to adjustment for confounding. However, since information was missing for at least 30% of all the potential confounders listed in Table 1, that adjustment was inadequate.

In other studies,^{27–31} when individual factors have been allowed for, the effect on the RR estimates has generally been minor (see, for example, Newcomb *et al.*³¹). In the CR, however, full allowance for the combined effect of multiple variables could not be made because of missing information, and substantial confounding could not be ruled out. In addition, the RRs were so small (see Statistical stability and strength of association) that even relatively minor uncontrolled confounding could have accounted for the findings.

Statistical stability and strength of association

In order to interpret the statistical significance of the findings in the CR report it is helpful first to consider two hypothetical studies, one small and one massive. Assume that the same RR is observed in both. In the small study, if it is small enough, the RR is not statistically significant; in the massive study, if it is massive enough, it is. Assume further that the association is not causal, but due to bias or confounding – and in a meta-analysis that all or most studies share much the same biases. Then if a massive study is sufficiently massive, any deviation of a RR from 1.0, no matter how small, becomes ‘significant’.

The strength of an association is also a determinant of statistical significance. If a RR is markedly elevated, say 5.0 or greater (‘large RR’), it can readily be shown to be significant in a relatively small study, and be confirmed in a few more studies. But if it is only slightly elevated, say well under 2.0 (‘small RR’),

it takes a massive study to show that the association is significant.³²

No epidemiological study is perfect, and it is almost never possible to be confident that bias or confounding can be ruled out entirely. However, 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 CR the RRs for HRT use were small, mostly <1.5. As mentioned above, for ever-use the unadjusted RR was 0.84, and the adjusted RR was 1.14. Since the shift in the RR of 0.3 (1.14–0.84) indicates that controlled sources of confounding of that magnitude occurred in the CR data, it is reasonable to infer that other uncontrolled sources could have changed the RR estimate, upwards or downwards, by about the same amount. It follows that confounding due to incompletely controlled or uncontrolled factors, such as obesity, socioeconomic status, or ethnic group, could have accounted for the association (see above). It also follows that information or detection bias could have accounted for it (see above).

Dose/duration-response

Under a promotional hypothesis it might reasonably have been 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.

Dose

The doses in the different studies changed over time, and varied geographically. Often the doses were not recorded; sometimes not even the names of the compounds were recorded. The compounds that were used also changed over time, and they varied in their estrogenic and progestogenic potency, so that stipulation of dose equivalence would in any case not have been feasible. In short, dose-response was not evaluated.

Duration

Among women who last used HRT within 5 years of diagnosis it was claimed that the risk of breast cancer increased with increasing duration of use (for use that ended ≥ 5 years previously no duration effect was apparent). However, the statistical significance of the duration effect may have been incorrectly estimated.^{33–34} In any analysis of HRT use versus non-use the relevant reference category should, of course, be never-use. However, in further assessing whether increasing duration of use is associated with a progressively increasing RR, the question is not whether, relative to *never-use* there is a statistically significant trend, but whether, relative to the *shortest duration of use* (in the present instance

<1 year), there is a significant duration-dependent gradient of increasing risk. For the latter analysis never-use is irrelevant, and it should be omitted.

Not only was the duration trend incorrectly assessed, but among women last exposed within 5 years of diagnosis all 99% floated CIs overlapped, and there were no statistically significant differences between any two duration categories. At the extreme, the RR of 1.56 for ≥ 15 years of use did not differ significantly from the RR of 0.99 for <1 year of use. In addition, the trend was not monotonic: the RR of 1.24 for 10–14 years of use was lower than the RR of 1.31 for 5–9 years of use. And still further, even if the trend was correctly estimated, it could still have been accounted for by detection bias due to repeated screening (see above).

It follows that the claim that among women last exposed within 5 years of diagnosis there was a 2.3% cumulative increase in the incidence of breast cancer attributable to the use of HRT with each additional year of use was not supported by the evidence.

Internal consistency

As described above, among women who used HRT in the 5 years before the index date the RRs among categories of BMI, and among women with localised and advanced breast cancer, were inconsistent.

External consistency

Among the 51 studies included in the CR the RRs were not statistically heterogeneous. However, tests for heterogeneity among studies are not robust.^{10–35} Studies vary in quality, quality cannot be quantified, and in the absence of statistically significant variability, any judgement as to whether or not they are heterogeneous must necessarily be qualitative. Bush and her colleagues conducted a qualitative review of all studies (largely the same as those included in the CR) of breast cancer risk in relation to the use of HRT, published from 1975 to 2000.² They judged the findings for ever-use of HRT to be inconsistent. They concluded that “although a small increase in breast cancer risk . . . or an increased risk with long duration of use (≥ 15 years) cannot be ruled out, the likelihood of this must be small, given the large number of studies conducted to date”.

In the CR >80% of the HRT use was conjugated estrogen without an added progestogen, and the increased risk identified in the CR was inconsistent with the WHI study of estrogen use,³⁶ in which the risk of breast cancer was significantly *decreased* among women who adhered to treatment.

Biological plausibility

For causation to be fully confirmed, any observed association should be compatible with established pathological mechanisms. Under a promotional hypothesis, the underlying assumption is that HRT accelerates the multiplication of initiated cells, so that clinically evident breast cancer develops sooner than would otherwise be the case (and it is speculated that HRT may also

accelerate the growth of otherwise slowly growing cancer that is already present).^{18 19} Therefore estrogens, and probably progestogens as well, as known proliferative factors could possibly enhance the median tumour doubling time, assessed for the most aggressively multiplying cells to be about 50–100 days,^{14 15 37} and it is generally accepted that 30–35 doublings are required to attain a tumour diameter of 1 cm, which is about the smallest lesion that can be diagnosed clinically. Thus on average at least 5–10 years will have elapsed from tumour initiation in a single cell to the time of diagnosis.

Once malignant transformation has occurred, breast cancer, especially if already invasive, cannot ‘untransform’ when HRT is stopped. It is important to stress that the outcome under study was *invasive*, not *preinvasive*, breast cancer. In the CR, the risk of *invasive* cancer began to decline immediately, and it was no longer increased 5 years after stopping HRT. At most, it can perhaps be speculated that carcinoma *in situ* may occasionally revert to a premalignant state, although this has still to be proven.^{18 19} But it is not conceivable that *invasive* cancer can do so. Alternatively, it may be speculated that some *invasive* tumours may cease to grow after stopping HRT, or only grow slowly,^{18 19} but in that case such tumours would still be detected by mammography – and women who stop would not at the same time stop having mammograms.

In one respect the findings were perhaps biologically plausible, since increased radiological breast density is associated with an increased risk of breast cancer.^{24–26} However, it has not been shown that the increased density specifically induced by HRT increases the risk.^{25 38–40} In addition, as noted above it is likely that detection bias was most marked among women with increased breast density who also used HRT.

Conclusions

The meta-analysis reviewed here was termed a ‘CR’ because the ‘raw data’ from the individual studies were obtained, standardised to the degree possible, and only then synthesised. In carrying out those procedures the original investigators were consulted. The implication is that the CR was therefore more valid than meta-analyses, which have synthesised only published data. The evidence, however, is that it was not valid, despite standardisation and consultation. In terms of time order, bias, confounding, statistical stability and strength of association, dose/duration-response, internal consistency, external consistency and biological plausibility, the study was defective.

We conclude that CR did not adequately satisfy the principles of causation. HRT may or may not increase the risk of having breast cancer diagnosed, but the CR did not establish that it does.

Competing interests Samuel Shapiro, John Stevenson and Alfred Mueck presently consult, and in the past have consulted, with manufacturers of products

discussed in this article. Richard Farmer has consulted with some manufacturers in the past.

Provenance and peer review Not commissioned; externally peer reviewed.

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 Bush TL, Whiteman M, Flaws JA. Hormone replacement therapy and breast cancer: a qualitative review. *Obstet Gynecol* 2001;98:498–508.
- 3 Rossouw JE, Anderson GL, Prentice RL, *et al.*; 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;288:321–323.
- 4 Beral V; Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003;362:419–427.
- 5 Hill AB. The environment and disease: association or causation? *Proc Royal Soc Med* 1965;58:295–300.
- 6 Susser M. *Causal Thinking in the Health Sciences*. New York, NY: Oxford University Press, 1979.
- 7 US Department 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: U.S. Government Printing Office, 1964.
- 8 Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991;133:635–648.
- 9 Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994;140:771–778.
- 10 Greenland S. Can meta-analysis be salvaged? *Am J Epidemiol* 1994;140:783–787.
- 11 Petitti DB. Of babies and bathwater. *Am J Epidemiol* 1994;140:779–782.
- 12 Petitti D. *Meta-Analysis, Decision Analysis and Cost-Effectiveness in Medicine: Methods for Quantitative Synthesis in Medicine* (2nd edn). New York, NY: Oxford University Press, 1994.
- 13 Shapiro S. Is meta-analysis a valid approach to the evaluation of small effects in observational studies? *J Clin Epidemiol* 1997;50:223–229.
- 14 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.
- 15 Mueck AO, Seeger H, Shapiro S. Risk of breast cancer during hormone replacement therapy. *Horm Mol Biol Clin Invest* 2011;3:329–339.
- 16 Hutchison GB, Shapiro S. Lead time gained by diagnostic screening for breast cancer. *J Natl Cancer Inst* 1968;41:665–681.
- 17 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.
- 18 Santen RJ. Does menopausal hormone therapy initiate new breast cancers or promote the growth of existing ones? *Women’s Health (Lond Engl)* 2008;4:207–210.

- 19 Santen RJ, Allred DC, Ardoin SP, *et al.* Postmenopausal hormone therapy: an Endocrine Society scientific statement. *J Clin Endocrinol Metab* 2010;**95**(Suppl. 1):S7–S66.
- 20 Skegg DC. Potential for bias in case-control studies of oral contraceptives and breast cancer. *Am J Epidemiol* 1988;**127**:205–212.
- 21 Morabia A, Szklo M, Stewart W, *et al.* Consistent lack of association between breast cancer and oral contraceptives using either hospital or neighborhood controls. *Prev Med* 1993;**22**:178–186.
- 22 La Vecchia C, Negri E, Franceschi S, *et al.* Hormone replacement treatment and breast cancer risk: a cooperative Italian study. *Br J Cancer* 1995;**72**:244–248.
- 23 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;doi:10.1186/bcr418. <http://breast-cancer-research.com/content/4/1/R1> [accessed 14 July 2010].
- 24 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.
- 25 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.
- 26 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.
- 27 Pukkala E, Weiderpass E. Time trends in socio-economic differences in incidence rates of cancers of the breast and female genital organs (Finland, 1971–1995). *Int J Cancer* 1999;**81**:56–61.
- 28 Danø H, Andersen O, Ewertz M, *et al.* Socioeconomic status and breast cancer in Denmark. *Int J Epidemiol* 2003;**32**:218–224.
- 29 Brown SB, Hole DJ, Cooke TG. Breast cancer incidence trends in deprived and affluent Scottish women. *Breast Cancer Res Treat* 2007;**103**:233–238.
- 30 Brett KM, Madans JH. Use of postmenopausal hormone replacement therapy: estimates from a nationally representative cohort study. *Am J Epidemiol* 1997;**145**:536–545.
- 31 Newcomb PA, Longnecker MP, Storer BE, *et al.* Long-term hormone replacement therapy and risk of breast cancer in postmenopausal women. *Am J Epidemiol* 1995;**142**:788–795.
- 32 Shapiro S. Bias in the evaluation of low-magnitude associations: an empirical perspective. *Am J Epidemiol* 2000;**151**:939–945.
- 33 Maclure M, Greenland S. Tests for trend and dose response: misinterpretations and alternatives. *Am J Epidemiol* 1992;**135**:96–104.
- 34 Breslow NE, Day NE. *Statistical Methods in Cancer Research. Vol II – The Design and Analysis of Cohort Studies* (IARC Scientific Publication No. 82). Lyon, France: International Agency for Research on Cancer, 1987.
- 35 Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 1987;**9**:1–30.
- 36 Anderson GL, Limacher M, Assaf AR, *et al.*; 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.
- 37 Mueck AO, Seeger H, Shapiro S. Risk of breast cancer during hormone replacement therapy. *Horm Biol Clin Invest* 2010;**3**:329–339.
- 38 Greendale GA, Reboussin BA, Slone S, *et al.* Postmenopausal hormone therapy and change in mammographic density. *J Natl Cancer Inst* 2003;**95**:30–37.
- 39 Speroff L. The meaning of mammographic breast density in users of postmenopausal hormone therapy. *Maturitas* 2002;**41**:171–175.
- 40 North American Menopause Society. Role of progestogen in hormone therapy for postmenopausal women: position statement of the North American Menopause Society. *Menopause* 2003;**10**:113–132.