

Commentary: The hormone replacement–coronary heart disease conundrum: is this the death of observational epidemiology?

Debbie A Lawlor, George Davey Smith and Shah Ebrahim

Under its definition for the word ‘hindsight’ the *Oxford English Dictionary* includes the following statement ‘hindsight is always better than foresight’ (<http://dictionary.oed.com/>), and the slogan of a private survey and evaluation company, ingeniously called Hindsight, is ‘remember hindsight is always 20/20!’ (<http://www.hindsight.com/>). We have the benefit of the ‘hindsight’ from randomized controlled trials (RCT) when we comment on this meta-analysis of observational studies, but whether the conflicting results between the trial and observational evidence on the association between hormone replacement therapy (HRT) use and coronary heart disease (CHD) will lead to 20/20 vision remains to be seen.

The disparity between findings from observational studies and RCT of the effects of HRT on CHD,^{1–4} has created considerable debate among researchers, practitioners and postmenopausal women. The authors of the meta-analysis reprinted in this issue of the *International Journal of Epidemiology* concluded that the pooled estimate of effect from the best quality observational studies (internally controlled prospective and angiographic studies) inferred a relative reduction of 50% with ever use of HRT and stated that ‘overall, the bulk of the evidence strongly supports a protective effect of estrogens that is unlikely to be explained by confounding factors’.⁴ By contrast, recent randomized trials among both women with established CHD and healthy women have found HRT to be associated with slightly increased risk of CHD or null effects.^{1,2} For example, the large Women’s Health Initiative (WHI) randomized trial found that the hazards ratio for CHD associated with being allocated to combined HRT was 1.29 (95% CI: 1.02, 1.63), after 5.2 years of follow-up.¹

These marked differences between observational findings and trials are important for two reasons. First, and foremost, is the clinical impact. As another commentator on the same subject remarked:

‘Does HRT decrease or increase the risk of heart disease? At least every woman, every gynecologist and every primary care doctor want to know the “correct” answer.’⁵

Second, is the broader implication for observational epidemiology. Prior to the publication of the WHI it was suggested that well conducted observational studies produced similar estimates of treatment effects as RCT, and that the notion of a hierarchy of evidence with the RCT on top could not be supported.^{6,7}

The differing results between observational studies and RCT in the association between HRT and CHD throw this idea into question and may signify the death of observational epidemiology.⁸ It is important, therefore, to determine why the results from the trials and observational studies are so different.

A number of explanations have been suggested for these disparities. Whilst some have suggested that the results of the trials were biased because of contamination, and in the case of the WHI, early termination of the arm assessing the effect of combined HRT, the consistency across a number of trials of a null effect make these explanations unlikely. More plausible explanations are that women who participated in the trials were importantly different from those who participated in the observational studies, or that the observational study results were confounded.

Changing goalposts

Once a hypothesis is shown to be shaky, the protagonists have several options: either dismiss the new evidence as inadequate, re-adjust the focus of the hypothesis, re-adjust their original data in order to fit with the new evidence or exceptionally, drop the hypothesis. Soon after publication of the RCT demonstrating no CHD benefits from taking HRT, re-analyses of the observational epidemiological data began to appear in attempts to demonstrate that the observational studies were not flawed but showed, essentially the same findings as the trials.

In one of their original publications, prior to the trial evidence, investigators from the Nurses’ Health Study showed that the protective effect of HRT was stronger among women who were at highest risk.⁹ In June 1990, the FDA Advisory Committee were considering a request by drug companies to approve a label change that would permit prevention of heart disease to be included as an indication for hormone replacement use. Elizabeth Barrett-Connor told the committee that the label change should not be agreed without trial evidence. However, Meir Stampfer told the committee ‘I believe that the data are quite substantial in showing a protective effect of estrogens for heart disease, and I believe that it is a cause and effect relationship’. The label change was approved (http://www.curedisease.com/internal_medicine.pdf, page 15–16; last accessed 19 May 2004).

Following publication of the Heart and Estrogen/progestin Replacement Study (HERS) (a secondary prevention trial), in 1998,² showing a 52% increased risk of CHD in the first year of use, the Nurse’s Health Study investigators re-analysed their data examining effects of short-term use of HRT among 2489 women (drawn from the total sample of 121 700) with a prior history of

Department of Social Medicine, University of Bristol, UK. E-mail: d.a.lawlor@bristol.ac.uk

cardiovascular disease and found a higher rate of recurrent events, the opposite of what they had claimed earlier, but now fitting with the trial data.¹⁰ Co-incident with publication of the WHI¹ a re-analysis of observational data demonstrated an almost identical increased relative risk (1.28) for a total cardiovascular incidence endpoint (CHD and stroke) as was found in the trial.¹¹

These shifting goalposts, however, fail to explain why earlier claims from observational studies were so strongly supportive of a protective effect of HRT.

Were trial participants importantly different from women studied in observational studies?

Women in the WHI trial were older than the typical age at which women take HRT and were more obese than the women who have been included in the observational studies.¹ These women may be more likely to have established atherosclerosis than younger and leaner women and therefore may be more prone to the pro-thrombotic effects of HRT.⁵ HRT may be protective for the general population of women but detrimental in sub-groups at higher risk of atherosclerosis. The results from the Framingham data, commented on in detail by Stampfer and Colditz,⁴ provide some support for a differential effect by age, with increased risk in older age groups. But Stampfer and Colditz also point out that other studies found the opposite age effect and yet other studies found no differential effect by age. Further, there was no evidence of interactions of treatment assignment with age, prior hormone use, or body mass index, for any cardiovascular outcomes in the WHI.^{1,12}

One of the commonest reasons for women seeking HRT at the time that most observational studies included in this review were conducted was that they had unpleasant menopausal symptoms; in the UK this continues to be one of the commonest reasons for use of HRT.^{13,14} A potentially important difference between trial participants and women included in observational studies is that the latter will have included many women with symptomatic menopause who want treatment whereas the former are, by definition, women who are prepared to take a 50:50 chance of being allocated to HRT. It is plausible therefore that if menopausal symptoms are a good indicator of relative or absolute oestrogen deficiency, and that this is associated with increased CHD risk, HRT may be protective against CHD in such women but not in those who are asymptomatic. However, menopausal symptoms are affected by a wide range of social, psychological and cultural factors and are not necessarily a good reflection of hormonal status. Further, the role of oestrogen as an important determinant of CHD is unclear.^{15,16}

Were observational studies confounded?

One of the interesting things about the Stampfer and Colditz paper⁴ is their assertion that the protective effect of oestrogen '... is unlikely to be explained by confounding factors'. This is based on their comparison of behavioural and physiological risk factors for CHD among HRT users and non-users, with no mention of possible confounding by socioeconomic position. Despite the fact that use of HRT is strongly socially patterned,¹⁷ and that socioeconomic position is associated with CHD,¹⁸ in many observational studies adjustment for adult socioeconomic position has failed to have a marked impact on the HRT-CHD

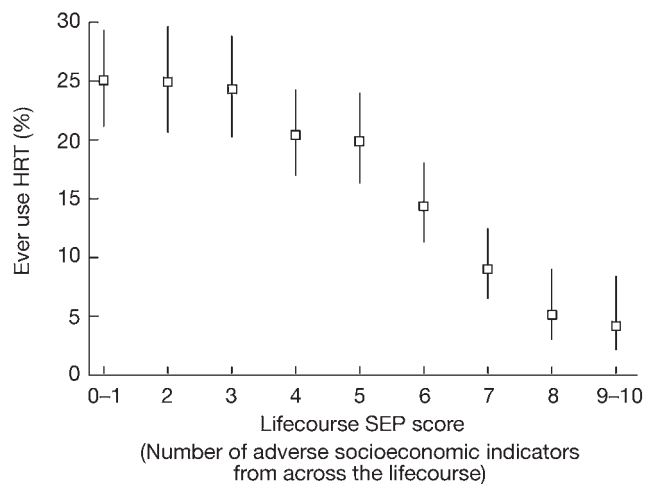


Figure Fully adjusted (for adult behavioural and physiological CHD risk factors*) prevalence (95% confidence interval) of ever use of HRT by cumulative score of life course socioeconomic position

* Fully adjusted association: systolic blood pressure, high density lipoprotein cholesterol, triglyceride levels, insulin resistance-diabetes, body mass index, waist to hip ratio, age at menopause, hysterectomy/oophorectomy, physical activity, smoking

association.^{19,20} We have recently shown that indicators of childhood socioeconomic position are associated with HRT use, independently of adult socioeconomic position, behavioural and physiological risk factors.²¹ Further, we found a cumulative effect of socioeconomic position from across the life course (Figure), indicating that not only does life course socioeconomic position need to be accounted for in observational studies of HRT-CHD, but that a single measure of socioeconomic position is unlikely to be adequate. Finally, we found that whilst use of HRT was associated with reduced odds of CHD in simple age-adjusted analyses, and remained protective with additional adjustment for four measures of adult socioeconomic position and behavioural and physiological risk factors, when we adjusted for all measures of socioeconomic position across the life course there was a slightly higher risk of CHD associated with HRT use, consistent with RCT evidence.²¹ We concluded that the protective effect of HRT found in previous observational studies was likely to be influenced by residual confounding. Inadequate adjustment for socioeconomic position from across the life course would be one source of this residual confounding.^{20,21} Other factors such as the tendency for doctors to select those women who are already at low risk of CHD for HRT will also be important.²²

Why were the observational and RCT results consistent for other outcomes?

One issue which remains intriguing in this debate is that the conflict between observational and RCT evidence is specific for CHD. For other health outcomes such as breast cancer, colon cancer, hip fracture, and stroke, results from observational studies and trials were similar, suggesting that if confounding explains the apparent protective effect of HRT against CHD, the associations with these other outcomes were not similarly confounded. Poor socioeconomic circumstances are not strongly associated with

breast and bowel cancer and therefore life course socioeconomic position, and risk factors associated with socioeconomic position, are probably not important confounders for these associations.^{23–25} Fewer studies have assessed the association of socioeconomic position with hip fracture and results are inconsistent, but one large community-based case-control study suggests that measures of adult adverse socioeconomic position are associated with increased risk of hip fracture.²⁶ If childhood socioeconomic position is not independently (of adult socioeconomic position and behavioural risk factors for hip fracture such as smoking) associated with hip fracture then adjustment for adult social class and behavioural risk factors may provide adequate adjustment for the HRT–hip fracture association.

Although results have been inconsistent, on the whole observational studies have not found an association between HRT with stroke.³ Clearly, if studies are not finding a protective effect with respect to stroke then one cannot begin to argue about the role of residual confounding in explaining this (non-existent) association. However, the fact that observational studies tended to find no consistent effect of HRT on stroke whilst in the same studies a protective effect against CHD is found requires further consideration, given the similarities between these two conditions with respect to risk factor profiles.^{3,26} Interestingly, in their discussion of the paper by Thompson *et al.*, which used a combined endpoint of myocardial infarction and stroke, Stampfer and Colditz hint at finding this combination problematic and indeed decrease the weight of this study in their pooled estimate because of the inclusion of strokes in the overall outcome. However, they make no comment about why they felt including strokes with CHD might underestimate the overall effect of HRT on their main outcome come of interest—CHD.

An important difficulty with respect to epidemiological studies of stroke is the ability to distinguish between stroke sub-types,²⁷ and any differential associations between HRT and stroke sub-types may obscure the real picture. Indeed, in the WHI trial, exactly this differential pattern of increased risk of ischaemic stroke (i.e. a similar association to that found for CHD in this trial, as one would expect from their similar pathophysiology) and decreased risk of haemorrhagic stroke was found,²⁸ suggesting that this is a plausible explanation. Attempts to assess the association of HRT with stroke sub-types in observational studies have yielded inconsistent results between studies for both sub-types.²⁷ Heterogeneity between studies may in part reflect the differing extent of misclassification of stroke sub-type in these studies since routine death certificate classifications tend to be inaccurate and clinical diagnoses will only have reasonable accuracy where there has been widespread use of scans.²⁷ In the WHI trial strokes were classified as ischaemic or haemorrhagic based on brain imaging.

Should we call it a day for observational epidemiology?

Well-conducted RCT have the clear advantage over observational studies in that they control for both known and unknown or unmeasured confounding factors, such as life course socioeconomic position and doctor selection practices. However, they are not always feasible, and because of the expense and ethical concerns of randomized trials, it is important that observational studies are used to effectively direct investigators to the interventions most appropriately assessed by trials. Abandoning observational studies

for RCT would not therefore be a panacea. A more careful approach to the design and analysis of observational epidemiological studies should ensure that they remain a useful methodology for generating and testing hypotheses that ultimately may improve the health of the public. Future observational studies, in this and other areas, should aim to collect (even retrospectively) information on socioeconomic circumstances from across the life course in order to be able to adjust as fully as possible for potential confounding factors. Sensitivity analyses to assess the possibility of residual confounding should also become routine practice.^{30,31} In addition, specificity of association should be considered.^{30,32} As long ago as 1986 Diana Petitti and colleagues showed that HRT was apparently equally protective against accidental and violent deaths in an observational study as it was against cardiovascular disease deaths.³³ They pointed out that given the lack of any biologically plausible link between HRT and these external causes of death both associations should be suspected of suffering from residual confounding.³³ We discuss approaches to strengthening inferences from observational studies in more detail elsewhere.^{34–36}

References

- 1 Writing Committee for the Women's Health Initiative randomized controlled trial. 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.
- 2 Hulley S, Grady D, Bush T *et al.* Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;**280**:605–13.
- 3 Grady D, Rubin SM, Petitti DB *et al.* Hormone therapy to prevent disease and prolong life in postmenopausal women [see comments]. *Ann Intern Med* 1992;**117**:1016–37.
- 4 Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991;**20**:47–63. (Reprinted *Int J Epidemiol* 2004;**33**:445–53)
- 5 Michels KB. Hormone replacement therapy in epidemiologic studies and randomized clinical trials— are we checkmate? *Epidemiology* 2003;**14**:3–5.
- 6 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *New Engl J Med* 2000;**342**:1878–86.
- 7 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New Engl J Med* 2000;**342**:1887–92.
- 8 Davey Smith G, Ebrahim S. Epidemiology—is it time to call it a day? *Int J Epidemiol* 2001;**30**:1–11.
- 9 Grodstein F, Stampfer MJ, Colditz GA *et al.* Postmenopausal hormone therapy and mortality. *New Engl J Med* 1997;**336**:1769–75.
- 10 Grodstein F, Manson JE, Stampfer MJ. Postmenopausal hormone use and secondary prevention of coronary events in the nurses' health study. a prospective, observational study. *Ann Intern Med* 2001;**135**:1–8.
- 11 Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 2002;**137**:273–84.
- 12 The WHI Steering Committee and Writing Group. Risks of postmenopausal hormone replacement therapy. *JAMA* 2002;**288**:2823–24.
- 13 Griffiths F, Convery B. Women's use of hormone replacement therapy for relief of menopausal symptoms, for prevention of osteoporosis, and after hysterectomy. *Br J Gen Pract* 1995;**45**:355–58.
- 14 Griffiths F, Jones K. The use of hormone replacement therapy; results of a community survey. *Fam Pract* 1995;**12**:163–65.

- ¹⁵ Barrett-Connor E. Sex differences in coronary heart disease. Why are women so superior? The 1995 Ancel Keys Lecture. *Circulation* 1997; **95**:252–64.
- ¹⁶ Lawlor DA, Ebrahim S, Davey Smith G. Sex matters: secular and geographic trends in sex differences in ischaemic heart disease mortality. *BMJ* 2001; **323**:541–45.
- ¹⁷ Barrett-Connor E. Hormone replacement therapy. *BMJ* 1998; **317**: 457–61.
- ¹⁸ Davey Smith G, Ben-Shlomo Y, Lynch J. Life course approaches to inequalities in coronary heart disease risk. In: Stansfeld S, Marmot M (eds). *Stress and the Heart*. London: BMJ books, 2002, pp. 20–49.
- ¹⁹ Grodstein F, Manson JE. Relationship between hormone replacement therapy, socioeconomic status, and coronary heart disease. *JAMA* 2003; **289**:44–45.
- ²⁰ Krieger N. Postmenopausal hormone therapy. *New Engl J Med* 2003; **348**:2363–64.
- ²¹ Lawlor DA, Davey Smith G, Ebrahim S. The association of socioeconomic position from across the life course with HRT use: an explanation for the discrepancy between observational and RCT evidence on HRT and CHD? *Am J Public Health* 2004 (In press)
- ²² Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004: (In press)
- ²³ Barbone F, Filiberti R, Franceschi S *et al.* Socioeconomic status, migration and the risk of breast cancer in Italy. *Int J Epidemiol* 1996; **25**:479–87.
- ²⁴ Claussen B, Davey Smith G, Thelle D. Impact of childhood and adulthood socioeconomic position on cause specific mortality: the Oslo Mortality Study. *J Epidemiol Community Health* 2003; **57**:40–45.
- ²⁵ Dedman DJ, Gunnell D, Davey Smith G, Frankel S. Childhood housing conditions and later mortality in the Boyd Orr cohort. *J Epidemiol Community Health* 2001; **55**:10–15.
- ²⁶ Farahmand BY, Persson PG, Michaelsson K *et al.* Socioeconomic status, marital status and hip fracture risk: a population-based case-control study. *Osteoporosis Int* 2000; **11**:803–08.
- ²⁷ Lawlor DA, Davey Smith G, Leon DA, Sterne JA, Ebrahim S. Secular trends in mortality by stroke subtype in the 20th century: a retrospective analysis. *Lancet* 2002; **360**:1818–23.
- ²⁸ Wassertheil-Smoller S, Hendrix SL, Limacher M *et al.* Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. *JAMA* 2003; **289**:2673–84.
- ²⁹ Paganini-Hill A. Hormone replacement therapy and stroke: risk, protection or no effect? *Maturitas* 2001; **38**:243–61.
- ³⁰ Davey Smith G, Ebrahim S. Data dredging, bias, or confounding. *BMJ* 2002; **325**:1437–38.
- ³¹ Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol* 1996; **25**:1107–16.
- ³² Weiss NS. Can the 'specificity' of an association be rehabilitated as a basis for supporting a causal hypothesis? *Epidemiology* 2002; **13**:6–8.
- ³³ Petitti DB, Perlman JA, Sidney S. Postmenopausal estrogen use and heart disease. *New Engl J Med* 1986; **315**:131–32.
- ³⁴ Davey Smith G, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* 2003; **32**:1–22.
- ³⁵ Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomized trial evidence? *Lancet* 2004 (In press).
- ³⁶ Davey Smith G, Ebrahim S. Mendelian randomization: prospects, potentials, and limitations. *Int J Epidemiol* 2004; **33**:30–42.