IJE vol.33 no.3 © International Epidemiological Association 2004; all rights reserved.
Advance Access publication 27 May 2004

International Journal of Epidemiology 2004;**33**:456–457
DOI: 10.1093/ije/dyh121

Commentary: The HRT story: vindication of old epidemiological theory

Jan P Vandenbroucke

In 1995, a long time before any result of any randomized controlled trial would be known, I wrote that about half of the alleged 35–45% reduction of myocardial infarction with hormone replacement therapy (HRT) in observational studies would not be real.[1] Still, I should confess that the total reversal of the effect—a slight increase of myocardial infarction in the first years of use—took me also by surprise.

Why was I so certain that a large part of the seeming protection from myocardial infarction would not be true—even in the face of an overwhelming number of epidemiological studies?[2] I guess, in essence, because during my training in epidemiology at the Harvard School of Public Health in 1978/79, I was much influenced by Miettinen's argument about the difference between studying the 'intended and unintended' effects of treatments.[3] The view of the believers of the HRT protection was summarized in a lecture by a leading gynaecologist that I attended in the 1990s: he showed beautiful coloured slides about the protective effect of HRT, telling his audience that all possibility of bias and confounding was ruled out by careful statistical adjustments in well-designed studies. He also had one slide in black and white, showing the possibility of an increase of venous thrombosis with HRT, but added that this would most certainly be due to bias. Actually, the theory on the difference between studying intended and unintended effects predicts exactly the opposite. Adverse effects are unintended, therefore generally unexpected and unpredictable—at least in the usual clinical consultation. Thus, there is little likelihood of bias and confounding for adverse effects of HRT, such as venous thrombosis. One can even increase the unpredictability of adverse effects by limiting a study to people without any risk factor for the adverse effect, as was described already in 1978 by Jick and Vessey[4]—for example, by limiting studies on venous thrombosis and different oral contraceptives to otherwise healthy young women without any risk factor for thrombosis.

In contrast, signs were all over the wall that prescription of HRT was highly selective against risk factors of myocardial infarction. Physicians who prescribed HRT in the 1970s and 1980s were risk avoiding about coronary risk factors, as was described by many authors (see ref. 1 for overview of critical opinions from the time before the RCT on HRT). This risk averse prescribing behaviour has persisted until the year 2000.[5] Coronary disease has many clinically recognisable risk factors, such as hypertension, hypercholesterolaemia, and diabetes. If prescription is with an eye on such risk factors, and in particular if many complex and interdepedent risk factors come into play, the confounding bias introduced by this prescription process cannot be grasped any more by simply adjusting for a few variables in a logistic regression.[3] A similar argument was used by Rubin in defence of randomization: he explained that a clinician's treatment decisions are just too complex to be modelled, and that therefore randomization is always preferable.[6]

In later works, Rubin and Rosenbaum coined the concept of 'strong ignorability' of the allocation mechanism: in specific subgroups, or in specific circumstances, the allocation mechanism might be 'ignored': it can be left out of a model when it plays no role in determining the outcome.[7] In general, adverse effects will meet that condition because they are most of the time unpredictable, or not taken into account, when prescribing.[3,4] Thus, for adverse effects observational studies will be as good as randomized trials. Indeed, for the adverse effects of HRT, such as venous thrombosis and breast cancer, observational and randomized studies concurred almost 'on the dot', at least in terms of relative risk.[8,9]

The main lesson of this episode to me is a vindication of the old theories on 'intended and unintended' effects and the investigation of adverse drug reactions.[3,4] To study the effect of a treatment that is prescribed on the basis of complex rules of indication and contraindications, observational studies run a high risk of getting the answer wrong. If there is no form of sufficiently 'haphazard allocation' of the exposure at baseline, it will not help to throw a few variables into a model for adjustment, and then hope that the causal risk has been quantified.[10] In observational research, such haphazard allocations—i.e. allocations that are not intrinsically tied to prognosis—are most likely for unexpected and undesired effects. 'Haphazard allocation' is not the same as physical randomization. However, as stated by Rosenbaum:

> *'Haphazard is not random ...*
>
> *Still, haphazard or ostensibly irrelevant assignments are to be preferred to assignments which are known to be biased in ways that cannot be measured and removed analytically.'*[11]

I have generalized these arguments in a separate paper proposing a 'three-pronged restriction' to give observational studies the best chances to be as credible as randomized trials.[12] In thinking about the evaluation of quality of research, the distinction between 'intended and unintended' effects—rather than a mere hierarchy with the randomized trial on top—might also be the way forward.[13]

References

[1] Vandenbroucke JP. How much of the cardioprotective effect of postmenopausal estrogens is real? *Epidemiology* 1995;**6:**207–08.

Department of Clinical Epidemioloy, Leiden University Medical Centre, Bldg 1 C9-P, 2300 RC Leiden, The Netherlands. E-mail: J.P.Vandenbroucke@lumc.nl

[2] 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.)

[3] Miettinen OS. The need for randomization in the study of intended effects. *Stat Med* 1983;**2:**267–71.

[4] Jick H, Vessey MP. Case-control studies of drug induced illness. *Am J Epidemiol* 1978;**107:**1–7.

[5] The Million Women Study. Patterns of use of hormone replacement therapy in one million women in Britain, 1996–2000. *Br J Obstet Gynaecol* 2002;**109:**1319–30.

[6] Rubin DB. Bayesian inference for causal effects: the role of randomization. *Ann Stat* 1978;**6:**34–58.

[7] Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983;**70:**41–55.

[8] Beral V, Banks E, Reeves G. Evidence from randomised trials on the long-term effects of hormone replacement therapy. *Lancet* 2002;**360:**942–44.

[9] Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med* 2003; **348:**645–50.

[10] Goldthorpe JH. Causation, statistics and sociology. *Eur Sociol Rev* 2001;**17:**1–20.

[11] Rosenbaum PR. *Observational studies 2nd Edn*. New York: Springer, 2002, p. 345.

[12] Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004 (In press).

[13] Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004:**328:**39–41.

IJE vol.33 no.3 © International Epidemiological Association 2004; all rights reserved.
Advance Access publication 27 May 2004

International Journal of Epidemiology 2004;**33**:457–459
DOI: 10.1093/ije/dyh123

Commentary: Observation versus intervention—what's different?

Elizabeth Barrett-Connor

Women have coronary heart disease (CHD) later than men in every country.[1] This universal sex difference has been attributed to a cardioprotective effect of premenopausal oestrogen levels. Many other lines of evidence including laboratory studies support this thesis. Belief in the preventive power of endogenous oestrogen was transformed to action mainly by meta-analyses of epidemiological studies of 'oestrogen replacement therapy' and heart disease.

Meta-analyses pool data from separate studies weighted for sample size, thereby increasing the total number of events, increasing power, and potentially providing statistically significant results (and narrow CI) not observed in individual studies. One of the first meta-analyses of hormone therapy and CHD is the 1991 paper by Stampfer and Colditz[2] reproduced here. The authors reported an overall relative risk (RR) of 0.56 (95% CI: 0.50, 0.61) based on pooled data from all 31 publications reviewed, and an RR of 0.50 (95% CI: 0.43, 0.56) based on pooled data from 13 prospective cohort studies and 3 cross-sectional angiographic studies. They concluded that these results were 'unlikely to be explained by confounding factors.' Other than the authors, few noted that the results for individual studies were inconsistent (test for heterogeneity $P < 0.001$), and that more than half of the associations were not statistically significant. The Nurses Health Study,[3] being the largest study, carried considerable weight in this and subsequent meta-analyses.

The promotion of hormone therapy for the prevention of heart disease *for nearly all postmenopausal women* came from a meta-analysis suggesting that the expected CHD risk reduction exceeded any expected increased risk for cancer.[4] Within 10 years many leading US medical organizations had endorsed the concept that all postmenopausal women should be offered hormone therapy to prevent heart disease.

By 2002, however, publications from two large, randomized, placebo-controlled clinical trials[5,6] and several smaller trials[7] failed to show a reduced risk of CHD in women assigned to oral oestrogen therapy, and some suggested an early excess risk.

Can we explain such diametrically opposed results? This could happen if trial participants differ in important ways from the populations in observational studies, or if the trial treatment differs from the regimen(s) found to be useful in observational studies.

Participants

Age

One popular thesis is that the women in the trials were too old. Many of the epidemiological studies reviewed by Stampfer and

Division of Epidemiology, Department of Family and Preventive Medicine, University of California, San Diego, La Jolla, CA 92093–0607, USA. E-mail: ebarrettconnor@ucsd.edu