COMMENTARY:

DATA ANALYSIS METHODS AND THE RELIABILITY

OF ANALYTIC EPIDEMIOLOGIC RESEARCH

Ross L. Prentice

Fred Hutchinson Cancer Research Center

1100 Fairview Avenue North, M3-A410, POB 19024, Seattle, WA 98109-1024

Ph: 206-667-4264; Fax: 206-667-4142; E-mail: rprentic@fhcrc.org

Running Title: Analytic epidemiology reliability

ABSTRACT

Publications that compare randomized controlled trial and cohort study results on the effects of postmenopausal estrogen plus progestin therapy are reviewed. The two types of studies agree in identifying an early elevation in coronary heart disease risk, and a later developing elevation in breast cancer risk. Effects among women who begin hormone therapy within a few years following the menopause may be comparatively more favorable for coronary heart disease and less favorable for breast cancer. These analyses illustrate the potential of modern data analysis methods to enhance the reliability and interpretation of epidemiologic data.

1.   INTRODUCTION

There is a pressing need to find ways to assess, and enhance as necessary, the reliability of findings from observational studies (OSs). Available methods for controlling confounding, measurement error and other biases can be expected to provide adjustments in the desired direction, but objective means of assessing bias avoidance are generally lacking. Randomized controlled trials (CTs) include an objective assignment of a study treatment or intervention, and avoid confounding by pre-randomization factors. However, CTs tend to be expensive and typically cannot be conducted in a manner that powerfully addresses subset hypotheses, or treatment effects over a lengthy exposure period. Hence, the population science research agenda must rely heavily on observational studies for the development and initial testing of disease prevention hypotheses, with CTs typically conducted only for well-established hypotheses having strong public health potential.

Settings in which both CTs and OSs are available provide a particularly opportunity to examine consistency of results from the two types of studies, and to identify improvements in study design, conduct, or analysis that may help to explain any discrepancy in results. Such data exist for postmenopausal hormone therapy (HT) in relation to several important clinical outcomes, and few topics having generated more interest and controversy in recent years, in part because CT and OS findings appeared to be strongly discrepant.

2.   BENEFITS AND RISKS OF POSTMENOPAUSAL HORMONE THERAPY

A substantial body of cohort and case-control studies suggested that postmenopausal hormone therapy would reduce coronary heart disease (CHD) risk perhaps by about 40-50%, with little indication for a difference in effects between estrogen-alone or estrogen plus progestin.[1, 2] A later developing extensive observational literature also suggested elevations in breast cancer risk, by about 30% for estrogen, and 50-100% for estrogen plus progestin.[3, 4] Reports that were available by the early 1990s informed the design of the Women's Health Initiative (WHI) CTs of 0.625 mg daily conjugated equine estrogen (CEE) among 10,739 women who were post-hysterectomy, and of this same estrogen regimen plus 2.5 mg/day medroxyprogesterone acetate (CEE/MPA) among 16,608 women with uterus. CHD was the designated primary outcome with breast cancer as the primary 'safety' outcome in both trials. A recruitment age range of 50-79 was specified to examine whether health benefits and risk would apply broadly to postmenopausal women. At the time these trials were initiated the CEE and CEE/MPA regimens under study were used by about 8 and 6 million women respectively in the United States.

With this background it came as quite a surprise when the CEE/MPA trial was stopped prematurely in 2002 when it was judged that health risks exceeded benefits over its 5.6-year average follow-up period. The health risks included elevations in breast cancer, stroke, venous thromboembolism (VT), and CHD, which were only partially offset by reductions in fracture and colorectal cancer.[5] Though breast cancer was a trigger for early stopping, the hazard ratio (HR) estimate was a moderate 1.24 with 95% confidence interval (CI) from 1.01 to 1.54.[6] More surprising was the HR of 1.24 (95% CI from 1.00 to 1.54) for CHD, with an HR of 1.81 (95% CI from 1.09 to 3.01) during the first year of CEE/MPA use.[7]

WHI investigators undertook joint analyses of data from this CT with that from a corresponding subset of the WHI observational Study (OS), which was comprised of women recruited from the same population as the CT, with much commonality in eligibility criteria, baseline data collection, and outcome ascertainment. HRs from the OS alone were considerably lower than for the RCT and similar to those from other cohort studies following confounding control, for each of CHD, stroke, and VT.[8] However, CHD HRs were found to agree closely following control for time from hormone therapy initiation (duration of use among adherent women). The same analytic techniques, however, did not appear to fully explain the lower stroke HR from the OS compared to the CT.

The WHI CEE trial also ended early in 2004, based on a stroke elevation of similar magnitude to that for CEE/MPA (HR of about 1.3), and a limited power to establish a CHD effect prior to the trial's planned termination.[9] The HR (95% CI) for CHD was 0.95 (0.79, 1.15), while that for

breast cancer was a rather surprising 0.80 (0.62, 1.04) over the trial's follow-up period that averaged 7.1 years.[10, 11]

Comparative analysis of WHI CT and OS analyses for CHD, stroke, and VT yielded almost identical results for CEE as for CEE/MPA. HRs from the two sources agreed closely for CHD and VT, and not so closely for stroke, after confounding control upon allowing the HR to depend on time from CEE initiation.[12] In fact, the ratio of CT to OS HRs was about 0.9 for CHD and VT, and about 0.7 for stroke for both CEE and CEE/MPA, presumably suggesting some residual bias for stroke.

Recently analyses of this type have also been presented for breast cancer.[13, 14] HRs from the OS were somewhat higher than those from the RCT for both CEE/MPA and CEE even after control confounding and accommodating time since HT initiation. These HRs, however, were found to be higher among women who first initiated HT within a few years following the menopause compared to women having larger gap times, and HRs agreed closely between the two data sources after allowing effect modification by this gap time variable. Among women having gap times of less than 5 years the breast cancer HR increased to about 2.0 following two or more years of CEE/MPA, while that for CEE was about 1.0.

The types of modeling and comparative analyses just described achieve some robustness by virtue of similar findings between CEE/MPA and CEE, but it is also of great interest to compare WHI CT results with results from other observational studies, including the Nurses Health Study

(NHS) which played an important role in the generation and initial testing of hypotheses related to HT effects.

3.   CEE/MPA AND CHD IN THE NURSES HEALTH STUDY

In this issue Hernán et al[15] provide a reanalysis of the association between CEE/MPA and CHD in the NHS. These authors are to be congratulated on a careful matching of the NHS subset used (34,575 women) to the set of women enrolled in the WHI CEE/MPA trial, and for a series of analyses that elucidate the impact of various analytic definitions and estimation procedures on the resulting HRs. Also, the participating NHS coauthors are to be congratulated for allowing their data to be subjected to the novel analytic approaches employed. Compared to the WHI OS, the NHS has the distinct advantage that much of the CEE/MPA use was initiated following cohort enrollment, potentially allowing precise assessment of benefits and risks during the early months following HT initiation. Previous analyses of NHS data evidently relied on a biennial snapshot of current HT user status. This was evidently an important analytic limitation for estimation of an early HR increase that substantially dissipated within a year or two following CEE/MPA initiation. For example, women who initiate CEE/MPA would be classified based on this snapshot as non-users until their biennial data collection time, and permanently as non-users if they stop usage prior to such collection.[16] In the present analysis the authors recover 'estimates' of the date of HT initiation to the extent possible through a fuller use of available data, presumably substantially mitigating this source of bias. They also attempt to emulate a CT by defining a multivariate response for each woman by classifying her as an initiator or non-initiator in each two-year follow-up interval and estimating an initiator vs. non-initiator HR from the follow-up of each such 'stratum' with appropriate provision for dependencies that arise from

individual woman contributing to several (up to 8) HR estimates. There was little evidence that

such HR estimates differed among strata, and the resulting common HR estimates agreed closely

with corresponding estimates from the WHI CT with HR estimate (95% CI) of 1.42 (0.92, 2.20)

for the first two years of use, and 0.96 (0.78, 1.18) over the entire follow-up period. Note that

this type of CT emulation methodology was not needed for analysis of the WHI OS, since there

were few HT initiators following cohort enrollment.

Hernán and colleagues go on to describe a possible interaction (p=0.08) of HR with years from

menopause to CEE/MPA initiation. Among women having fewer than 10 years from menopause

to HT initiation the HR (95% CI) was 1.28 (0.62, 0.84) in the first two years of follow-up, and

0.81 (0.56, 1.17) thereafter. Such an interaction was not evident in the WHI trial, and would

benefit from study in other settings.

4. INTENTION-TO-TREAT AND ADHERENCE ADJUSTMENT

Hernán and colleagues include some rather harsh criticisms of intention-to-treat (ITT) analyses,

indicating that ITT estimates 'may be unsatisfactory when studying efficacy, and inappropriate

when studying the safety, of an active treatment compared to no treatment'. It seems worth

reiterating that of the various analyses discussed here only for the CT ITT comparisons can we

be sure that the treated and untreated groups are fully comparable at enrollment. Hence, if the

clinical outcomes are equally ascertained between the active and placebo groups, a causal

interpretation for the treatment and its sequelae is justified for any differences that emerge. By

comparison, what Hernán and colleagues refer to as an ITT analysis of the NHS data attempts to

argue toward a causal interpretation by virtue of careful confounding control, and

accommodation of time of HT initiation, and time since HT initiation, and there is limited ability in the absence of corresponding CT data, to assess the success of these efforts.

However, there are important questions to answer beyond ITT comparisons, including estimating the magnitude of treatment effects among study subjects who adhere to the treatment regimen. Even the CT setting does not allow a HR function for adherent women to be estimated without making additional assumptions. Women who adhere to treatment or non-treatment status may have many biobehavioral differences from those who do not, and these characteristics may differ between treated and non-treated groups. A CT that is able to maintain an effective blinding of active versus placebo status may yield fairly comparable groups of adherent women.[5] Nevertheless, WHI investigators describe comparisons between women adherent to active and placebo pill taking as sensitivity analysis to alert the reader to possible non-comparability between these groups.

Some adherence-adjusted analyses in WHI have simply censored the follow-up of women soon after they become non-adherent. Including inverse censoring probability weighting (ICPW) as in Hernán et al, could presumably enhance these comparisons, by attempting to restore a contrast that is theoretically applicable to the entire randomized group. While this ICPW method is a useful step forward, it is worth noting that the justification for the adherence-adjusted HRs that emerge depends directly on the ability to adequately model the non-adherence process. Doing so is analogous to modeling to control confounding. One presumably needs detailed knowledge of, and accurate measurement and modeling of, the factors that may determine adherence to each treatment group in the study population. It would seem that the knowledge base for this type of

activity is still limited, arguing for a suitably circumspect interpretation of resulting HR estimates. HRs among adherent women tend to be more extreme in their departure from the null than do ITT analyses for both the cardiovascular disease and breast cancer outcomes for each of the data sources considered here.

5.   CONCLUDING COMMENT

Excellent progress has been made in recent decades on the development of data analytic methods for trials and observational studies emanating, in part, from the Cox[17] hazard ratio regression model and its multivariate extensions. The reanalysis of Hernán et al strongly suggest that the use of these methods can strengthen the analysis and interpretation of observational studies. Still, it seems evident that CTs are needed when preventive interventions are widely used, or when the public health implications are sufficiently large. In the special case of postmenopausal HT the state of knowledge of health benefits and risks is quite different following the WHI trials than had been assumed in advance, and it is interesting to question whether an early elevation in CHD risk, or a more sustained elevation in stroke and dementia risk,[18-20] would have been identified in the absence of CT data?

REFERENCES

1.   Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med.* 1991; **20**:47–63.

2.   Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med.* 1992; **117**:1016–1036.

3.   Beral V, Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet.* 2003; **362**:419–427.

4.   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. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 1997; **350**:1047–1059.

5.   Rossouw JE, Anderson GL, Prentice RL, et al; for the Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA.* 2002; **288**:321-333.

6.   Chlebowski RT, Hendrix SL, Langer RD, et al; for the WHI Investigators. 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.

7.   Manson JE, Hsia J, Johnson KC, et al; for the Women's Health Initiative Investigators. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med.* 2003; **349**:523-534.

8.   Prentice RL, Langer R, Stefanick ML, et al; for the Women's Health Initiative Investigators. Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observational studies and the Women's Health Initiative clinical trial. *Am J Epidemiol.* 2005; **162**:404-414.

9.   Anderson GL, Limacher M, Assaf AR, et al; for 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**:1701-1712.

10.   Hsia J, Langer RD, Manson JE, et al; for the Women's Health Initiative Investigators. Conjugated equine estrogens and coronary heart disease: the Women's Health Initiative. *Arch Intern Med.* 2006; **166**:357-365.

11.   Stefanick ML, Anderson GL, Margolis KL, et al; for the WHI Investigators. Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA.* 2006; **295**:1647-1657.

12.   Prentice RL, Langer RD, Stefanick ML, et al; for the Women's Health Initiative Investigators. Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *Am J Epidemiol*. 2006; **163**:589-599.

13.   Prentice RL, Chlebowski RT, Stefanick ML, et al. Estrogen plus progestin therapy and breast cancer in recently postmenopausal women. *Am J Epidemiol*. 2008; **167**:1207-1216 [Epub 2008 Mar 27].

14.   Prentice RL, Chlebowski RT, Stefanick ML, et al. Conjugated equine estrogens and breast cancer risk in the Women's Health Initiative clinical trial and observational study. *Am J Epidemiol.* 2008; **167**:1407-1415 [Epub 2008 Apr 29].

15.   Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. In press, *Epidemiology,* 2008.

16.   Prentice RL, Pettinger M, Anderson GL. Statistical issues arising in the Women's Health Initiative (with discussion). *Biometrics.* 2005; **61**:899-941.

17.   Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B.* 1972; **34**:187-220.

18.   Wassertheil-Smoller S, Hendrix SL, Limacher M, et al; for the WHI Investigators. Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative: a randomized trial. *JAMA.* 2003; **289**:2673-2684.

19.   Hendrix SL, Wassertheil-Smoller S, Johnson KC, et al; for the WHI Investigators. Effects of conjugated equine estrogen on stroke in the Women's Health Initiative. *Circulation.* 2006; **113**:2425-2434.

20.   Shumaker SA, Legault C, Rapp SR, et al; for the WHIMS Investigators. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: the Women's Health Initiative Memory Study: a randomized controlled trial. *JAMA.* 2003; **289**:2651-2662.