Data Analysis Methods and the Reliability of Analytic Epidemiologic Research

Ross L. Prentice

Abstract: Publications that compare randomized controlled trial and cohort study results on the effects of postmenopausal estrogen-plus-progestin therapy are reviewed. The 2 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 after 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.

(*Epidemiology* 2008;19: 785–788)

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

Settings in which both trials and observations studies are available provide a particular opportunity to examine consistency of results from the 2 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 in relation to several important clinical outcomes. Few topics have generated more interest and controversy in recent years, in part because findings from a clinical trial and observation studies seemed to be strongly discrepant.

Benefits and Risks of Postmenopausal Hormone Therapy

A substantial body of cohort and case-control studies has suggested that postmenopausal hormone therapy reduces coronary heart disease (CHD) risk by about 40% to 50%, with little indication for a difference in effects between estrogen-alone or estrogen-plus-progestin.[1,2] A subsequent and extensive observational literature has also suggested elevations in breast cancer risk, by about 30% for estrogen and 50% to 100% for estrogen-plus-progestin.[3,4] Reports available by the early 1990s informed the design of the Women's Health Initiative (WHI) clinical trial, which randomized 10,739 posthysterectomy women to 0.625 mg daily conjugated equine estrogen, and 16,608 women with intact uterus to this same estrogen regimen plus 2.5 mg/d medroxyprogesterone acetate. CHD was the designated primary outcome with breast cancer as the primary "safety" outcome in both trials. A recruitment age range of 50 to 79 years was specified to examine whether health benefits and risk would apply broadly to postmenopausal women. At the time these trials were initiated the estrogen and estrogen-plus-progesterone 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 estrogen-plus-progesterone trial was stopped prematurely in 2002. Health risks were judged to exceed benefits over its 5.6-year average follow-up period. The health risks included elevations in breast cancer, stroke, venous thromboembolism, and CHD, which were only partially offset by reductions in fractures and colorectal cancer.[5] Although breast cancer was a trigger for early stopping, the hazard ratio (HR) estimate was a moderate 1.24 with 95% confidence interval (CI) = 1.01–1.54.[6] More surprising was the HR of 1.24 (95% CI from 1.00–1.54) for CHD, with an HR of 1.81 (95% CI = 1.09–3.01) during the first year of combined-hormone use.[7]

WHI investigators undertook joint analyses of data from the clinical trial with data from a corresponding subset

of the WHI observational study, which was composed of women recruited from the same population as the trail, with much commonality in eligibility criteria, baseline data collection, and outcome ascertainment. HRs from the observational study alone were considerably lower than for the trial and similar to those from other cohort studies following confounding control, for each of CHD, stroke, and thromboembolism.[8] However, HRs for CHD agreed closely following control for time from hormone therapy initiation (duration of use among adherent women). The same analytic techniques did not seem to explain fully the lower risk of stroke in the observational study compared with the trial.

The WHI estrogen trial was also ended early (in 2004), based on an elevation in stroke risk similar to that for estrogen-plus-progesterone (HR ~1.3), and a limited power to establish a CHD effect before the trial's planned termination.[9] The HR (95% CI) for CHD was 0.95 (0.79–1.15), whereas 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 year.[10,11]

Comparative analyses of WHI trial and observational study data for CHD, stroke, and thromboembolism yielded almost identical results for estrogen as for estrogen-plus-progesterone. HRs from the 2 sources agreed closely for CHD and thromboembolism, and not so closely for stroke, after confounding control on allowing the HR to depend on time from estrogen initiation.[12] In fact, the ratio of HRs from the trial and the observational study was about 0.9 for CHD and thromboembolism, and about 0.7 for stroke for both hormone regimens, presumably suggesting some residual bias for stroke.

Comparatively analyses of this type were recently presented for breast cancer.[13,14] HRs from the observational study were somewhat higher than those from the trial for both hormone regimens, even after control confounding and accommodating time since initiation of hormone therapy. These HRs, however, were higher among women who first initiated hormones within a few years after menopause compared with women having larger gap times. The HRs agreed closely between the 2 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 2 or more years of estrogen-plus-progesterone, whereas that for estrogen alone was about 1.0.

The types of modeling and comparative analyses just described achieve some robustness by virtue of similar findings between the 2 hormone regimens, but it is also of great interest to compare the clinical trial 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 hormone treatment effects.

Estrogen-Plus-Progesterone Therapy and CHD in the NHS

In this issue Hernán et al[15] provide a reanalysis of the association between estrogen-plus-progesterone 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 trial of combined hormones, 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 these novel analytic approaches. Compared with the WHI observational study, the NHS has the distinct advantage that much of the hormone use was initiated after cohort enrollment, potentially allowing precise assessment of benefits and risks during the early months after hormone initiation. Previous analyses of NHS data relied on a biennial snapshot of current hormone user status. This was evidently an important analytic limitation for estimation of an early HR increase that substantially dissipated within a year or 2 after initiation of hormone therapy. For example, women who initiate hormones would be classified based on this snapshot as nonusers until their biennial data collection time, and permanently as nonusers if they stopped usage before such collection.[16] In the present analysis, the authors recover "estimates" of the date of hormone 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 clinical trial by defining a multivariate response for each woman. Each woman is classified as an initiator or noninitiator in each 2-year follow-up interval. An initiator versus noninitiator HR is then estimated from the follow-up of each such "stratum" with appropriate provision for dependencies that arise from an 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 trial with HR estimate (95% CI) of 1.42 (0.92–2.20) for the first 2 years of use, and 0.96 (0.78–1.18) over the entire follow-up period. Note that this type of methodology for emulating clinical trials was not needed for analysis of the WHI observational study because there were few hormone initiators after cohort enrollment.

Hernán et al go on to describe a possible interaction (P = 0.08) of HR with years from menopause to initiation of hormone therapy. Among women having fewer than 10 years from menopause to initiation of hormones, the HR (95% CI) was 1.28 (0.62–0.84) in the first 2 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.

Intention-to-Treat and Adherence Adjustment

Hernán et al 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 with no treatment." It seems worth reiterating that of the various analyses discussed here, it is only for the ITT comparisons in the clinical trial that we can be sure the treated and untreated groups were 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 et al 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 hormone initiation, and time since hormone initiation. There is limited ability in the absence of corresponding clinical trial data to assess the success of these efforts.

However, there are important questions to answer beyond ITT comparisons. One is the magnitude of treatment effects among study subjects who adhere to the treatment regimen. Even the clinical trial setting does not allow an estimate of risk for adherent women without making additional assumptions. Women who adhere to treatment or nontreatment status may have many biobehavioral differences from those who do not, and these characteristics may differ between treated and nontreated groups. A trial 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 noncomparability between these groups.

Some adherence-adjusted analyses in WHI have simply censored the follow-up of women soon after they become nonadherent. Including inverse censoring probability weighting as in Hernán et al, could presumably enhance these comparisons by restoring a contrast that is theoretically applicable to the entire randomized group. Although this method is a useful step forward, the justification for the adherence-adjusted HRs that emerge depends directly on the ability to model the nonadherence process. Doing so is analogous to modeling for control of confounding. The factors that determine adherence to each treatment group in the study population must be accurately measured and correctly modeled. It would seem that the knowledge base for this type of analysis 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.

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] HR regression model and its multivariate extensions. The reanalysis by Hernán et al strongly suggests that the use of these methods can strengthen the analysis and interpretation of observational studies. Still, it seems evident that clinical trials are needed when preventive interventions are widely used or when the public health implications are sufficiently large. In the special case of postmenopausal hormone therapy, the state of knowledge of health benefits and risks is quite different following the WHI trials than had been assumed in advance. 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 clinical trial 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. 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.
5. Rossouw JE, Anderson GL, Prentice RL, et al. 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. 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. 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. 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. 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. 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. 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. 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.
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.

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. *Epidemiology*. 2008;19:766–779.

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. 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. 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. 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.

A Call for Nominations:
The 2009 Rothman EPIDEMIOLOGY Prize

EPIDEMIOLOGY presents an annual award for the best paper published by the ournal during the previous year. The prize of $3,000 and a plaque goes to the author whose paper is selected by the Editors and the Editorial Board for its originality, importnace, clarity of thought, and excellence in writing.

With this issue, we close our 2008 volume. We invite our readers to nominate papers published during the past year. Please email your nominations to Allen Wilcox, Editor-in-Chief: EDITOR@EPIJOURNAL.ORG.

Nominations must be received no later than 31 December 2008. The winner will be announced in our September 2009 issue and at the annual meeting of the American College of Epidemiology.

This award is made possible by an endowment from Hoffman-LaRoche Ltd., managed by the American College of Epidemiology.