

The Sound and the Fury

Was It All Worth It?

Robert N. Hoover

Abstract: The initial report of coronary heart disease (CHD) results from the trial of menopausal hormone therapy within the Women's Health Initiative precipitated substantial surprise and concern in the epidemiology research community over the apparent differences between the trial results and those of observational studies. What followed was 6 years of discussion and debate, frequently acrimonious, along with intense methodologic and substantive research attempting to reconcile or explain the apparent differences. The results have been an impressive improvement in methods to contrast and combine studies of differing designs, dramatic illustrations of some central epidemiologic principles, insights into likely mechanisms of CHD, and increasing clarity of the public health message about menopausal hormone therapy.

(*Epidemiology* 2008;19: 780–782)

The early termination of the Women's Health Initiative (WHI) trial of combined estrogen/progestin menopausal hormone therapy precipitated a profound reaction by the media, among menopausal women, and in the clinical community.¹ The trial result, that cumulative risks outweighed benefits, overturned conventional wisdom and decades of clinical practice. Particularly disturbing was the increased risk of coronary heart disease (CHD) in the face of widely accepted results from epidemiologic and clinical investigations suggesting protective effects of hormone therapy. Profound angst ensued for many epidemiologists and biostatisticians, leading to years of attempts to reconcile or explain the apparent differences between the trial results and those of the observational studies.^{2–5} The intensity of this reaction was surprising, because epidemiologists spend years learning, and careers teaching, the litany of relative strengths and weaknesses of different study designs, and the distinct possibility of differing results because of differing designs. That said, the methodologic advances and enhanced understanding of the

effects of hormone therapy resulting from this level of concern have been major contributions to our discipline.

From the methodologic perspective, investigators from the WHI⁴ and Nurses' Health Study (NHS)⁵ have proposed innovative methods for analyzing data from clinical trials and observational cohort studies in a comparable manner. This not only allows assessment of similarities and differences in findings, but more importantly, these approaches provide ways to combine data coherently from different designs. Such pooled analyses not only increase power but also can leverage the distinct strengths of each design. The result is a whole that is truly greater than the sum of its parts.

The paper by Hernán et al⁵ in this issue of the journal confirms previous analyses and speculations that many of the apparent differences in overall CHD risk between the 2 studies lie not in classic confounding by other causal variables, but in differing risks by duration of follow-up, and in groups of women defined by duration of interval between menopause and initiation of hormone therapy, along with differing distributions in these variables between the 2 study populations. Because the strongest evidence for an adverse effect occurs in the years immediately following the initiation of hormone therapy, correcting for the misclassification bias in this time interval within previous NHS analyses (associated with counting exposure from the date of return of the first questionnaire after initiation of MHT rather than an estimated time of initiation itself) further narrows apparent differences in this interval. As might be expected when subgroups of 2 studies are compared, confidence intervals are wide. Thus, comparability is hard to assess quantitatively; qualitatively, both the WHI and the NHS appear to show excess risk of heart disease immediately following initiation of hormones. Furthermore, this risk is, more pronounced among those who initiate such therapy more than a decade after menopause than in those who initiate it earlier. The cumulative excess risk dissipates with time within both studies, perhaps progressing to a decreased risk with long-term use.

Hernán et al believe that the aforementioned differences largely explain any discrepancies between the WHI and NHS results, and that residual confounding for the initiation of therapy in the NHS plays small role. The major suspect when observational study results and trial results disagree is unknown risk factors that could be controlled only by

From the Epidemiology and Biostatistics Program, Division of Epidemiology and Genetics, National Cancer Institute, Bethesda, Maryland.

Correspondence: Robert Hoover, MD, Epidemiology and Biostatistics Program, Division of Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd., EPS/8094, Bethesda, MD 20892. E-mail: hooverr@mail.nih.gov.

Copyright © 2008 by Lippincott Williams & Wilkins

ISSN: 1044-3983/08/1906-0780

DOI: 10.1097/EDE.0b013e318188e21d

randomization. With regard to hormone therapy, this concern has focused primarily on potential compliance and survivor biases for longer-term users (a healthy persistent-medication-user effect).^{6,7} As Hernán et al point out, it is harder to assess the consistency of results with an assumption of no unmeasured confounders for treatment discontinuation than for initiation. If some of the remaining discrepancies (magnitude of the initial excess risk, slope of the decline in excess risk with duration of use, and the magnitude of protection, if any, with long-term use) are not simply because of chance, the possibility of unrecognized confounding remains. The plausibility of this could be contested but will probably not be resolved by further methodologic work, now that the trial is over.

Other important methodologic observations have emerged from the comparisons and contrasts associated with these reconciliation processes. Few have been new methodologic insights per se, but several are compelling examples and illustrations of established epidemiologic principles, namely the relative strengths and weaknesses of various study designs alluded to in the first paragraph. These illustrations of central principles have the potential to become particularly effective didactic tools in educating epidemiology students and colleagues from other disciplines about the richness of our discipline's strategies to contribute to understanding health and disease. For example, one weakness of a trial typically lies in the area of its generalizability. In MacMahon's words of almost 50 years ago, "the interpretation of results in terms of general applicability may be limited."⁸ Examples of this abound for therapeutic effects, which can vary by characteristics of populations chosen for study (eg, age, sex, race, and general health status). Now we have an example of an etiologic factor the effect of which apparently varies substantially with only a 10-year difference in time of exposure. Not only does this point out the potential limitations of generalizability of a trial, and the need to define from its outset the population to which you wish it to relate, but also the implications for statistical power. Trialists understandably are excited to explore subgroup effects but rarely power their studies with this in mind. This can work for a therapy trial, because another trial to explore an interesting subgroup observation can often be launched. For prevention trials, which are larger and more complex to start with, and entail added ethical issues, further trials are rarely feasible.

Interestingly, the current circumstance points out that this "weakness" may be turned to an advantage. This can happen when important biologic insights result from assessing the effect of an exposure in an understudied group. Indeed, the nonrepresentative nature of the WHI study group was recognized at the start of the trial, and a review of its protocol by a National Academy of Sciences committee suggested that the inclusion of women a decade after menopause was one of the strengths of the study design for precisely this reason.⁹ This turned out to be the case. The

differences in risk of clinical CHD by timing of exposure in the trial, and the subsequent exploration of this in the limited observational data available, has contributed directly to speculation about a mechanism that would involve increased clotting risk in women with preexisting subclinical CHD and different effects and mechanisms in those without.

Another key design effect is illustrated by the difficulty in unbiased assessment of the risks of hormone therapy shortly after beginning use (exposures beginning between follow-up efforts) in the NHS. Although a general strength of the cohort design is in unbiased exposure assessment, it frequently presents challenges in assessment of the details of timing of exposure and risk. The multiple endpoints under study, multiple risk factors being assessed, time intervals between exposure assessments, and deaths and losses to follow-up within these intervals (the probability of which may relate to exposure), all conspire to obscure detailed temporal relations between exposure and disease. This contrasts with the relative ease of collecting data in a case-control design, with its focus on 1 disease and a limited set of potential risk factors, with total exposure histories for these factors up to diagnosis. If these potential misclassification biases are an issue in even the NHS, with its compliant population and 2-year follow-up intervals, the hormone therapy story must be a cautionary tale for epidemiologists working with cohort studies that have higher drop-out rates and longer intervals between exposure updates.

When differing results emanate from different disciplines—or from different high-quality studies within the same discipline—good public health practice can be hard to define. Fortunately in this instance, the public health message for menopausal women is quite clear, despite any residual methodologic questions or differences of opinion. In addition to the patterns of risk of CHD, hormone therapy increases the risk of breast cancer, with a pattern of risk a mirror image of that for CHD—high levels of risk emerging with longer-term use, and in those initiating therapy near the time of menopause.¹⁰ Hormone therapy is also associated with increased risks of stroke,¹¹ blood clots,¹² dementia,¹³ gall bladder disease,¹⁴ and urinary incontinence¹⁵ and shows no clinically significant benefit for health-related quality of life.¹⁶ These cumulative risks so overwhelm the protective effects for osteoporosis that a clear consensus of epidemiologists and trialists has emerged. Hormonal therapy may be appropriate for the short-term treatment of moderate-to-severe vasomotor symptoms in recently menopausal women, but it should not be used long-term for the prevention of chronic disease.

In the immediate aftermath of the controversy 6 years ago, the editors of *Epidemiology* called on our community to address the issues raised by the WHI results, arguing that such an effort "cannot help but sharpen the peculiar and rigorous qualities of thinking demanded by observational

data.”¹⁷ Many of our best epidemiologists and biostatisticians have responded, and the intensity with which they have contemplated the navel of our discipline since that time has clearly achieved this goal.

ABOUT THE AUTHOR

ROBERT HOOVER is Director of the intramural Epidemiology and Biostatistics Program at the National Cancer Institute. He published the first epidemiologic study suggesting menopausal hormone therapy as a cause of breast cancer. He is continually humbled by how much he still has to learn about the epidemiologic method after nearly 40 years of trying to practice it.

REFERENCES

- Kolata G. Hormone studies: what went wrong? *The New York Times*. April 22, 2003.
- Grodstein F, Manson JE, Stampfer MJ. Hormone therapy and coronary heart disease: the role of time since menopause and age at hormone initiation. *J Women's Health*. 2006;15:1, 35–44.
- Rossouw JE. Postmenopausal hormone therapy for disease prevention: have we learned any lessons from the past? *Nature*. 2008;83:14–16.
- 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.
- 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.
- Sturgeon SR, Schairer C, Brinton LA, et al. Evidence of a healthy estrogen user survivor effect. *Epidemiology*. 1995;6:227–231.
- Egeland GM, Kuller LH, Mathews KA, et al. Premenopausal determinants of menopausal estrogen use. *Prev Med*. 1991;20:343–349.
- MacMahon B. Experimental epidemiology. In: MacMahon, Pugh, Ipsen, eds. *Epidemiologic Methods*. Boston, MA: Little Brown and Co; 1960.
- Thaul S, Hotra D, eds. *An Assessment of the NIH Women's Health Initiative*. Washington, DC: National Academy Press; 1993.
- 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.
- Wassertheil-Smoller S, Hendrix SL, Limacher M, et al. Effect of estrogen plus progestin on stroke in postmenopausal women. *JAMA*. 2003;289:2673–2684.
- Cushman M, Kuller LH, Prentice R, et al. Estrogen plus progestin and risk of venous thrombosis. *JAMA*. 2004;292:1573–1580.
- Shumaker SA, Legault C, Rapp SR, et al. Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women. *JAMA*. 2003;289:2651–2662.
- Cirillo DJ, Wallace RB, Rodabough RJ, et al. Effect of estrogen therapy on gallbladder disease. *JAMA*. 2005;293:330–339.
- Hendrix SL, Cochrane BB, Nygaard IE, et al. Effects of estrogen with and without progestin on urinary incontinence. *JAMA*. 2005;293:935–948.
- Hays J, Ockene JK, Brunner RL, et al. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med*. 2003;348:1839–1854.
- Editors. Epidemiology and randomized clinical trials. *Epidemiology*. 2003;14:2.