

THE **B** FILES

Case studies of bias in real life epidemiologic studies

Bias File 1. The Rise and Fall of Hormone Replacement Therapy

Compiled by

Madhukar Pai, MD, PhD

Jay S Kaufman, PhD

Department of Epidemiology, Biostatistics & Occupational Health

McGill University, Montreal, Canada

madhukar.pai@mcgill.ca & jay.kaufman@mcgill.ca



THIS CASE STUDY CAN BE FREELY USED FOR EDUCATIONAL PURPOSES WITH DUE CREDIT

Bias File 1. The Rise and Fall of Hormone Replacement Therapy

The story

By the mid-1990s, hormone replacement therapy (HRT) had become one of the most widely prescribed medications for women, especially in North America. Several observational studies had shown that women who took long-term estrogen replacement therapy had lower risk of cardiovascular disease. In the late 1990s, a clinical trial called HERS [Heart and Estrogen-progestin Replacement Study], found that estrogen therapy increased, rather than decreased, the likelihood that women who already had heart disease would suffer a heart attack. In 2002, a second trial, the Women's Health Initiative [WHI], concluded that HRT constituted a potential health risk for all postmenopausal women. Randomized trials had suddenly over-turned the long-held belief (from observational studies) that HRT was beneficial for prevention of heart disease. Subsequently, the use of HRT declined worldwide. So, what went wrong and why?

The study

Several observational studies, including large cohort studies, showed a cardiovascular benefit for HRT. For example, in the Nurses' Health Study [NHS] (Stampfer et al. 1991), investigators followed 48,470 postmenopausal women, 30 to 63 years old, and who did not have a history of cancer or cardiovascular disease at base line. During up to 10 years of follow-up (337,854 person-years), they documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes. After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80); the risk was significantly reduced among women with either natural or surgical menopause. The investigators concluded that "current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease."

A widely cited systematic review of several observation studies, published in 1992, by Grady et al, estimated the pooled relative risk, using meta-analysis, to be 0.65, which translated to about 35% reduction in coronary heart disease. The authors concluded that "there is evidence that estrogen therapy decreases risk for coronary heart disease... and hormone therapy should probably be recommended for women who have had a hysterectomy and for those with coronary heart disease or at high risk for coronary heart disease."

The bias

One of the best, clearest descriptions of the HRT story is by an article in NY Times by Gary Taubes entitled "Do We Really Know What Makes Us Healthy?" [Sept 2007]. A more technical, expert review is by Barrett-Connor et al, entitled "The rise and fall of menopausal hormone therapy" [Annu Rev Public Health 2005].

There are several inter-related biases that may explain why observational studies were wrong about

Do We Really Know What Makes Us Healthy?



By GARY TAUBES
Published: September 16, 2007

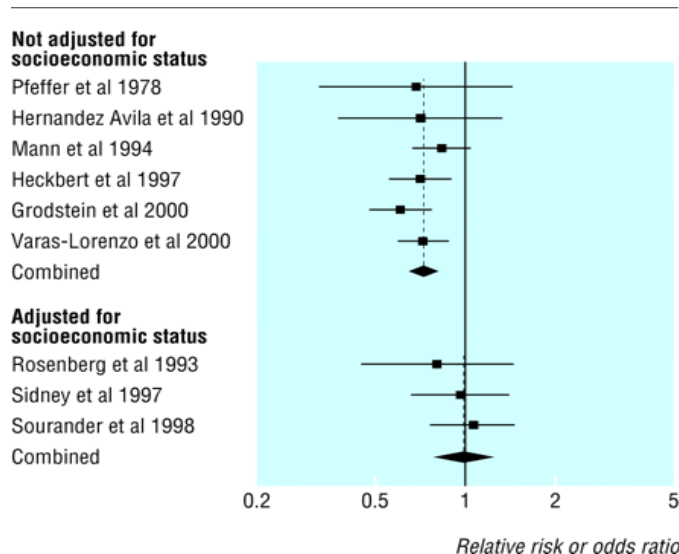
Reinhard Hunger

© 2007 NY TIMES

HRT and heart disease. The first is called "healthy user bias." As Gary Taubes described nicely, "people who faithfully engage in activities that are good for them — taking a drug as prescribed, for instance, or eating what they believe is a healthy diet — are fundamentally different from those who don't. One thing epidemiologists have established with certainty, for example, is that women who take HRT differ from those who don't in many ways, virtually all of which associate with lower heart-disease risk: they're thinner; they have fewer risk factors for heart disease to begin with; they tend to be more educated and wealthier; to exercise more; and to be generally more health conscious."

Next, there is another subtle component of healthy-user (or "healthy continuer") bias. This is the "compliance or adherer effect or bias". Individuals who comply or adhere with their doctors' orders when given a prescription are different and healthier than people who don't. Those who took HRT every day, in all likelihood, did other things that may have reduced their risk of heart disease (avoid smoking, daily exercise, better diet, etc.).

The last related issue is lack of adequate adjustment for bias due to socioeconomic status. Observational studies did adjust for confounding, but probably residual confounding remained. In a BMJ editorial entitled "The scandal of poor epidemiological research", the authors pointed out that "a protective effect of HRT was evident in studies that did not control for socioeconomic status, but not in studies that did (shown in the figure below). Higher socioeconomic position is strongly associated with both more frequent use of hormone replacement therapy and lower risk of coronary heart disease. In the large (unconfounded) Women's Health Initiative randomized trial HRT had no beneficial effect on cardiovascular disease.



Meta-analysis of cohort studies and case-control studies of hormone replacement therapy and coronary heart disease. There is little evidence for a protective effect when analyses are adjusted for, in contrast to studies not adjusted for, socioeconomic status.

Source: BMJ 2004;329:868-869.

The lesson and the evolving saga

Observational epidemiologic studies should always be interpreted cautiously, because confounding is almost always likely, and not all studies are able to prevent or adjust for confounding adequately. The HRT story also reminds us that repeated observational studies can consistently show the same effect, but all can be consistently biased! Lastly, new therapies and interventions must be subjected to rigorous randomized controlled trials, before they become widely used. Observational evidence alone may be

inadequate or even misleading. In the case of HRT, the story has evolved since the first RCTs on HRT. A 2009 paper by Vandenbroucke entitled "The HRT controversy: observational studies and RCTs fall in line" provides a nice snap shot of current thinking on this topic. According to this new paper, "For coronary heart disease, the results of observational data and trials fell in line, mainly by analysing the data according to time since start of HRT. For randomised trials, this is the natural analysis because therapy starts at randomisation. In the Women's Health Initiative and other trials, the first years of hormone replacement by combined oestrogen-progestin did increase coronary heart disease, which then waned. The analysis of the observational studies, however, had mostly been a contrast between current users at the time of enrolment to never users. Most current users were past the window wherein coronary heart disease risk was increased and were in a phase of decreased incidence. When cohort data from the observational part of the Women's Health Initiative were reanalysed according to time since start of therapy, the same pattern emerged of an initial increase in risk, followed by a decrease. Thus nothing was intrinsically wrong with the observational data; what went wrong was an analysis that had not taken into account that the effect of HRT might be different over time. The piece of evidence that closes the case is the recent reanalysis of the Nurses' Health Study on combined oestrogen-progestin and coronary heart disease, which finds the same pattern of an initial increase in risk by contrast with the original analysis which showed overall protection."

The re-analysis of NHS, cited by Vandenbroucke, was published by Hernan and colleagues (2008) and entitled "Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease." The WHI investigators found a greater CHD risk in the estrogen plus progestin therapy arm than in the placebo arm of the trial (hazard ratio: 1.24, 95% CI: 1.00–1.54). In contrast, the NHS investigators found a lower CHD risk in current users of combined hormone therapy than in never users (HR: 0.68, 95% CI: 0.55–0.83). To reconcile these differences, Hernan and colleagues reanalyzed the NHS data by using a novel approach that conceptualizes a follow-up observational study as a sequence of "trials." The reanalysis showed that there is a short-term increase in CHD incidence after initiation of combined hormone therapy among all NHS women. The analysis also suggested effect modification of the hazard ratio for combined hormone therapy by years since menopause. This finding is consistent with the so-called "timing hypothesis," which states that the increased CHD risk is concentrated in women who start combined hormone therapy many years after menopause. The authors of the re-analysis claimed that "discrepancies between previous NHS and WHI results in regard to the 2 results above appear to be due to the NHS analytic approach and not to any inherent problems in the NHS data." (Hernan 2008). Their paper generated heated debate (Hoover 2008; Stampfer 2008; Prentice 2008), and while the dust is yet to settle, this example illustrates the importance of attempting newer statistical approaches that may overcome some of the limitations of currently used analytic methods in observational studies.

Sources and suggested readings*

1. Stampfer M et al. Postmenopausal estrogen therapy and cardiovascular disease. Ten-year follow-up from the nurses' health study. *N Engl J Med*. 1991 Sep 12;325(11):756-62.
2. Grady D, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med*. 1992 Dec 15;117(12):1016-37.
3. Hulley S, 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.

4. Rossouw JE, 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;288:321–33.
5. Taubes, G. Do We Really Know What Makes Us Healthy? *New York Times*, September 16, 2007.
6. Barrett-Connor E et al. The rise and fall of menopausal hormone therapy. *Annu Rev Public Health*. 2005;26:115-40.
7. von Elm E et al. The scandal of poor epidemiological research. *BMJ* 2004;329:868-869.
8. Vandenbroucke JP. The HRT controversy: observational studies and RCTs fall in line. *Lancet* 2009; Apr 11;373(9671):1233-5.
9. Hernan M et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008 Nov;19(6):766-79.
10. Hoover RN. The sound and the fury: was it all worth it? *Epidemiology*. 2008 Nov;19(6):780-
11. Stampfer MJ. ITT for observational data: worst of both worlds? *Epidemiology* 2008 Nov;19(6):783-4.
12. Prentice RL. Data analysis methods and the reliability of analytic epidemiologic research. *Epidemiology*. 2008 Nov;19(6):785-8

*From this readings list, the most relevant papers are enclosed.

POSTMENOPAUSAL ESTROGEN THERAPY AND CARDIOVASCULAR DISEASE

Ten-Year Follow-up from the Nurses' Health Study

MEIR J. STAMPFER, M.D., GRAHAM A. COLDITZ, M.B., B.S., WALTER C. WILLET, M.D.,
JOANN E. MANSON, M.D., BERNARD ROSNER, PH.D., FRANK E. SPEIZER, M.D.,
AND CHARLES H. HENNEKENS, M.D.

Abstract Background. The effect of postmenopausal estrogen therapy on the risk of cardiovascular disease remains controversial. Our 1985 report in the *Journal*, based on four years of follow-up, suggested that estrogen therapy reduced the risk of coronary heart disease, but a report published simultaneously from the Framingham Study suggested that the risk was increased. In addition, studies of the effect of estrogens on stroke have yielded conflicting results.

Methods. We followed 48,470 postmenopausal women, 30 to 63 years old, who were participants in the Nurses' Health Study and who did not have a history of cancer or cardiovascular disease at base line. During up to 10 years of follow-up (337,854 person-years), we documented 224 strokes, 405 cases of major coronary disease (nonfatal myocardial infarctions or deaths from coronary causes), and 1263 deaths from all causes.

Results. After adjustment for age and other risk factors, the overall relative risk of major coronary disease in women currently taking estrogen was 0.56 (95 percent confidence interval, 0.40 to 0.80); the risk was significantly reduced among women with either natural or surgical

menopause. We observed no effect of the duration of estrogen use independent of age. The findings were similar in analyses limited to women who had recently visited their physicians (relative risk, 0.45; 95 percent confidence interval, 0.31 to 0.66) and in a low-risk group that excluded women reporting current cigarette smoking, diabetes, hypertension, hypercholesterolemia, or a Quetelet index above the 90th percentile (relative risk, 0.53; 95 percent confidence interval, 0.31 to 0.91). The relative risk for current and former users of estrogen as compared with those who had never used it was 0.89 (95 percent confidence interval, 0.78 to 1.00) for total mortality and 0.72 (95 percent confidence interval, 0.55 to 0.95) for mortality from cardiovascular disease. The relative risk of stroke when current users were compared with those who had never used estrogen was 0.97 (95 percent confidence interval, 0.65 to 1.45), with no marked differences according to type of stroke.

Conclusions. Current estrogen use is associated with a reduction in the incidence of coronary heart disease as well as in mortality from cardiovascular disease, but it is not associated with any change in the risk of stroke. (N Engl J Med 1991; 325:756-62.)

THE influence of exogenous hormones on the risk of cardiovascular disease has long been controversial. More than 20 studies published in the past decade have addressed the issue of postmenopausal estrogen use and coronary disease.¹ Our earlier report of a benefit from estrogen use in terms of the risk of coronary disease, based on four years of follow-up,² was accompanied by a report from the Framingham Study that came to the opposite conclusion.³ These disparate findings led to considerable confusion.⁴ We now report results for both coronary disease and stroke, based on 10 years of follow-up in the Nurses' Health Study, a large cohort study that included 48,470 postmenopausal women with 337,252 person-years of follow-up.

METHODS

The Nurses' Health Study Cohort

The Nurses' Health Study began in 1976, when 121,700 female registered nurses in the United States completed questionnaires sent to them by mail about their medical history, including previous cardiovascular disease, menopause, diabetes, hypertension, high serum cholesterol levels, and parental myocardial infarction. We included questions on height, weight, smoking, the use of postmeno-

pausal hormones, and the use of oral contraceptives.⁵ Every two years, follow-up questionnaires were mailed to obtain updated information and identify newly diagnosed major illnesses. A dietary questionnaire was added in 1980.⁶

Ascertainment of Estrogen Use

In 1976 the women were asked whether they had taken hormone supplements after menopause, and if so, for how long. Information on hormone use, including the type taken, was updated in the subsequent questionnaires sent every two years through 1986, with explicit questions about current use and duration of use in the intervening period. Because no information on current use was explicitly requested on the 1976 questionnaire, we considered women to have been current estrogen users for the 1976-1978 period if the duration of their estrogen use was equal (within 12 months) to the interval between menopause and the date of completion of the questionnaire. Women whose duration of hormone use was more than 12 months shorter than this interval were considered former users. The daily dose of conjugated estrogens was obtained beginning in 1980.

Identification and Confirmation of Cardiovascular End Points

The study end points included nonfatal myocardial infarction, fatal coronary heart disease, coronary-artery bypass grafting or angioplasty, fatal and nonfatal stroke, total cardiovascular mortality, and deaths from all causes after the return of the 1976 questionnaire but before June 1, 1986. Nurses who reported having a nonfatal myocardial infarction or stroke on a follow-up questionnaire were asked for permission for a study investigator to review their medical records. Nonfatal myocardial infarctions were considered confirmed by hospital records if they met the World Health Organization criteria⁷ (i.e., symptoms plus either cardiac-enzyme elevations or diagnostic electrocardiographic changes). Myocardial infarctions that required hospitalization and for which confirmatory information was obtained by interview or letter, but for which no medical records were obtainable, were designated as probable. Thus, infarc-

From the Channing Laboratory, Departments of Medicine (M.J.S., G.A.C., W.C.W., J.E.M., B.R., F.E.S., C.H.H.) and Preventive Medicine (B.R., C.H.H.), Harvard Medical School and Brigham and Women's Hospital, and the Departments of Epidemiology (M.J.S., G.A.C., W.C.W.), Nutrition (W.C.W.), Biostatistics (B.R.), and Environmental Health (F.E.S.), Harvard School of Public Health, all in Boston. Address reprint requests to Dr. Stampfer at Channing Laboratory, 180 Longwood Ave., Boston, MA 02115.

Supported by research grants (HL 34594 and CA 40356) from the National Institutes of Health.

tions of indeterminate duration discovered on routine examination were not included. Coronary-artery surgery was ascertained by the participants' reports alone.

Nonfatal strokes were considered confirmed by a review of medical records if they were characterized by a typical neurologic deficit, rapid in onset and lasting at least 24 hours, and if they met the criteria of the National Survey of Stroke.⁸ We classified strokes as ischemic strokes (thrombotic or embolic occlusion of a cerebral artery), subarachnoid hemorrhages, or intraparenchymal hemorrhages. We excluded subdural hematomas and strokes caused by infection or neoplasia. Strokes reported on the questionnaires that required hospitalization and were confirmed by information from a letter or telephone call, but for which the medical records were unavailable, were designated as probable.

Most deaths were reported by the participants' families. We used the National Death Index⁹ to identify deaths among the nonrespondents to each two-year questionnaire; the mortality follow-up was more than 98 percent complete. For all deaths possibly attributable to cardiovascular causes, we requested permission from the next of kin (subject to state regulations) to review the medical records. Deaths were considered to be due to coronary disease if the medical records or autopsy findings confirmed that a fatal myocardial infarction had occurred. The category of coronary death also included cases in which coronary disease was listed on the death certificate as the underlying cause without another, more plausible cause and in which the nurse was known (e.g., on the basis of the hospital record or an interview with her next of kin) to have had coronary disease before death. In no case was the cause listed on the death certificate used as the sole criterion for a determination of coronary death. We classified strokes as fatal if they were documented by autopsy findings or hospital records or if stroke was listed as the underlying cause of death on the death certificate.

The category of cardiovascular mortality included deaths from stroke, deaths from coronary disease, sudden deaths (death within one hour of the onset of symptoms in an apparently healthy woman), and deaths for which coronary disease was listed as the underlying cause and no more plausible cause could be assigned, but for which confirmation was lacking. Major cardiovascular disease was defined to include both death from cardiovascular disease and nonfatal myocardial infarction and stroke. All the interviews and reviews of medical records were conducted without the investigators' knowledge of the category of estrogen use.

Population for Analysis

Women for whom information on hormone use was missing (3.6 percent of all respondents) were excluded from the analysis. Because women with diagnosed cardiovascular disease may alter their hormone use and are also at increased risk for progression of the disease, their inclusion could distort the results. We therefore excluded from the analysis all women who reported a diagnosis of any cardiovascular disease or cancer (except skin cancer other than melanoma) on the 1976 questionnaire. Similarly, women who reported such a diagnosis on a subsequent questionnaire were excluded from further analysis. Thus, at the start of each two-year interval, the base population included no women reporting these diagnoses. For the analyses of mortality from all causes, however, these women were included, so that deaths due to illnesses lasting more than two years could be considered.

We classified women as postmenopausal from the time they reported having a natural menopause or undergoing hysterectomy with bilateral oophorectomy. Women who underwent hysterectomy without bilateral oophorectomy were considered postmenopausal when they reached the age at which natural menopause had occurred in 90 percent of the cohort (54 years for smokers and 56 for nonsmokers). The women's reports of reaching menopause were highly accurate in this cohort.¹⁰

In 1976, a total of 22,950 postmenopausal women entered the analysis for the 1976–1978 period. The population was expanded to include women who became postmenopausal subsequently and were free of cancer and cardiovascular disease. During the 10-year period from 1976 through June 1, 1986, we accrued 337,854 person-years of follow-up among 48,470 women. The follow-up of the cohort, calculated as a percentage of the total potential person-years

of follow-up, was 88.4 percent complete for nonfatal outcomes; for mortality, it was more than 98 percent complete. Follow-up rates were quite similar within the different categories of hormone use.

Statistical Analysis

For each participant, person-months were allocated to the categories of hormone use according to the data reported in 1976 and updated at each two-year interval according to information obtained subsequently. Follow-up for a participant ended with a diagnosis of cardiovascular disease or death. If no questionnaire was returned for a two-year follow-up period, the most recent data were applied to the subsequent follow-up interval. If a woman's previous status had been current hormone use, however, she was classified in the update as having used hormones at some time, but current or former use was not specified.

We calculated the relative risk associated with hormone use, defined as the incidence rate of cardiovascular disease among hormone users (estimated as the number of events divided by the person-time of follow-up for the hormone users) divided by the corresponding rate among women who had never used hormones. Age-specific rates of cardiovascular disease for users and nonusers were calculated in five-year categories and used to compute age-adjusted relative risks with 95 percent confidence intervals.¹¹ To adjust for a number of risk factors simultaneously, we used proportional-hazards models.¹² All P values are two-tailed.

RESULTS

Women currently using postmenopausal hormones accounted for 21.8 percent of the total follow-up time of 337,854 person-years. Former hormone users accounted for 25.2 percent of the time, and women who had never used hormones 53 percent. In all three groups, potential risk factors for cardiovascular disease were distributed in generally similar patterns. Table 1 shows the age-standardized proportions of

Table 1. Distribution of Characteristics and Coronary Risk Factors Reported by the Women in the Cohort, According to Postmenopausal Hormone Use, with Standardization for Age.*

VARIABLE	HORMONE USE		
	CURRENT	FORMER	NONE
	percent of subjects		
Parental MI before the age of 60	10.6	10.0	9.3
Hypertension	23.2	25.0	21.8
Diabetes mellitus	2.7	3.8	3.5
High serum cholesterol	9.9	11.2	7.6
Current smoker (15–24 cigarettes/day)	11.2	14.7	14.5
Quetelet index ≥ 29 †	9.8	13.3	15.0
Bilateral oophorectomy	50.3	39.3	9.3
Past use of oral contraceptives	34.0	27.6	23.9
Vigorous physical activity ≥ 1 time/week‡	48.2	43.1	42.4
	grams per day		
Mean dietary intake‡§			
Saturated fat	27.6	26.2	26.7
Cholesterol	0.32	0.32	0.32
Polyunsaturated fat	8.9	8.7	8.7
Dietary fiber	17.3	17.2	16.8
Alcohol	7.9	7.5	7.3

*Data are standardized to the age distribution of the person-years of follow-up for the cohort, from 1976 through 1986. MI denotes myocardial infarction.

†The Quetelet index was calculated by dividing the weight in kilograms by the square of the height in meters.

‡As assessed in 1980 and standardized to the age distribution of the cohort at that time.

§Adjusted for energy intake.

Table 2. Relative Risk of Cardiovascular Disease among Current and Former Postmenopausal Hormone Users, as Compared with Those Who Never Used Postmenopausal Hormones, after Adjustment for Age and Multiple Risk Factors.*

GROUP†	NO. OF PERSON-YEARS	MAJOR CORONARY DISEASE		FATAL CARDIOVASCULAR DISEASE		TOTAL STROKE		ISCHEMIC STROKE		SUBARACHNOID HEMORRHAGE	
		NO. OF CASES	RR (95% CI)	NO. OF CASES	RR (95% CI)	NO. OF CASES	RR (95% CI)	NO. OF CASES	RR (95% CI)	NO. OF CASES	RR (95% CI)
No hormone use	179,194	250	1.0	129	1.0	123	1.0	56	1.0	19	1.0
Current hormone use	73,532										
Adjusted for age	—	45	0.51 (0.37–0.70)	21	0.48 (0.31–0.74)	39	0.96 (0.67–1.37)	23	1.26 (0.78–2.02)	5	0.80 (0.30–2.10)
Adjusted for age and risk factors	—	—	0.56 (0.40–0.80)	—	0.61 (0.37–1.00)	—	0.97 (0.65–1.45)	—	1.46 (0.85–2.51)	—	0.53 (0.18–1.57)
Former hormone use	85,128										
Adjusted for age	—	110	0.91 (0.73–1.14)	55	0.84 (0.61–1.15)	62	1.00 (0.74–1.36)	34	1.14 (0.75–1.74)	12	1.42 (0.70–2.90)
Adjusted for age and risk factors	—	—	0.83 (0.65–1.05)	—	0.79 (0.56–1.10)	—	0.99 (0.72–1.36)	—	1.19 (0.77–1.86)	—	1.03 (0.47–2.25)

*RR denotes relative risk, and CI confidence interval.

†Women with no hormone use served as the reference category in this analysis. The risk factors included in the multivariate models were age (in five-year categories), cigarette smoking (none, former, current [1 to 14, 15 to 24, and ≥ 25 cigarettes per day]), hypertension (yes, no), diabetes (yes, no), high serum cholesterol level (yes, no), parental myocardial infarction before the age of 60 (yes, no), Quetelet index (in five categories), past use of oral contraceptives (yes, no), and time period (in five two-year periods).

women who reported various characteristics and coronary risk factors according to their estrogen-use status, on the basis of cumulative person-years from 1976 through 1986. Table 1 also shows the mean intake of various nutrients, with adjustment for energy intake,¹³ and the proportion of women reporting a period of vigorous exercise at least once per week, both of which were ascertained in 1980. Estrogen users were less likely to have diabetes and more likely to be lean, to engage in regular, vigorous physical activity, to have had a surgical menopause, and to have used oral contraceptives in the past.

Among the postmenopausal women who reported no previous cardiovascular disease, we documented 293 nonfatal myocardial infarctions (228 confirmed and 65 probable), 112 confirmed deaths from coronary disease, and 224 strokes (52 fatal and 172 nonfatal; 177 confirmed and 47 probable) during the 10 years of follow-up. Of the strokes, 113 were ischemic strokes and 36 were subarachnoid hemorrhages; the remaining strokes were of other or unknown types. There were 41 other deaths from cardiovascular causes, for a total of 205 cardiovascular deaths. Coronary-artery surgery or angioplasty was reported by 185 women. In the analyses of total mortality, which included women in whom illnesses developed during follow-up, there were 1263 deaths from all causes. No material differences were observed in any of the analyses between the confirmed and the probable categories of myocardial infarction and stroke or between the fatal and the nonfatal categories of coronary disease or stroke; we therefore merged these categories into two larger categories: major coronary disease (nonfatal myocardial infarction and death from coronary causes) and total stroke.

Overall, the age-adjusted risk of major coronary disease among current estrogen users was about half that of women who had never used estrogen, with a relative risk of 0.51 (95 percent confidence interval, 0.37 to 0.70; $P < 0.0001$) (Table 2). For former users, the age-adjusted relative risk was 0.91 (95 per-

cent confidence interval, 0.73 to 1.14; $P = 0.42$). In contrast, we observed no association between current estrogen use and total stroke. The age-adjusted relative risk was 0.96 (95 percent confidence interval, 0.67 to 1.37) and was virtually unchanged after further adjustment for other cardiovascular risk factors. No material associations were observed for ischemic stroke or subarachnoid hemorrhage; there were too few cases of intraparenchymal hemorrhage for analysis.

We observed no apparent association between estrogen use and the incidence of coronary-artery surgery. Among the current users, the age-adjusted relative risk was 1.21 (95 percent confidence interval, 0.84 to 1.73), and for former users it was 0.86 (95 percent confidence interval, 0.60 to 1.22). Among the former users, there were no notable trends with regard to duration of use or time since most recent use. Simultaneous adjustment for other risk factors in multivariate analyses had virtually no effect on these estimates. We found no evidence to suggest that the degree of protection associated with current estrogen use was related to the duration of use, independent of age, for any of the end points; among the former users, the period of time since the cessation of estrogen use was not consistently related to the risk of cardiovascular outcomes (data available elsewhere*).

The study had insufficient statistical power to determine the effects of specific forms of hormone therapy other than unopposed oral conjugated estrogen. Of the 57,570 person-years of follow-up for current hormone users from 1978 through 1986, 71.5 percent involved the use of unopposed oral conjugated estrogen, 11.5 percent other estrogens, 2.7 percent estrogens with progestin, 2.2 percent other hormones, and 12

*See NAPS document no. 04890 for eight pages of supplementary material. Order from NAPS c/o Microfiche Publications, P.O. Box 3513, Grand Central Station, New York, NY 10163-3513. Remit in advance (in U.S. funds only) \$7.75 for photocopies or \$4 for microfiche. Outside the U.S. and Canada add postage of \$4.50 (\$1.50 for microfiche postage). There is a \$15 invoicing charge on all orders filled before payment.

percent hormones of unknown type (or information was missing). The age-adjusted relative risk of major coronary disease with current use of unopposed oral conjugated estrogen was 0.40 (95 percent confidence interval, 0.26 to 0.62).

Information about the dose of conjugated estrogen was available for the period from 1980 through 1986. The only marked difference in the association observed with different dose levels was an apparent increase in the risk of coronary disease among women taking more than 1.25 mg per day (relative risk, 2.8; 95 percent confidence interval, 0.9 to 8.2), as compared with the substantial decrease in risk among those taking lower doses. The use of estrogen at doses of more than 1.25 mg per day was very uncommon (4 percent of the cohort), however, and the relative risk is based on only three cases.

We assessed whether the inverse association of estrogen use with the risk of coronary disease differed for women with different characteristics. We observed few marked differences in the associations. The age-specific relative risk appeared to show a nonsignificant trend ($P = 0.19$) toward more protection from coronary disease among younger postmenopausal hormone users. For the oldest age group, women 60 to 64 years of age, the relative risk was 1.35 (95 percent confidence interval, 0.65 to 2.82). We noted possible tendencies toward more protection among smokers than among nonsmokers, among women without a parental history of myocardial infarction before the age of 60, and among the leanest women, but these differences in relative risks were not statistically significant.* Among the women who had a natural menopause, the age-adjusted relative risk of major coronary disease for current estrogen users was 0.62 (95 percent confidence interval, 0.39 to 0.97), not as low as the risk for women who underwent bilateral oophorectomy (relative risk, 0.40; 95 percent confidence interval, 0.22 to 0.73).

To evaluate the effect of estrogen use among women at low risk, we defined a subgroup of women who were not current smokers; had no hypertension, diabetes, or high serum cholesterol level; and had a Quetelet index below 32, the 90th percentile for this cohort. For this group, the age-adjusted relative risk of major coronary disease among current hormone users was 0.53 (95 percent confidence interval, 0.31 to 0.91).

To adjust for the effects of several potential risk factors simultaneously, we used proportional-hazards models to estimate the relative risks associated with current and former use of estrogens, controlling for age, follow-up period, and the characteristics listed in Table 1. Because the current estrogen users were slightly healthier, this adjustment attenuated the apparent benefit slightly. The results (shown as adjusted for age and risk factors) were similar to those obtained after adjustment for age alone; for major coronary disease, the relative risk among current users was 0.56 (95 percent confidence interval, 0.40 to 0.80) (Table 2). A model that also included age at menopause as a continuous variable yielded virtually the

same estimates. Similar models that included the data on dietary intake and physical activity yielded similar findings, although the estimates were less precise because only data for 1980 through 1986 could be included.

To assess whether receiving more medical care might account for the benefit in postmenopausal estrogen users, we repeated the analysis, limiting it to women who reported having visited a physician in 1978 (65 percent of the cohort). The results were similar to those for the population as a whole: the age-adjusted relative risks of major coronary heart disease were 0.45 (95 percent confidence interval, 0.31 to 0.66) for current estrogen users and 0.79 (95 percent confidence interval, 0.60 to 1.05) for former users. For cardiovascular mortality, the age-adjusted relative risks were 0.52 (95 percent confidence interval, 0.40 to 0.69) for current users and 0.77 (95 percent confidence interval, 0.62 to 0.95) for former users.

In analyzing mortality from all causes, we focused primarily on women who had used estrogen at any time, in order to avoid the potential problem created by shifts in status from current to former use as a result of a diagnosis of disease. We also eliminated the requirement that the cohort be free of diagnosed cancer and heart disease at the beginning of each two-year period; this allowed us to include deaths due to illnesses lasting more than two years. Thus, in this analysis, the cohort was free from diagnosed cancer and heart disease at base line in 1976 (or at entry into the analysis, for those who became postmenopausal later) and was followed until death or the cutoff date of May 31, 1986. For women who had used estrogens at any time, the age-adjusted relative risk of mortality from all causes was 0.81 (95 percent confidence interval, 0.72 to 0.91; $P = 0.0004$); for cardiovascular mortality, it was 0.68 (95 percent confidence interval, 0.52 to 0.90). After adjustment for other risk factors, the relative risks were slightly attenuated, but they remained statistically significant; for total mortality, the risk was 0.89 (95 percent confidence interval, 0.78 to 1.00), and for cardiovascular mortality it was 0.72 (95 percent confidence interval, 0.55 to 0.95; $P = 0.02$). Because in the earlier analyses benefits had been found to be attributable to current estrogen use, this analysis underestimated the benefit of estrogen by including former users with current users. To remove this bias in part, we excluded women who had already discontinued estrogen use at base line but not those who used estrogen at base line and discontinued it later. The exclusion of the latter group would have led to an overestimate of the benefit, because estrogen therapy is often discontinued in women who have potentially fatal illnesses, such as breast cancer.

DISCUSSION

In this prospective study of 48,470 women, we observed that when current postmenopausal estrogen users were compared with women who had never used estrogen, they had about half the risk of major coronary disease or fatal cardiovascular disease and no

increase in the risk of stroke. The prospective study design virtually eliminated the biases in recall and selection that can affect case-control studies. The follow-up rate was high, particularly for fatal outcomes, reducing the likelihood that differential follow-up could have affected the results.

Information on exposure to estrogen and other potential risk factors was derived from reports by the women themselves, but we believe them to be reliable. The reports have been validated by a review of the medical records and by direct measurement with respect to several conditions.^{6,14} Also, the risk factors reported by the subjects were strong predictors of subsequent cardiovascular disease,^{2,5,15,16} and the subjects were all registered nurses with a demonstrated interest in medical research.

The most plausible alternative to a cause-effect relation between estrogen use and the reduced risk of coronary disease is that healthier women are selected for such therapy. In this cohort, however, the estrogen users appeared only slightly healthier than the non-users and were generally similar to them with respect to most cardiovascular risk factors. The estrogen users had a much higher incidence of bilateral oophorectomy, a coronary risk factor only for women not receiving estrogen-replacement therapy.^{17,18} The estrogen users also tended to be leaner, which may result in lower levels of estrogen from adipose tissue.¹⁹ The likelihood of lower levels of endogenous estrogen in thinner women is consistent with the trend toward a greater benefit from postmenopausal estrogen with respect to coronary disease in that group, but the protection associated with estrogen use was present in women in all categories of the Quetelet index. The stratified and multivariate proportional-hazards models indicated only minor overall confounding, as judged by the similarity of the relative risks after adjustment for age alone with those that took account of other risk factors. The similar benefit in the analysis limited to women who reported a recent visit to a physician suggests that access to medical care appeared to have little effect on estimates of the effect of estrogen on the risk of cardiovascular disease.

The apparent marked benefit of estrogen in reducing the risk of coronary disease is consistent with previous evidence. Of 15 other prospective studies, 14 found decreased risks among estrogen users.¹ The Framingham Study alone found an elevated risk,³ which was not statistically significant when women with angina were omitted. A subsequent reanalysis of the Framingham data showed a nonsignificant protective effect among younger women but a nonsignificant adverse effect among older women.²⁰ Similarly, all three cross-sectional studies of coronary angiography showed substantially less atherosclerosis among estrogen users.²¹⁻²³ A quantitative overview of previous studies taken together yielded a relative risk of 0.56 (95 percent confidence interval, 0.50 to 0.61); when only the analytic prospective and angiographic studies

were considered, the relative risk was 0.50 (95 percent confidence interval, 0.43 to 0.56).¹

The nonsignificant trend in our data toward a decreasing benefit of estrogen with increasing age is consistent with the Framingham data,^{3,20} but Henderson et al. found a substantial reduction in risk among women in their 70s.²⁸ Future follow-up will clarify this issue, but the weight of the evidence suggests a protective effect among postmenopausal women of all ages. In the analysis of women with a favorable risk-factor profile, the observed age-adjusted relative risk of major coronary disease, 0.53, was virtually identical to that for the whole cohort. This implies that women at lower risk enjoy the same relative benefit from estrogen as women in general. Because rates of coronary disease were lower among the low-risk women, however, the same relative decrease corresponded to a smaller reduction in the number of events.

As in other studies,¹ we found that the benefit of therapy was evident primarily among current estrogen users, and there was no indication of an effect of the duration of use independent of age. The best-supported mechanism is the markedly favorable effect of estrogen on serum lipids: estrogens raise the level of high-density lipoprotein cholesterol and lower that of low-density lipoprotein cholesterol. Although estrogen-induced changes in lipid metabolism are sufficient to explain a large reduction in the risk of coronary disease,²⁴ other plausible mechanisms have been proposed.²⁵ We observed less benefit, and perhaps an adverse effect, among women taking more than 1.25 mg of estrogen daily. Such high doses were common in the Framingham cohort, which may partly explain their discrepant results.

The absence of an association between estrogen use and the incidence of coronary-artery surgery was unexpected, particularly in view of evidence from cross-sectional angiographic studies showing a strong association of estrogen use with a reduction in atherosclerosis.²¹⁻²³ Perhaps women taking estrogens under closer medical supervision are more likely to undergo coronary surgery when they have a given level of symptoms than women not taking estrogens.

We found no effect of estrogens on the incidence of total stroke or that of ischemic stroke and subarachnoid hemorrhage. In the Leisure World Study, Paganini-Hill et al.²⁶ did find a decrease in risk, but the benefit may have been overestimated because patients with previous cardiovascular disease, who may be more prone to strokes and less likely to have estrogen prescribed, were not excluded. However, this can explain only part of the observed benefit. The Framingham Study³ found an adverse effect of using estrogen at any time on the risk of stroke, whereas the large, prospective Copenhagen Study²⁷ found little effect either way. Both the Copenhagen Study and our own study included mostly middle-aged women, as compared with the Leisure World Study,²⁶ in which the median age was 73; perhaps the protective effect is limited to older women.

In analyses of total mortality, it is important to exclude subjects at the start of follow-up who have life-threatening diseases. Because women with such diseases (e.g., breast cancer) are both less likely to be prescribed estrogen and more likely to die within a given period, their inclusion in an analysis would exaggerate any benefit of estrogen. Similarly, an analysis restricted to women who continue to use estrogen would have the same effect, because women who acquire certain life-threatening conditions may be advised to cease hormone use. In our analysis, which took those considerations into account, we observed an age-adjusted relative risk of 0.81 (95 percent confidence interval, 0.72 to 0.91) and a multivariate relative risk of 0.89 (95 percent confidence interval, 0.78 to 1.00). The relative risk of cardiovascular mortality in women with any estrogen use, after adjustment for other risk factors, was 0.72 (95 percent confidence interval, 0.55 to 0.95). The benefit with respect to mortality from all causes is likely to be an underestimate, because the effects of estrogen-induced protection from hip fracture and its associated mortality would be more pronounced at an older age. Indeed, in the Leisure World Study, a relative risk of 0.64 (95 percent confidence interval, 0.52 to 0.78) was found for total mortality among current estrogen users,²⁸ but this was probably an overestimate of the true benefit, because women with prevalent disease were not omitted at base line. Bush et al. reported a relative risk of 0.54 among estrogen users for mortality from all causes,²⁹ but their result may also have overestimated the benefit, for the same reason. Petitti et al. reported a relative risk of total mortality of 0.8 (95 percent confidence interval, 0.6 to 1.1) in a follow-up study of healthy women.³⁰

The consistency of the epidemiologic data, the apparent absence of important confounding or selection bias, and biologic plausibility^{24,25} all suggest a causal association between estrogen use and a reduced risk of coronary disease. Further work is needed to identify the women most likely to benefit from hormone therapy, as well as the effect of added progestins. Proposed clinical trials among women with established coronary disease will be useful. The findings regarding mortality from all causes as well as risk-benefit analyses³¹⁻³³ suggest that, overall, the benefits of postmenopausal estrogen therapy outweigh the risks,³⁴ even apart from the substantial benefits in alleviating menopausal symptoms. These risks include an increase in the rate of endometrial cancer, which can be completely or largely blocked by the addition of a progestin, and possibly some increase in the incidence of breast cancer.

The risk-benefit assessment will differ according to a given woman's medical condition and nonmedical characteristics (including the fear of cancer), so we make no global recommendations. The decision must be made by the individual woman and her physician after they evaluate all the relevant benefits and risks.

We are indebted to the participants in the Nurses' Health Study for their continuing cooperation, and to Stefanie Bechtel, Karen Corsano, Gary Chase, Sue-Wei Chiang, Barbara Egan, Marion McPhee, Mark Shneyder, Debbie O'Sullivan, and Susan Newman for their expert help.

REFERENCES

- Stampfer MJ, Colditz GA. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 1991; 20:47-63.
- Stampfer MJ, Willett WC, Colditz GA, Rosner R, Speizer FE, Hennekens CH. A prospective study of postmenopausal estrogen therapy and coronary heart disease. *N Engl J Med* 1985; 313:1044-9.
- Wilson PWF, Garrison RJ, Castelli WP. Postmenopausal estrogen use, cigarette smoking, and cardiovascular morbidity in women over 50: the Framingham Study. *N Engl J Med* 1985; 313:1038-43.
- Bailar JC III. When research results are in conflict. *N Engl J Med* 1985; 313:1080-1.
- Colditz GA, Stampfer MJ, Willett WC, Rosner B, Speizer FE, Hennekens CH. A prospective study of parental history of myocardial infarction and coronary heart disease in women. *Am J Epidemiol* 1986; 123:48-58.
- Willett WC, Sampson L, Stampfer MJ, et al. Reproducibility and validity of a semiquantitative food frequency questionnaire. *Am J Epidemiol* 1985; 122:51-65.
- World Health Organization. IHD registers: report of the fifth working group. Copenhagen, Denmark: World Health Organization, 1971.
- Walker AE, Robins M, Wienfeld FD. The National Survey of Stroke: clinical findings. *Stroke* 1981; 12:Suppl 1:1-13-144.
- Stampfer MJ, Willett WC, Speizer FE, et al. Test of the National Death Index. *Am J Epidemiol* 1984; 119:837-9.
- Colditz GA, Stampfer MJ, Willett WC, et al. Reproducibility and validity of self-reported menopausal status in a prospective cohort study. *Am J Epidemiol* 1987; 126:319-25.
- Rothman KJ, Boice JD Jr. Epidemiologic analysis with a programmable calculator. Washington, D.C.: Public Health Service, 1979. (NIH publication no. 79-1649.)
- Cox DR. Regression models and life-tables. *J R Stat Soc [B]* 1972; 34:187-220.
- Willett WC, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* 1986; 124:17-27.
- Colditz GA, Martin P, Stampfer MJ, et al. Validation of questionnaire information on risk factors and disease outcomes in a prospective cohort study in women. *Am J Epidemiol* 1986; 123:894-900.
- Stampfer MJ, Colditz GA, Willett WC, Rosner B, Speizer FE, Hennekens CH. Coronary heart disease risk factors in women: the Nurses' Health Study experience. In: Eaker E, Packard B, Wenger NK, Clarkson TB, Tyroler HA, eds. Coronary heart disease in women. New York: Haymarket Doyma, 1987:112-6.
- Colditz GA. The Nurses' Health Study: findings during 10 years of follow-up of a cohort of U.S. women. *Curr Probl Obstet Gynecol Fertil* 1990; 13:129-74.
- Colditz GA, Willett WC, Stampfer MJ, Rosner B, Speizer FE, Hennekens CH. Menopause and the risk of coronary heart disease in women. *N Engl J Med* 1987; 316:1105-10.
- Stampfer MJ, Colditz GA, Willett WC. Menopause and heart disease: a review. *Ann N Y Acad Sci* 1990; 592:193-203.
- Judd HL, Davidson BJ, Frumar AM, Shamonki IM, Lagasse LD, Ballon SC. Serum androgens and estrogens in postmenopausal women with and without endometrial cancer. *Am J Obstet Gynecol* 1980; 136:859-71.
- Eaker ED, Castelli WP. Coronary heart disease and its risk factors among women in the Framingham Study. In: Eaker E, Packard B, Wenger NK, Clarkson TB, Tyroler HA, eds. Coronary heart disease in women. New York: Haymarket Doyma, 1987:122-32.
- Sullivan JM, Zwagg RV, Lemp GF, et al. Postmenopausal estrogen use and coronary atherosclerosis. *Ann Intern Med* 1988; 108:358-63.
- Gruchow HW, Anderson AJ, Barboriak JJ, Sobocinski KA. Postmenopausal use of estrogen and occlusion of coronary arteries. *Am Heart J* 1988; 115:954-63.
- McFarland KF, Boniface ME, Hornung CA, Earnhardt W, Humphries JO. Risk factors and noncontraceptive estrogen use in women with and without coronary disease. *Am Heart J* 1989; 117:1209-14.
- Bush TL, Miller VT. Effects of pharmacologic agents used during menopause: impact on lipids and lipoproteins. In: Mishell D, ed. Menopause: physiology and pharmacology. Chicago: Year Book Medical, 1986:187-208.
- Lobo RA. Estrogen and cardiovascular disease. *Ann N Y Acad Sci* 1990; 592:286-94.
- Paganini-Hill A, Ross RK, Henderson BE. Postmenopausal oestrogen treatment and stroke: a prospective study. *BMJ* 1988; 297:519-22.

27. Boysen G, Nyboe J, Appleyard M, et al. Stroke incidence and risk factors for stroke in Copenhagen, Denmark. *Stroke* 1988; 19:1345-53.
28. Henderson BE, Paganini-Hill A, Ross RK. Decreased mortality in users of estrogen replacement therapy. *Arch Intern Med* 1991; 151:75-8.
29. Bush TL, Barrett-Connor E, Cowan LD, et al. Cardiovascular mortality and noncontraceptive use of estrogen in women: results from the Lipid Research Clinics Program Follow-up Study. *Circulation* 1987; 75:1102-9.
30. Petitti DB, Perlman JA, Sidney S. Noncontraceptive estrogens and mortality: long-term follow-up of women in the Walnut Creek Study. *Obstet Gynecol* 1987; 70:289-93.
31. Henderson BE, Ross RK, Paganini-Hill A, Mack TM. Estrogen use and cardiovascular disease. *Am J Obstet Gynecol* 1986; 154:1181-6.
32. Hillner BE, Hollenberg JP, Pauker SG. Postmenopausal estrogens in prevention of osteoporosis: benefit virtually without risk if cardiovascular effects are considered. *Am J Med* 1986; 80:1115-27.
33. Weinstein MC, Schiff I. Cost-effectiveness of hormone replacement therapy in the menopause. *Obstet Gynecol Surv* 1983; 38:445-55.
34. Colditz GA, Stampfer MJ, Willett WC, Hennekens CH, Rosner B, Speizer FE. Prospective study of estrogen replacement therapy and risk of breast cancer in postmenopausal women. *JAMA* 1990; 264:2648-53.

**Massachusetts Medical Society
Registry on Continuing Medical Education**

To obtain information on continuing medical education courses in the New England area, call between 9:00 a.m. and 12:00 noon, Monday through Friday, (617) 893-4610 or in Massachusetts 1-800-322-2303, ext. 1342. If writing, direct correspondence to: Program Registrar, Massachusetts Medical Society, 1440 Main St., Waltham, MA 02154-1649. The booklet is free to MMS members, \$5.00 for nonmembers.

Randomized Trial of Estrogen Plus Progestin for Secondary Prevention of Coronary Heart Disease in Postmenopausal Women

Stephen Hulley, MD; Deborah Grady, MD; Trudy Bush, PhD; Curt Furberg, MD, PhD;
David Herrington, MD; Betty Riggs, MD; Eric Vittinghoff, PhD;
for the Heart and Estrogen/progestin Replacement Study (HERS) Research Group

Context.—Observational studies have found lower rates of coronary heart disease (CHD) in postmenopausal women who take estrogen than in women who do not, but this potential benefit has not been confirmed in clinical trials.

Objective.—To determine if estrogen plus progestin therapy alters the risk for CHD events in postmenopausal women with established coronary disease.

Design.—Randomized, blinded, placebo-controlled secondary prevention trial.

Setting.—Outpatient and community settings at 20 US clinical centers.

Participants.—A total of 2763 women with coronary disease, younger than 80 years, and postmenopausal with an intact uterus. Mean age was 66.7 years.

Intervention.—Either 0.625 mg of conjugated equine estrogens plus 2.5 mg of medroxyprogesterone acetate in 1 tablet daily ($n = 1380$) or a placebo of identical appearance ($n = 1383$). Follow-up averaged 4.1 years; 82% of those assigned to hormone treatment were taking it at the end of 1 year, and 75% at the end of 3 years.

Main Outcome Measures.—The primary outcome was the occurrence of non-fatal myocardial infarction (MI) or CHD death. Secondary cardiovascular outcomes included coronary revascularization, unstable angina, congestive heart failure, resuscitated cardiac arrest, stroke or transient ischemic attack, and peripheral arterial disease. All-cause mortality was also considered.

Results.—Overall, there were no significant differences between groups in the primary outcome or in any of the secondary cardiovascular outcomes: 172 women in the hormone group and 176 women in the placebo group had MI or CHD death (relative hazard [RH], 0.99; 95% confidence interval [CI], 0.80-1.22). The lack of an overall effect occurred despite a net 11% lower low-density lipoprotein cholesterol level and 10% higher high-density lipoprotein cholesterol level in the hormone group compared with the placebo group (each $P < .001$). Within the overall null effect, there was a statistically significant time trend, with more CHD events in the hormone group than in the placebo group in year 1 and fewer in years 4 and 5. More women in the hormone group than in the placebo group experienced venous thromboembolic events (34 vs 12; RH, 2.89; 95% CI, 1.50-5.58) and gallbladder disease (84 vs 62; RH, 1.38; 95% CI, 1.00-1.92). There were no significant differences in several other end points for which power was limited, including fracture, cancer, and total mortality (131 vs 123 deaths; RH, 1.08; 95% CI, 0.84-1.38).

Conclusions.—During an average follow-up of 4.1 years, treatment with oral conjugated equine estrogen plus medroxyprogesterone acetate did not reduce the overall rate of CHD events in postmenopausal women with established coronary disease. The treatment did increase the rate of thromboembolic events and gallbladder disease. Based on the finding of no overall cardiovascular benefit and a pattern of early increase in risk of CHD events, we do not recommend starting this treatment for the purpose of secondary prevention of CHD. However, given the favorable pattern of CHD events after several years of therapy, it could be appropriate for women already receiving this treatment to continue.

MANY OBSERVATIONAL studies have found lower rates of coronary heart disease (CHD) in women who take postmenopausal estrogen than in women not receiving this therapy.¹⁻⁵ This association has been reported to be especially strong for secondary prevention in women with CHD, with hormone users having 35% to 80% fewer recurrent events than nonusers.⁶⁻¹² If this association is causal, estrogen therapy could be an

For editorial comment see p 650.

important method for preventing CHD in postmenopausal women. However, the observed association between estrogen therapy and reduced CHD risk might be attributable to selection bias if women who choose to take hormones are healthier and have a more favorable CHD profile than those who do not.¹³⁻¹⁵ Observational studies cannot resolve this uncertainty.

Only a randomized trial can establish the efficacy and safety of postmenopausal hormone therapy for preventing CHD.

From the University of California, San Francisco (Drs Hulley, Grady, and Vittinghoff); The Johns Hopkins University, Baltimore, Md (Dr Bush); Wake Forest University School of Medicine, Winston-Salem, NC (Drs Furberg and Herrington); and Wyeth-Ayerst Research, Radnor, Pa (Dr Riggs).

A complete list of the HERS Research Group participants appears at the end of this article.

HERS was funded by Wyeth-Ayerst Research. Dr Grady has been a consultant to Eli Lilly, and she and Dr Hulley receive research support from that company. Dr Bush has received honoraria and/or research support from Wyeth-Ayerst Research, Upjohn, Merck, Rhone-Poulenc Rorer, and Solvay. She is a board member for Women First HealthCare, Inc. Dr Furberg is a consultant to Wyeth-Ayerst Research. Dr Herrington has received research support and honoraria from Wyeth-Ayerst Research, Pfizer, and Eli Lilly. Dr Riggs is an employee of Wyeth-Ayerst Research.

Reprints: Stephen Hulley, MD, UCSF Box 0886, San Francisco, CA 94143.

JAMA. 1998;280:605-613

The Heart and Estrogen/progestin Replacement Study (HERS) was a randomized, double-blind, placebo-controlled trial of daily use of conjugated equine estrogens plus medroxyprogesterone acetate (progestin) on the combined rate of nonfatal myocardial infarction (MI) and CHD death among postmenopausal women with coronary disease. We enrolled women with established coronary disease because their high risk for CHD events and the strong reported association between hormone use and risk of these events make this an important and efficient study population in which to evaluate the effect of hormone therapy.

METHODS

Study Participants

The design, methods, and baseline findings of the study have been published.¹⁶ Briefly, participants were postmenopausal women younger than 80 years with established coronary disease who had not had a hysterectomy. Postmenopausal was defined as age at least 55 years and no natural menses for at least 5 years, or no natural menses for at least 1 year and serum follicle-stimulating hormone (FSH) level more than 40 IU/L, or documented bilateral oophorectomy, or reported bilateral oophorectomy with FSH level more than 40 IU/L and estradiol level less than 92 pmol/L (25 pg/mL). Established coronary disease was defined as evidence of 1 or more of the following: MI, coronary artery bypass graft surgery, percutaneous coronary revascularization, or angiographic evidence of at least a 50% occlusion of 1 or more major coronary arteries.

Women were excluded for the following reasons: CHD event within 6 months of randomization; serum triglyceride level higher than 3.39 mmol/L (300 mg/dL); use of oral, parenteral, vaginal, or transdermal sex hormones within 3 months of the screening visit; history of deep vein thrombosis or pulmonary embolism; history of breast cancer or breast examination or mammogram suggestive of breast cancer; history of endometrial cancer; abnormal uterine bleeding, endometrial hyperplasia, or endometrium thickness greater than 5 mm on baseline evaluation; abnormal or unobtainable Papanicolaou test result; serum aspartate aminotransferase level more than 1.2 times normal; unlikely to remain geographically accessible for study visits for at least 4 years; disease (other than CHD) judged likely to be fatal within 4 years; New York Heart Association class IV or severe class III congestive heart failure; alcoholism or other drug abuse; uncontrolled hypertension (diastolic blood pressure ≥ 105 mm Hg or systolic blood pressure ≥ 200 mm Hg);

uncontrolled diabetes (fasting blood glucose level ≥ 16.7 mmol/L [300 mg/dL]); participation in another investigational drug or device study; less than 80% compliance with a placebo run-in prior to randomization; or history of intolerance to hormone therapy.

Baseline Measurements

At 2 baseline clinic visits we collected data on demographic characteristics, reproductive and health history, risk factors for CHD, quality of life, and medication use. Participants had a clinical examination, including breast examination and pelvic examination with Papanicolaou test and endometrial evaluation (endometrial aspiration biopsy if possible or otherwise transvaginal ultrasound measurement of endometrial thickness), and a screening mammogram. Standardized 12-lead electrocardiograms (ECGs) were obtained using the Mac PC (Marquette Electronics, Milwaukee, Wis) and transmitted electronically to EPICARE (Wake Forest University School of Medicine, Winston-Salem, NC) where they were analyzed using computer protocols.¹⁷ Fasting total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, and triglyceride levels were determined by the Lipoprotein Analytical Laboratory at Johns Hopkins Hospital.¹⁸

Randomization and Blinding

The randomization code was prepared using computer-generated random numbers. Eligible participants were assigned with equal probability to the 2 treatment groups using tamper-proof blocked randomization stratified by clinical center. At each center, women who met the entry criteria were logged and assigned the next available sequential randomized treatment assignment.

Study medication consisted of 1 tablet daily containing both conjugated equine estrogens, 0.625 mg, and medroxyprogesterone acetate, 2.5 mg (estrogen plus progestin [Prempro]), or 1 placebo tablet of identical appearance. Chemical analysis of tablets confirmed the composition of the tablets and the accuracy of the blinded medication assignment.

With the exception of 3 persons at the Coordinating Center at the University of California, San Francisco, who prepared analyses for the Data and Safety Monitoring Board and for the final report, investigators and staff at the clinical centers, Wyeth-Ayerst Research, the Coordinating Center, and the independent Morbidity and Mortality Subcommittee were blinded to individual participant assignment throughout the study. To prevent unblinding of clinical center staff, breast discomfort

and vaginal bleeding were reported directly to gynecology staff who were located separate from the clinical center, did not communicate with clinical center personnel about gynecologic symptoms, and did not participate in ascertainment of cardiovascular outcomes. Sealed treatment allocation envelopes were available to the study center gynecologist. To determine if endometrial biopsy was necessary, the gynecologists could open a treatment assignment envelope in limited, defined situations with prior approval of a Coordinating Center physician. Unblinding in this fashion, generally to assist in the management of persistent vaginal bleeding, occurred in 34 women (30 in the hormone group, among whom 1 primary CHD event occurred).

Follow-up

Follow-up visits to the clinical center occurred every 4 months to assess and enhance compliance, provide study medication refills, and obtain outcome and adverse event data. Annual evaluations at the clinical center included general and cardiac examinations, an ECG, and venipuncture at the first, third, and final annual visits. Separate annual follow-up visits to the study gynecologist included repeat breast examination, pelvic examination with Papanicolaou test, screening mammogram, and a repeat endometrial evaluation at the second and final annual visits.

We used extensive quality assurance procedures for clinical management and data collection. All procedures were defined by the Coordinating Center in the HERS procedure manual, with formalized updates and clarifications. The Coordinating Center monitored the degree to which procedures at the clinics conformed with those described in the procedure manual during annual site visits. All data were entered twice and checked by computer algorithms.

Study treatment was discontinued (but follow-up continued) for women who developed any of the following conditions: simple endometrial hyperplasia without atypia that did not respond to treatment with progestin; endometrial hyperplasia with atypia; endometrial, cervical, breast, or ovarian cancer; deep vein thrombosis; pulmonary embolism; prolonged immobilization; or active gallbladder disease.

Outcome Ascertainment

The CHD events (nonfatal MI or CHD death) that occurred between the date of randomization and the closeout date were the primary outcome of the trial; nonfatal MI could be either symptomatic or silent, and CHD death could be a fatal documented MI, sudden death within 1 hour of onset of symptoms, un-

observed death that occurred out of the hospital in the absence of other known cause, or death due to coronary revascularization procedure or congestive heart failure. The diagnosis of nonfatal MI was based on an algorithm¹⁶ that took into account 3 categories of clinical information from the acute event: ischemic symptoms, ECG abnormalities, and elevated cardiac enzyme levels. The diagnosis could also be made if there was evidence of fresh MI at autopsy. All ECGs obtained electronically were compared with the ECG obtained at baseline for changes indicating new MI.

Secondary cardiovascular outcomes included coronary artery bypass graft surgery, percutaneous coronary revascularization, hospitalization for unstable angina, resuscitated cardiac arrest, congestive heart failure, stroke or transient ischemic attack, and peripheral arterial disease. Other prespecified secondary outcomes were total mortality; cancer death; non-CHD, noncancer death; breast, endometrial, and other cancer; deep vein thrombosis; pulmonary embolism; hip and other fracture; and gallbladder disease.¹⁶

The primary and secondary outcomes of HERS were addressed at each follow-up contact. Suspected outcome events were reported within 24 hours to the Coordinating Center, which had primary responsibility for the outcome database, and to Wyeth-Ayerst Research as a cross-check. Clinics obtained and sent to the Coordinating Center specified documentation that included (depending on the suspected event) hospital discharge summaries, ECGs, cardiac enzyme levels and other test results, and reports of tissue pathology, procedures, and x-ray examinations. Data from all deaths and suspected primary outcome events were reviewed and classified according to prespecified criteria by an independent Morbidity and Mortality Subcommittee blinded to treatment assignment. Secondary events were classified by Coordinating Center physicians blinded to treatment assignment. Every event (whether primary or secondary) was classified independently by 2 reviewers, and discordant classifications were resolved in discussions between the reviewers. Problematic potential primary events were discussed on conference calls or meetings involving the entire subcommittee.

Vital status is known for all 2763 women, and all deaths are included in this report. We are still in the process of collecting hospital records and adjudicating recent events. Included in this report are 99% of all primary CHD events reported to have occurred by the closeout visit (April-July 1998) and 97% of all secondary events. Adjudication is final

for 96% of included primary events (the remaining classifications are provisional), and it is final for 99% of included secondary events.

Statistical Power and Analyses

We estimated that we needed to enroll 2340 women, assuming a primary CHD event rate in the placebo group of 5% per year, a combined non-CHD death and loss to follow-up rate of 2% per year, crossovers from active to placebo of 5%, 4%, and 3% in the first 3 years and 2% per year thereafter, crossovers from placebo to active of 1% each year, and average follow-up of 4.75 years.¹⁶ We assumed that half the reduction in primary CHD events would operate through nonlipid mechanisms (and therefore be immediate), and half would operate through lipid changes (and therefore begin after a 2-year lag period). These assumptions resulted in 90% power at a 2-tailed α of .05 to detect an intention-to-treat effect size of 24%. In the actual study, the event rate was only 3.3%, compliance was less than expected, and treatment duration averaged 4.1 years. The chief reason for the shorter-than-expected treatment duration, despite ending the study at the planned time, was the fact that most women were enrolled toward the end of the recruitment period. The reduction in power caused by these deviations from prestudy assumptions was partially offset by the fact that we recruited 18% more participants than planned.

The primary analysis compares the rate of CHD events among women assigned to active medication with the rate among women assigned to placebo using an unadjusted Cox proportional hazards model for time to first CHD event; this is equivalent to the log rank test. The analysis was by intention to treat, categorizing participants according to randomized treatment assignment regardless of compliance. Participants who asked to drop out of the study and had not had a nonfatal MI were censored for nonfatal events at their last visit (this occurred for 31 women in the hormone group and 38 women in the placebo group); however, vital status was assessed at the end of the trial for 100% of the cohort, and all deaths are included in this report.

Secondary analyses used multivariate proportional hazards models to investigate study findings. Possible confounding was examined by controlling for important baseline covariates. To identify potential postrandomization confounders, treatment effect estimates were compared in nested models with and without measures of postrandomization lipid-lowering drug use and lipid change. These covariates were also included in an as-treated model, where inclusion in

the risk sets was limited to women in both treatment groups whose average pill-count compliance since randomization was at least 80%; this model included 74% of the primary events. Relative hazards were estimated by year since randomization (censoring women with events in earlier years), and continuous trend in the log relative hazard was examined in a companion model. Time-dependent indicators were used to assess risk by treatment assignment among women who had recently stopped taking study medication.

Data and Safety Monitoring Board

Interim monitoring of study events every 3 to 6 months was performed by an independent HERS Data and Safety Monitoring Board. Early in the trial the board noted adverse trends in primary CHD events, which conflicted with existing evidence and did not cross the stopping boundaries.¹⁹ In the middle years of the trial, an increased risk of venous thromboembolic events in the hormone-treated group consistent with existing evidence did cross the stopping boundaries. As a consequence, the board advised HERS investigators to report the findings regarding increased risk of venous thrombosis and to institute additional measures to reduce risk in HERS participants.²⁰ Near the end of the trial, the board noted a trend toward lower rates of nonfatal MI in the hormone group. At its final meeting in December 1997, the board recommended against continuing the study beyond the scheduled closeout date, because at that time conditional power estimates for primary CHD events were low and because of uncertainty about whether a sufficient proportion of women would consent to continue blinded treatment. The board recommended closeout at the originally planned time (April-July 1998), continuation of disease event surveillance, and rapid publication of the findings to allow HERS participants to make timely informed decisions concerning their use of this specific hormone therapy.

RESULTS

Between January 1993 and September 1994, the 20 HERS clinical centers enrolled 2763 women; 1380 were assigned to the hormone group and 1383 to the placebo group (Figure 1). Participants ranged in age from 44 to 79 years, with a mean of 66.7 years (SD, 6.7 years) at baseline. Most participants were white (89%) and had completed high school (80%). Examination of the distribution of these and other variables revealed no significant differences between the treatment groups at baseline (Table 1).

Table 1.—Baseline Characteristics of HERS Participants (n=2763) by Treatment Group*

Characteristic	Treatment Group		P Value
	Estrogen-Progestin (n=1380)	Placebo (n=1383)	
Demographics			
Age, mean±SD, y	67±7	67±7	.32
White, %	88	90	.14
Education, mean±SD, y	13±3	13±3	.84
CHD risk factors			
Current smoker, %	13	13	.84
Diabetes on oral medication or insulin, %	19	18	.44
Systolic blood pressure, mean±SD, mm Hg	135±19	135±19	.88
Diastolic blood pressure, mean±SD, mm Hg	73±10	73±10	.89
LDL cholesterol, mean±SD, mmol/L (mg/dL)	3.75±0.96 (145±37)	3.75±0.98 (145±38)	.83
HDL cholesterol, mean±SD, mmol/L (mg/dL)	1.29±0.34 (50±13)	1.29±0.34 (50±13)	.41
Triglyceride, mean±SD, mmol/L (mg/dL)	1.90±0.72 (168±64)	1.86±0.72 (165±64)	.25
Time since last menstrual period, mean ± SD, y	18±8	18±8	.31
Body mass index >27 kg/m ² , %	57	55	.44
Exercise >3 times weekly, %	39	38	.72
No. of drinks per week, mean±SD	1.4±4	1.3±4	.83
General health poor or fair, %	24	24	.94
Postmenopausal estrogen use, %†	24	23	.43
CHD manifestations			
Signs of congestive heart failure, %‡	10	9	.38
Q-wave myocardial infarction, %	17	17	.94
Percutaneous coronary revascularization, %	45	45	.96
Coronary artery bypass graft surgery, %	42	41	.64
Medication use			
Aspirin, %	78	78	.73
β-Blockers, %	33	32	.72
Lipid-lowering medications, %	45	47	.26
Calcium channel blockers, %	55	55	.83
Angiotensin-converting enzyme inhibitors, %	17	18	.57
Diuretics, %	28	28	.79
Multivitamins, %	29	30	.45

*HERS indicates Heart and Estrogen/progestin Replacement Study; CHD, coronary heart disease; LDL, low-density lipoprotein; and HDL, high-density lipoprotein. P values are for difference between treatment groups by t test or χ^2 .

†Estrogen use refers to use after menopause but not within 3 months of HERS screening.

‡Presence of jugular venous distention more than 8 cm H₂O, S₃ heart sound, rales, or pitting peripheral edema.

At the end of the first year, the proportion who reported taking study medication was 82% in the hormone group and 91% in the placebo group; by the end of the third year, these proportions had declined to 75% and 81%. Pill counts revealed 79% of the women in the hormone group to be taking at least 80% of their study medication at the end of year 1 and 70% to be doing so at the end of year 3 (Figure 2). Among women who stopped taking HERS medications, 110 (8%) of those assigned to the placebo group and 36 (3%) of those assigned to the hormone group reported taking open-label oral or transdermal estrogen.

During the closeout period (April-July 1998), vital status was ascertained for all 2763 randomized women. Follow-up percentages were nearly the same in the 2 randomized groups (Figure 1).

Primary CHD Outcome

Primary CHD events occurred in 172 women in the hormone group (33.1/1000 women per year) and in 176 women in the placebo group (33.6/1000 women per

year) (relative hazard [RH], 0.99; 95% confidence interval [CI], 0.80-1.22) (Table 2). These primary events were composed of CHD deaths (RH, 1.24; 95% CI, 0.87-1.75) and nonfatal MIs (RH, 0.91; 95% CI, 0.71-1.17). None of these differences was statistically significant. The 71 CHD deaths in the hormone group and the 58 CHD deaths in the placebo group were distributed, respectively, as follows: sudden death within 1 hour of onset of symptoms, 19 and 20; myocardial infarction, 19 and 16; congestive heart failure, 9 and 6; coronary artery bypass graft surgery, 5 and 2; and other CHD death, 19 and 14.

Survival curves for the primary CHD outcome and its components (Figure 3) correspond with the findings in Table 2. The curves for CHD death diverged during the second year of observation. The curves for nonfatal MI diverged during the first year, then converged and crossed during the third year. This possible change in the RH with time since randomization is further examined in Table 3. The point estimates for the primary outcome

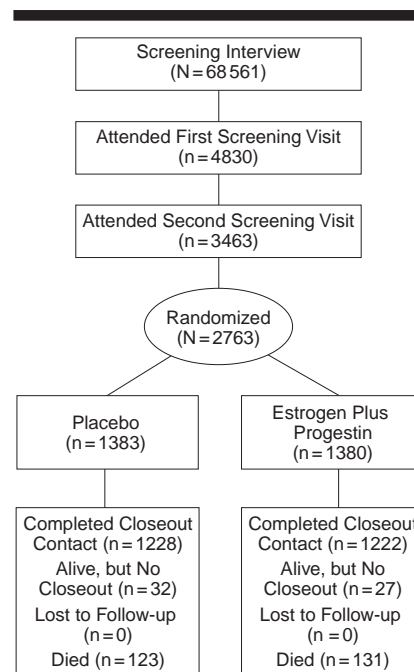


Figure 1.—The Heart and Estrogen/progestin Replacement Study trial profile, showing numbers of participants from screening to closeout.

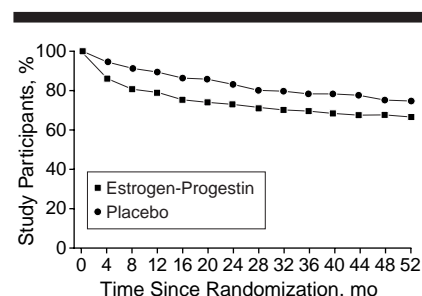


Figure 2.—Participants taking protocol medications and with pill count of 80% or more, as a percentage of all women at risk for a primary coronary heart disease event.

in the hormone group compared with the placebo group are 1.52 in year 1, 1.00 in year 2, 0.87 in year 3, and 0.67 in years 4 and 5 ($P = .009$ for trend in log RH); within the first year, the RH was 2.30 for the first 4 months, 1.46 for the second 4 months, and 1.18 for the third 4 months ($P = .33$ for trend). The difference over time was most pronounced for the nonfatal MI component of the primary CHD outcome (Table 3 and Figure 3).

In an as-treated analysis limited to women who had been at least 80% compliant with study medication by pill count, the RH comparing the primary CHD outcome in the hormone and placebo groups was 0.87 (95% CI, 0.67-1.11), lower than the intention-to-treat analysis but not statistically significant. For women who stopped taking HERS medication, risk of primary CHD events was elevated in the

first month after stopping use of the medication. However, there was no difference by group (RH in hormone group, 7.28; 95% CI, 4.45-11.93; RH in placebo group, 7.40; 95% CI, 4.23-12.95), suggesting that illness caused both the discontinuation of medication and the CHD event.

The RH comparing risk of the primary CHD outcome in the hormone and placebo groups was similar after adjusting for the small and nonsignificant differences between the groups in age and other baseline CHD risk factors (RH, 0.95; 95% CI, 0.76-1.17). We sought to identify differential effects of estrogen plus progestin therapy in women classified by baseline variables such as older age, ill health, history of MI, and so forth. There was no clear evidence of differential effects in 86 subgroups categorized by all the variables presented in Table 1 and others.

Other Cardiovascular Outcomes

There were no statistically significant differences between the randomized

groups in any of the other cardiovascular outcomes that we evaluated (Table 2). The survival curve for time to first occurrence of any coronary revascularization procedure or hospitalization for definite unstable angina (Figure 4) appeared to diverge, with lower rates in the hormone-treated group, although this difference did not achieve statistical significance (RH, 0.89; $P = .15$).

Plasma Lipids

By the end of the first year of treatment, mean LDL cholesterol levels had decreased by 14% from baseline to a level of 3.23 mmol/L (125 mg/dL) in the hormone group and by 3% to 3.62 mmol/L (140 mg/dL) in the placebo group ($P < .001$ for difference between groups) (Figure 5). During the same period, mean HDL cholesterol levels had increased by 8% to 1.40 mmol/L (54 mg/dL) in the hormone group and decreased by 2% to 1.27 mmol/L (49 mg/dL) in the placebo group ($p < .001$). Mean triglycer-

ide levels had increased by 10% to 2.04 mmol/L (181 mg/dL) in the hormone group and by 2% to 1.93 (170 mg/dL) in the placebo group ($P < .001$).

In proportional hazards analysis, high LDL cholesterol and low HDL cholesterol levels at baseline predicted subsequent primary CHD events in both univariate and multivariate (controlling for other baseline risk factors) models, but high triglyceride levels predicted primary CHD events only in univariate analyses. Changes in LDL cholesterol, HDL cholesterol, and triglyceride levels over the first year of the study were not significantly associated with subsequent primary CHD events, but the point estimates were in the expected direction and there was limited power to examine this effect.

More women in the placebo group than in the hormone group began treatment with lipid-lowering drugs, primarily statins, during the trial (22% vs 18%; $P = .004$), probably because the higher LDL cholesterol levels in placebo-treated women compared with hormone-treated women were noted by the women's personal physicians. Adjustment for this difference using regression analysis did not substantially change the overall estimate of the between-group difference in risk of primary CHD events (RH, 0.94; 95% CI, 0.76-1.17).

Other Secondary Outcomes

Cancer deaths and other deaths were nearly identical in the 2 study groups. Total mortality in the hormone group was not significantly different from that in the placebo group (131 vs 123 women; RH, 1.08; 95% CI, 0.84-1.38) (Table 4; Figure 6).

Confirmed venous thromboembolic events occurred in 34 women in the hormone group (6.3/1000 woman-years) and

Table 2.—Cardiovascular Outcomes by Treatment Group*

Outcomes	Treatment Group		RH (95% CI)	P Value
	Estrogen-Progestin (n=1380)	Placebo (n=1383)		
Primary CHD events†	172	176	0.99 (0.80-1.22)	.91
CHD death	71	58	1.24 (0.87-1.75)	.23
Nonfatal MI	116	129	0.91 (0.71-1.17)	.46
Other cardiovascular outcomes				
Coronary artery bypass graft surgery	88	101	0.87 (0.66-1.16)	.36
Percutaneous coronary revascularization	164	175	0.95 (0.77-1.17)	.62
Hospitalization for unstable angina	103	117	0.89 (0.68-1.16)	.38
Hospitalization for congestive heart failure	128	112	1.07 (0.84-1.38)	.58
Resuscitated cardiac arrest	19	13	1.48 (0.73-3.00)	.28
Other CHD event	3	1	3.03 (0.32-29.1)	.34
Peripheral arterial disease	94	108	0.87 (0.66-1.15)	.34
Stroke/transient ischemic attack	108	96	1.13 (0.85-1.48)	.40

*RH indicates relative hazard; CI, confidence interval; CHD, coronary heart disease; and MI, myocardial infarction. Each row represents the number of women with the designated event; women with more than 1 type of event may appear in more than 1 row.

†Primary CHD events include coronary death and nonfatal MI. Among the 245 nonfatal MIs, there were 7 silent MIs, found on annual electrocardiogram. There were 26 women with nonfatal MI who subsequently suffered CHD death.

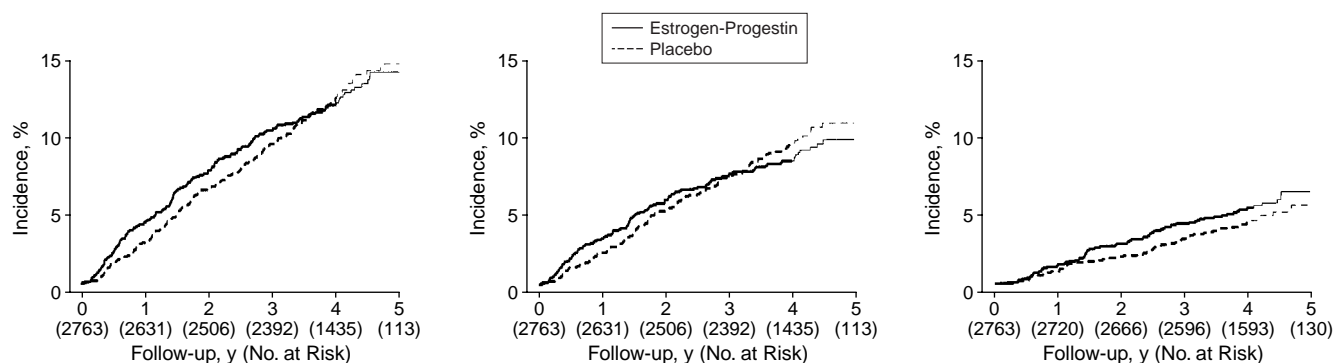


Figure 3.—Kaplan-Meier estimates of the cumulative incidence of primary coronary heart disease (CHD) events (left) and to its constituents: nonfatal myocardial infarction (MI) (center) and CHD death (right). The number of women observed at each year of follow-up and still free of an event are provided in parentheses, and the curves become fainter when this number drops below half of the cohort. Log rank P values are .91 for primary CHD events, .46 for nonfatal MI, and .23 for CHD death.

Table 3.—Outcomes by Treatment Group and Year Since Randomization*

Outcome and Period	Estrogen-Progestin		Placebo		RH (95% CI)	P Value‡
	No.	Rate†	No.	Rate†		
Primary CHD event§						
Year 1	57	42.5	38	28.0	1.52 (1.01-2.29)	.009
Year 2	47	37.0	48	37.1	1.00 (0.67-1.49)	
Year 3	35	28.8	41	33.1	0.87 (0.55-1.37)	
Years 4 and 5	33	23.0	49	34.4	0.67 (0.43-1.04)	
Nonfatal myocardial infarction						
Year 1	42	31.3	29	21.4	1.47 (0.91-2.36)	.01
Year 2	34	26.8	37	28.6	0.94 (0.59-1.49)	
Year 3	20	16.5	29	23.4	0.70 (0.40-1.24)	
Years 4 and 5	20	13.9	34	23.9	0.58 (0.34-1.02)	
CHD death						
Year 1	17	12.5	11	8.0	1.56 (0.73-3.32)	.34
Year 2	19	14.4	13	9.7	1.48 (0.73-2.99)	
Year 3	18	14.0	16	12.3	1.14 (0.58-2.24)	
Years 4 and 5	17	11.0	18	11.6	0.95 (0.49-1.84)	
Unstable angina or coronary revascularization¶						
Year 1	101	77.1	94	71.1	1.08 (0.82-1.44)	.42
Year 2	52	43.3	85	70.6	0.61 (0.43-0.87)	
Year 3	69	61.9	56	50.5	1.22 (0.86-1.74)	
Years 4 and 5	47	36.6	67	54.2	0.67 (0.46-0.98)	
Venous thromboembolic event						
Year 1	13	9.6	4	2.9	3.29 (1.07-10.08)	.28
Year 2	8	6.1	2	1.5	4.09 (0.87-19.27)	
Year 3	7	5.5	3	2.3	2.40 (0.62-9.28)	
Years 4 and 5	6	4.0	3	2.0	2.05 (0.15-8.18)	

*RH indicates relative hazard; CI, confidence interval; and CHD, coronary heart disease.

†Event rates per 1000 women-years in the estrogen plus progestin or placebo group.

‡P values for tests of continuous trend in log-relative hazard.

§Primary CHD events include nonfatal myocardial infarction and CHD death.

¶Coronary revascularization includes coronary artery bypass graft surgery and percutaneous coronary revascularization.

in 12 women in the placebo group (2.2/1000 woman-years) (RH, 2.89; 95% CI, 1.50-5.58; $P = .002$) (Table 4). More women in the hormone group experienced deep vein thromboses (25 vs 8; $P = .004$) and pulmonary emboli (11 vs 4; $P = .08$); 2 of the pulmonary emboli, both in the hormone group, were fatal. The RH in the hormone group relative to the placebo group remained elevated over the 4 years of observation but declined somewhat during the study (Table 3).

Gallbladder disease occurred in 84 women in the hormone group and in 62 women in the placebo group (RH, 1.38; 95% CI, 1.00-1.92). Gallbladder surgery accounted for 89% of these events, and the rest were symptomatic cholelithiasis. None of the gallbladder events was fatal.

There were no significant differences between the treatment groups in the rates of breast cancer, endometrial cancer, other cancers, or fracture (Table 4).

COMMENT

In this clinical trial, postmenopausal women younger than 80 years with established coronary disease who received estrogen plus progestin did not experience a reduction in overall risk of nonfatal MI and CHD death or of other car-

diovascular outcomes. How can this finding be reconciled with the large body of evidence from observational and pathophysiologic studies suggesting that estrogen therapy reduces risk for CHD?

Contrast With Findings of Observational Studies

Observational studies may be misleading because women who take postmenopausal hormones tend to have a better CHD risk profile^{13,21,22} and to obtain more preventive care¹⁴ than nonusers. The consistency of the apparent benefit in the observational studies could simply be attributable to the consistency of this selection bias. The lower rate of CHD in hormone users compared with nonusers persists after statistical adjustment for differences in CHD risk factors,²² but differences in unmeasured factors remain a possible explanation.

The discrepancy between the findings of HERS and the observational studies may also reflect important differences between the study populations and treatments. Most of the observational studies of postmenopausal hormone therapy enrolled postmenopausal women who were relatively young and healthy and who took unopposed estrogen.^{1-3,23} In contrast, participants in

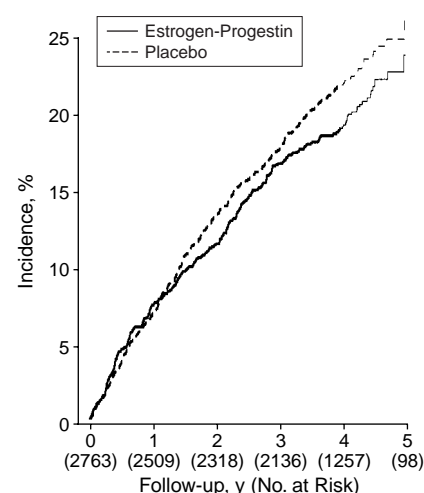


Figure 4.—Kaplan-Meier estimate of the cumulative incidence of definite unstable angina or coronary artery bypass graft or percutaneous coronary revascularization. The number of women observed at each year of follow-up and still free of an event are provided in parentheses, and the curves become fainter when this number drops below half the cohort. Log rank P value is .15.

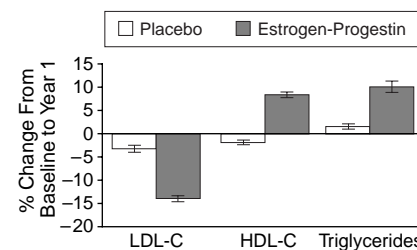


Figure 5.—Mean change in low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglyceride levels during the first year of the study, expressed as percent change \pm SEM.

HERS were older, had coronary disease at the outset, and were treated with estrogen plus progestin. However, some observational studies did examine women with prior CHD, and all of these reported a beneficial association with postmenopausal hormone therapy.⁶⁻¹² Similarly, some observational studies did examine the effect of postmenopausal estrogen plus progestin therapy on CHD risk in women, and these generally report a lower rate of CHD events in hormone users that is similar to that reported for estrogen alone^{4,5,22,24-27}; however, details in these studies about the specific progestin formulations and dosing regimens used are limited.

Possible Adverse Effects of Medroxyprogesterone Acetate

Several potential mechanisms whereby estrogen therapy might reduce risk for CHD have been proposed, including fa-

Table 4.—Death and Secondary Noncardiovascular Outcomes by Treatment Group*

Outcomes	Treatment Group		RH (95% CI)	P Value
	Estrogen-Progestin (n=1380)	Placebo (n=1383)		
Death				
CHD death	71	58	1.24 (0.87-1.75)	.23
Cancer death	19	24	0.80 (0.44-1.46)	.47
Non-CHD, noncancer death	37	36	1.04 (0.66-1.64)	.87
Unadjudicated death	4	5
Total deaths	131	123	1.08 (0.84-1.38)	.56
Venous thromboembolic event				
Deep vein thrombosis	25	8	3.18 (1.43-7.04)	.004
Pulmonary embolism	11	4	2.79 (0.89-8.75)	.08
Any thromboembolic event	34	12	2.89 (1.50-5.58)	.002
Cancer				
Breast	32	25	1.30 (0.77-2.19)	.33
Endometrial	2	4	0.49 (0.09-2.68)	.41
Other	63	58	1.10 (0.77-1.57)	.60
Any cancer	96	87	1.12 (0.84-1.50)	.44
Fracture				
Hip	12	11	1.10 (0.49-2.50)	.82
Other	119	129	0.93 (0.73-1.20)	.59
Any fracture	130	138	0.95 (0.75-1.21)	.70
Gallbladder disease	84	62	1.38 (1.00-1.92)	.05

*RH indicates relative hazard; CI, confidence interval; and CHD, coronary heart disease. Each row represents the number of women with the designated event; women with more than 1 type of event may appear in more than 1 row.

avorable effects on lipoproteins, coronary atherosclerosis, endothelial function, and arterial thrombosis.^{28,29} Progestins down-regulate estrogen receptors and may also have direct, progestin receptor-mediated effects that oppose these actions of estrogen³⁰; medroxyprogesterone acetate may do this to a greater extent than other progestins. In the Postmenopausal Estrogen-Progestin Interventions Trial, medroxyprogesterone acetate blunted the estrogen-associated increase in HDL cholesterol substantially more than did micronized progesterone.³¹ Oral medroxyprogesterone acetate appears to significantly attenuate the beneficial effects of estrogen on coronary atherosclerosis in nonhuman primates,³² while subcutaneous progesterone does not.³³ Animal data also suggest that medroxyprogesterone acetate may inhibit the beneficial effects of estrogen on endothelial-dependent vasodilation,³⁴ but this has not been documented in women.³⁵ Despite these mechanistic data suggesting an adverse effect of medroxyprogesterone acetate, observational studies show a similar reduction in CHD risk in women using medroxyprogesterone acetate plus estrogen as in women taking unopposed estrogen.⁴

Possible Differences in the Effects of Therapy Over Time

When the results were examined by year since randomization, the estrogen plus progestin regimen appeared to increase risk for primary CHD events in the first year of therapy but to decrease risk in subsequent years. This time trend

should be interpreted with caution. It could simply represent random variation, although the level of statistical significance makes this unlikely. More importantly, between-group contrasts that exclude the first several years are not true randomized comparisons, as the remaining study groups may no longer be comparable if, for example, treatment has caused high-risk individuals to have events early in the study.

On the other hand, the time trend is biologically plausible. The early increase in risk for CHD events might be attributable to an immediate prothrombotic, proarrhythmic, or proischemic effect of treatment that is gradually outweighed by a beneficial effect on the underlying progression of atherosclerosis, perhaps as a result of beneficial changes in lipoproteins. In trials of lipid interventions, the delay before CHD risk is reduced has ranged from 0 to 2 years.³⁶⁻⁴¹ After a lag period, the 11% net reduction in LDL cholesterol and 10% net increase in HDL cholesterol observed in the hormone group would be expected to reduce the risk of CHD events^{36,42} and may account for the trend toward a late benefit observed in HERS.

A pattern of early harm and later benefit could account for part of the discrepancy between the results of this trial and observational studies of estrogen and CHD. Attrition of susceptible individuals soon after starting estrogen replacement could increase the prevalence of survivors available for inclusion in observational studies; most observational studies are not designed to observe the

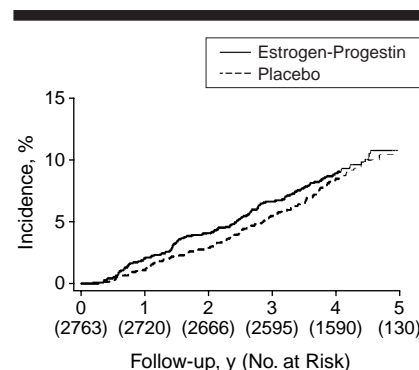


Figure 6.—Kaplan-Meier estimate of cumulative incidence of death from any cause. The number of women observed at each year of follow-up and still free of an event are provided in parentheses, and the curves become fainter when this number drops below half of the cohort. Log rank P value is .56.

onset of therapy or to detect an early adverse effect.

Previous Clinical Trial Evidence

The CHD data from previous hormone trials in women have been summarized⁴³ but are of limited value because the studies were small, short term, and not designed to examine CHD as an outcome. The only large prior trial of estrogen therapy to prevent CHD events was the Coronary Drug Project, which studied very high doses of estrogen (5.0 mg or 2.5 mg of conjugated equine estrogen daily) in men with preexisting CHD. The estrogen arms of this trial were stopped early because of an excess of MIs, thromboembolic events, and estrogenic symptoms in the 5.0-mg/d group⁴⁴ and the lack of benefit on the CHD end point and estrogenic symptoms in the 2.5-mg/d group.⁴⁵ The relevance of this trial of high-dose estrogen in men to postmenopausal hormone therapy in women is uncertain.

Safety and Other Noncardiovascular Outcomes

Venous thromboembolic events were 3 times more common in the hormone group than in the placebo group. Recent observational studies have reported similar relative risks for idiopathic venous thromboembolism among users of both unopposed estrogen⁴⁶⁻⁴⁹ and estrogen plus progestin therapy.^{47,49} The excess incidence of venous thrombotic events in HERS was 4.1 per 1000 woman-years of observation, an order of magnitude higher than the excess reported in the observational studies; the higher rate is probably a consequence of the facts that women enrolled in HERS were older and had multiple risk factors for venous thrombosis and that only idiopathic events were counted in the observational studies.

We found an increased risk of gallbladder disease in the hormone group that is

likely attributable to the estrogen therapy. Metabolic studies indicate that estrogen enhances hepatic lipoprotein uptake and inhibits bile acid synthesis, resulting in increased biliary cholesterol and cholelithiasis.⁵⁰

Observational studies have suggested that therapy with postmenopausal estrogen for 5 years or less is not associated with an increased risk of breast cancer but that longer duration of therapy might be associated with a small increase in risk.⁵¹ The HERS trial was not large enough and therapy did not continue for long enough to address this issue.

The incidence of fractures in the hormone group was only slightly lower than in the placebo group. Wide CIs around the fracture risk estimates reveal inadequate statistical power and do not exclude a reduction in risk of hip fracture of as much as 51% or a reduction in risk of other fracture of as much as 27%.

Strengths and Limitations of the Trial

The CHD risk factor profile of women enrolled in HERS is similar to that of a random sample of US women with probable heart disease, suggesting that the findings of HERS may be generalized to that population.⁵² However, HERS did not evaluate the effect of estrogen plus progestin therapy in women without CHD, and it is not known whether our findings apply to healthy women. It is also not known whether use of a different progestin or of estrogen alone would have been beneficial.

HERS exceeded the recruitment goal by 18%, carried out a successful randomization, collected objective, blindly adjudicated disease outcome data, and achieved 100% vital status ascertainment. Compliance with hormone treatment, while lower than projected, was sufficient to produce LDL and HDL cholesterol changes that compare favorably with previous studies.³¹ The 95% CIs for the effect of treatment assignment on primary CHD events (RH, 0.99; 95% CI, 0.80-1.22) make it unlikely that HERS missed a benefit of more than 20% for the overall 4.1-year period of observation. However, this statistic does not address the possible late benefit of treatment suggested by the time trend analysis, which is plausible based on the finding of a 1- to 2-year lag period observed in lipid trials³⁶⁻⁴¹; a longer study would be more definitive for investigating this possibility.

Future Directions

HERS is the first large trial of the effect of postmenopausal estrogen plus progestin therapy on risk for CHD events. The findings differ from those of observational studies and studies with surrogate outcomes, emphasizing the

importance of basing treatment policies on randomized controlled trials.⁵³ Other randomized trials of postmenopausal hormone therapy are likely to answer some of the questions raised by HERS. The Women's Health Initiative Randomized Trial⁵⁴ includes a group of women who have undergone hysterectomy and receive unopposed estrogen as well as women with intact uterus who receive the same estrogen plus progestin regimen used in HERS. Participants are not required to have CHD and are generally younger than the HERS cohort. The Women's Health Initiative Randomized Trial plans to enroll 27 500 women and to report the results in 2005 after 9 years of treatment. Further information will also emerge from HERS as we continue disease event surveillance.

Several interventions have been proven to reduce risk for CHD events in patients with coronary disease, including aspirin, β -blockers, lipid lowering, and smoking cessation.⁵⁵ The need for encouraging these interventions for women with coronary disease is illustrated by the facts that 90% of the HERS cohort had LDL cholesterol exceeding 2.59 mmol/L (100 mg/dL) at baseline and that only 32% were receiving β -blockers.

Conclusions

First, in the population studied in HERS, ie, postmenopausal women with established coronary disease and an average age of 66.7 years, daily use of conjugated equine estrogens and medroxyprogesterone acetate did not reduce the overall risk for MI and CHD death or any other cardiovascular outcome during an average of 4.1 years of follow-up. This therapy did increase the risk of venous thromboembolic events and gallbladder disease.

Second, we did not evaluate the cardiovascular effect of treatment with unopposed estrogen, commonly used in women who have had a hysterectomy, or other estrogen plus progestin formulations. We also did not study women without coronary disease.

Third, based on the finding of no overall cardiovascular benefit and a pattern of early increase in risk of CHD events, we do not recommend starting this treatment for the purpose of secondary prevention of CHD. However, given the favorable pattern of CHD events after several years of therapy, it could be appropriate for women already receiving hormone treatment to continue. Extended follow-up of the HERS cohort and additional randomized trials are needed to clarify the cardiovascular effects of postmenopausal hormone therapy.

Manuscript Reviewers: Donald Hunninghake, MD, Robert Califf, MD.

Statistical Analyst: Josephine Fong, MS.

Clinical Centers: *Emory University, Atlanta, Ga:* Nanette K. Wenger, MD, Sally McNagny, MD, MPH, Janice Parrott, RN, Dana Law-McKenzie, RN, Terri Selik, RN, Julia C. Gathe, MD; *The Johns Hopkins University, Baltimore, Md:* Trudy L. Bush, PhD, MHS, Roger Blumenthal, MD, Katherine Bass, MD, MHS, Susan R. Miller, MPH, DSc, Elizabeth Elder, MS, Janice Huth; *University of Alabama, Birmingham:* William J. Rogers, MD, Vera A. Bittner, MD, R. Edward Varner, MD, Robert L. Holley, MD, Brian P. Gleason, MD, Anthony R. Fuisz, MD, Brian G. Boatman, MD, David M. Clark, MD, G. Robert McDaniel, MD, Jerri Moody, RN, Brenda Vaughn, RN, Amanda Murphy, BSN, RN, Glenda Blackburn, LPN; *Chicago Center for Clinical Research, Chicago, Ill:* Michael Davidson, MD, Phyllis Marx, MD, Ann Maki, MS, RD, Marlene Wentworth, RNP; *Duke University Medical Center, Durham, NC:* Robert Califf, MD, L. Kristin Newby, MD, Jeff Andrews, MD, Donna Bowen, ANP, Diana Shenkin, MSN, FNP, Karen Philbrick, FNP, Lois Rittenhouse, BSN; *Hartford Hospital, Hartford, Conn:* David Waters, MD, Linda Chaffkin, MD, Judith Mace, Jennifer DeDominicis, RN, Marilyn Siwy, RN, JoAnn Pazdar, LPN; *Baylor College of Medicine, Houston, Tex:* J. Alan Herd, MD, Michael Snabes, MD, PhD, Melissa Kulkarni, RN, Diane Tanksley, RN, Ronald L. Young, MD, Harold Miller, MD, Amy Young, MD; *University of Iowa, Iowa City:* Helmut Schrott, MD, Ellen Gordon, MD, Dale Stovall, MD, Diane Meyerholz, RN, Pamela A. Terrill, ARNP, Regina Arthur, RN; *University of California, San Diego:* Elizabeth Barrett-Connor, MD, Cynthia Stuenkel, MD, Deborah Goodman-Gruen, MD, PhD, Sue Hawley, MSN, RN, Carol Anne Drastal, RN, MPH; *Cedars-Sinai Medical Center, Los Angeles, Calif:* Steven S. Khan, MD, Sanjay A. Agarwal, MD, T. Keta Hodgson, BSN, RN, Marina Chizhik; *University of Tennessee, Memphis:* William Applegate, MD, Karen C. Johnson, MD, Frank Ling, MD, Nancy Fowler, RN, Suzanne Sutterfield, MD, Nancy Miles, FNP, Beth McCammon, RN, Linda Sweeney, RN; *University of Miami School of Medicine, Miami, Fla:* Maureen Lowery, MD, Mike Janicek, MD, Ricardo Estape, MD, Jose A. Martel, MPH, Genevieve Dupuy, BSN, RN; *University of Minnesota, Heart Disease Prevention Clinic, Minneapolis:* Donald Hunninghake, MD, June LaValleur, MD, Jean Olson, RN, Sue Krook, PhD, Dawn Remme; *University of Pittsburgh, Pittsburgh, Pa (2 sites):* Jane Cauley, PhD, Alan Gradman, MD, Robert McDonald, Jr, MD, Michele Boyd, RN, Karen Southwick, CCRC, Melinda Everson, RN, MPH, Diana Stapleton, RN; *Northwest Lipid Research Clinic, Seattle, Wash:* Robert Knopp, MD, Jeanne Butler, RN, Kimberly K. Butler, Barbara J. Twaddell, RN, Jane Lau, MD, Leslie Miller, MD, Brenda L. Buck, RD; *Stanford Center for Research in Disease Prevention, Palo Alto, Calif:* William Haskell, PhD, Marcia Stefanick, PhD, Kathy Berra, ANP, RN, Linda Giudice, MD, PhD, Sarah French, CNP; *George Washington University, Washington, DC:* Valery Miller, MD, Judith Hsia, MD, Vanessa Barnabei, MD, PhD, Ginny Levin, MPH, Donna Embersit; *Wake Forest University School of Medicine, Winston-Salem, NC:* David Herrington, MD, MHS, Tom Snyder, MD, Penny Sharp, EdD, Marlene Baruch, RN, Karen Blinson, BS, Marcia Davis, RN, Judy Iannuzzi, RN, Vickie Wilson, RN, Lynda Doomy; *Wake Forest University School of Medicine, Greensboro, NC:* Jennifer Schaaf, MD, Kay Cheshire, MEd, Virginia Bradsher, MEd, Mary Boozer, LPN.

Central Lipid Laboratory: *Johns Hopkins University School of Medicine, Baltimore, Md:* Paul S. Bachorik, PhD, Kathleen L. Lovejoy.

Steering Committee Chairman: *Wake Forest University School of Medicine, Winston-Salem, NC:* Curt Furberg, MD, PhD.

Coordinating Center: *University of California, San Francisco:* Stephen B. Hulley, MD, MPH, Deborah Grady, MD, MPH, Steven R. Cummings, MD, Dennis Black, PhD, Rodger Shepherd, MD,

MPH, Eric Vittinghoff, PhD, Joel Simon, MD, MPH, Mark Hlatky, MD, Josephine Fong, MS, Christine C. Ireland, MPH, Christopher Gehrman.

Sponsor: Wyeth-Ayerst Research, Radnor, Pa: Betty Riggs, MD, Cynthia Richards, MD, James Pickar, MD, Jay Graepel, PhD, Kathleen Shields, MS, Judy Vittitoe, RN, MPH, Margaret Carr, RN, MBA, Tony Aranillo, MBA, Alease Winchester, RN, MS, Bernice Yost, MS.

Data and Safety Monitoring Board: O. Dale Williams, PhD, MPH, David DeMets, PhD, Genell Knatterud, PhD, Christine Cassel, MD, Judith Hochman, MD, Susan Johnson, MD, Neil Stone, MD.

Morbidity and Mortality Committee: Richard Conti, MD, David Faxon, MD, Nora F. Goldschlager, MD, Bruce Psaty, MD, PhD, W. Douglas Weaver, MD.

References

1. Bush TL. Noncontraceptive estrogen use and risk of cardiovascular disease: an overview and critique of the literature. In: Korenman SG, ed. *The Menopause: Biological and Clinical Consequences of Ovarian Failure: Evolution and Management*. Norwell, Mass: Serono Symposia; 1990:211-223.
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.
3. Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med*. 1992;117:1016-1037.
4. Psaty BM, Heckbert SR, Atkins D, et al. The risk of myocardial infarction associated with the combined use of estrogens and progestins in postmenopausal women. *Arch Intern Med*. 1994;154:1333-1339.
5. Sidney S, Petitti DB, Quesenberry CP Jr. Myocardial infarction and the use of estrogen and estrogen-progestogen in postmenopausal women. *Ann Intern Med*. 1997;127:501-508.
6. Bush TL, Barrett-Connor E, Cowan LD, et al. Cardiovascular mortality and noncontraceptive use of estrogen in women: results from the Lipid Research Clinics Program Follow-up Study. *Circulation*. 1987;75:1102-1109.
7. Sullivan JM, Vander Zwaag R, Hughes JP, et al. Estrogen replacement and coronary artery disease. *Arch Intern Med*. 1990;150:2557-2562.
8. Henderson BE, Paganini-Hill A, Ross RK. Decreased mortality in users of estrogen replacement therapy. *Arch Intern Med*. 1991;151:75-78.
9. O'Brien JE, Peterson ED, Keeler GP, et al. Relation between estrogen replacement therapy and stenosis after percutaneous coronary interventions. *J Am Coll Cardiol*. 1996;28:1111-1118.
10. Newton KM, LaCroix AZ, McKnight B, et al. Estrogen replacement therapy and prognosis after first myocardial infarction. *Am J Epidemiol*. 1997;145:269-277.
11. Sullivan JM, El-Zeky F, Vander Zwaag R, Ramanathan KB. Effect on survival of estrogen replacement therapy after coronary artery bypass grafting. *Am J Cardiol*. 1997;79:847-850.
12. O'Keefe JH, Kim SC, Hall RR, Cochran VC, Lawhorn SL, McCallister BD. Estrogen replacement therapy after coronary angioplasty in women. *J Am Coll Cardiol*. 1997;29:1-5.
13. Cauley JA, Cummings SR, Black DM, Mascioli SR, Seeley DG. Prevalence and determinants of estrogen replacement therapy in elderly women. *Am J Obstet Gynecol*. 1990;163:1438-1444.
14. Barrett-Connor E. Postmenopausal estrogen and prevention bias. *Ann Intern Med*. 1991;115:455-456.
15. Petitti DB. Coronary heart disease and estrogen replacement therapy: can compliance bias explain the results of observational studies? *Ann Epidemiol*. 1994;4:115-118.
16. Grady D, Applegate W, Bush T, et al. Heart and Estrogen/progestin Replacement Study (HERS):

design, methods and baseline characteristics. *Control Clin Trials*. 1998;19:314-335.

17. Rautaharju PM, Calhoun HP, Chaitman BR. Novacode serial ECG classification system for clinical trials and epidemiological studies. *J Electrocardiol*. 1992;34(suppl):179-187.
18. Schrott HG, Bittner V, Vittinghoff E, Herrington DM, Hulley SB. Adherence to National Cholesterol Education Program treatment goals in postmenopausal women with heart disease. *JAMA*. 1997;277:1281-1286.
19. Lan KKG, DeMets DL, Halperin M. More flexible sequential and non-sequential designs in long-term clinical trials. *Commun Stat Theory Method*. 1984;13:2339-2353.
20. Grady D, Hulley SB, Furberg C. Venous thromboembolic events associated with hormone replacement therapy. *JAMA*. 1997;278:477.
21. Matthews KA, Kuller LH, Wing RR, Meilahn EN, Plantinga P. Prior to use of estrogen replacement therapy, are users healthier than nonusers? *Am J Epidemiol*. 1996;143:971-978.
22. Grodstein F, Stampfer MJ, Manson JE, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. *N Engl J Med*. 1996;335:453-461.
23. Barrett-Connor E, Grady D. Hormone replacement therapy, heart disease, and other considerations. *Annu Rev Public Health*. 1998;19:55-72.
24. Falkeborn M, Persson I, Adami H, et al. The risk of acute myocardial infarction after oestrogen and oestrogen-progestogen replacement. *Br J Obstet Gynaecol*. 1992;99:821-828.
25. Thompson SG, Meade TW, Greenberg G. The use of hormonal replacement therapy and the risk of stroke and myocardial infarction in women. *J Epidemiol Community Health*. 1989;43:173-178.
26. Rosenberg L, Palmer JR, Shapiro S. A case-control study of myocardial infarction in relation to use of estrogen supplements. *Am J Epidemiol*. 1993;137:54-63.
27. Mann RD, Lis Y, Chukwujindu J, Chanter DO. A study of the association between hormone replacement therapy, smoking and the occurrence of myocardial infarction in women. *J Clin Epidemiol*. 1994;47:307-312.
28. Samaan SA, Crawford MH. Estrogen and cardiovascular function after menopause. *J Am Coll Cardiol*. 1995;26:1403-1410.
29. Guetta V, Cannon RO. Cardiovascular effects of estrogen and lipid-lowering therapies in postmenopausal women. *Circulation*. 1996;93:1928-1937.
30. Sarrel PM. How progestins compromise the cardioprotective effects of estrogens. *Menopause*. 1995;2:187-190.
31. Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women [published correction appears in *JAMA*. 1995;274:1676]. *JAMA*. 1995;273:199-208.
32. Adams MR, Register TC, Golden DL, Wagner JD, Williams JK. Medroxyprogesterone acetate antagonizes inhibitory effects of conjugated equine estrogens on coronary artery atherosclerosis. *Arterioscler Thromb Vasc Biol*. 1997;17:217-221.
33. Adams MR, Kaplan JR, Manuck SB, et al. Inhibition of coronary artery atherosclerosis by 17-beta estradiol in ovariectomized monkeys: lack of an effect of added progesterone. *Arteriosclerosis*. 1990;10:1051-1057.
34. Williams JK, Honoré EK, Washburn SA, Clarkson TB. Effects of hormone replacement therapy on reactivity of atherosclerotic coronary arteries in cynomolgus monkeys. *J Am Coll Cardiol*. 1994;24:1757-1761.
35. McCrohon JA, Adams MR, McCredie RJ, et al. Hormone replacement therapy is associated with improved arterial physiology in healthy postmenopausal women. *Clin Endocrinol*. 1996;45:435-441.
36. Lipid Research Clinics Program. The Lipid Research Clinics Coronary Primary Prevention Trial results, II: the relationship of reduction in incidence

of coronary heart disease to cholesterol lowering. *JAMA*. 1984;251:365-374.

37. Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*. 1995;333:1301-1307.
38. Sacks FM, Pfeffer MA, Moye LA, et al. The effect of pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. *N Engl J Med*. 1996;335:1001-1009.
39. Downs JR, Clearfield M, Weis S, et al. Primary prevention of acute coronary events with lovastatin in men and women with average cholesterol levels: results of AFCAPS/TexCAPS. *JAMA*. 1998;279:1615-1622.
40. Frick MH, Elo O, Haapa K, et al. Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia. *N Engl J Med*. 1987;317:1237-1245.
41. Buchwald H, Varco RL, Matts JP, et al. Effect of partial ileal bypass surgery on mortality and morbidity from coronary heart disease in patients with hypercholesterolemia. *N Engl J Med*. 1990;323:946-955.
42. Rosenenson RS, Tangney CC. Antiatherothrombotic properties of statins. *JAMA*. 1998;279:1643-1650.
43. Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *BMJ*. 1997;315:149-153.
44. Coronary Drug Project Research Group. The Coronary Drug Project: initial findings leading to modifications of its research protocol. *JAMA*. 1970;214:1303-1313.
45. Coronary Drug Project Research Group. Findings leading to discontinuation of the 2.5-mg/day estrogen group. *JAMA*. 1973;226:652-657.
46. Jick H, Derby LE, Myers MW, Vasilakis C, Newton KM. Risk of hospital admission for idiopathic venous thromboembolism among users of postmenopausal oestrogens. *Lancet*. 1996;348:981-983.
47. Daly E, Vessey MP, Hawkins MM, Carson JL, Gough P, Marsh S. Risk of venous thromboembolism in users of hormone replacement therapy. *Lancet*. 1996;348:977-980.
48. Grodstein F, Stampfer MJ, Goldhaber SZ, et al. Prospective study of exogenous hormones and risk of pulmonary embolism in women. *Lancet*. 1996;348:983-987.
49. Gutthann SP, Rodriguez LAG, Castellsague J, Oliart AD. Hormone replacement therapy and risk of venous thromboembolism: population based case-control study. *BMJ*. 1997;314:796-800.
50. Everson GT, McKinley C, Kern F Jr. Mechanisms of gallstone formation in women. *J Clin Invest*. 1991;87:237-246.
51. 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.
52. Herrington D, Fong J, Sempos CT, et al. Comparison of the HERS cohort with women with coronary disease from the NHANES III. *Am Heart J*. 1998;136:115-124.
53. Rossouw JE. Estrogens for prevention of coronary heart disease: putting the brakes on the bandwagon. *Circulation*. 1996;94:2982-2985.
54. Women's Health Initiative Study Group. Design of the Women's Health Initiative Clinical Trial and Observational Study. *Control Clin Trials*. 1998;19:61-109.
55. Ryan TJ, Anderson JL, Antman EM, et al. ACC/AHA Guidelines for the Management of Patients With Acute Myocardial Infarction: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (Committee on Management of Acute Myocardial Infarction). *J Am Coll Cardiol*. 1996;28:1328-1428.

Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women

Principal Results From the Women's Health Initiative Randomized Controlled Trial

Writing Group for the
Women's Health Initiative
Investigators

THE WOMEN'S HEALTH INITIATIVE (WHI) focuses on defining the risks and benefits of strategies that could potentially reduce the incidence of heart disease, breast and colorectal cancer, and fractures in postmenopausal women. Between 1993 and 1998, the WHI enrolled 161 809 postmenopausal women in the age range of 50 to 79 years into a set of clinical trials (trials of low-fat dietary pattern, calcium and vitamin D supplementation, and 2 trials of postmenopausal hormone use) and an observational study at 40 clinical centers in the United States.¹ This article reports principal results for the trial of combined estrogen and progestin in women with a uterus. The trial was stopped early based on health risks that exceeded health benefits over an average follow-up of 5.2 years. A parallel trial of estrogen alone in women who have had a hysterectomy is being continued, and the planned end of this trial is March 2005, by which time the average follow-up will be about 8.5 years.

The WHI clinical trials were designed in 1991-1992 using the accumulated evidence at that time. The primary outcome for the trial of estrogen plus progestin was designated as coronary heart disease (CHD). Potential cardioprotection was based on generally

Context Despite decades of accumulated observational evidence, the balance of risks and benefits for hormone use in healthy postmenopausal women remains uncertain.

Objective To assess the major health benefits and risks of the most commonly used combined hormone preparation in the United States.

Design Estrogen plus progestin component of the Women's Health Initiative, a randomized controlled primary prevention trial (planned duration, 8.5 years) in which 16 608 postmenopausal women aged 50-79 years with an intact uterus at baseline were recruited by 40 US clinical centers in 1993-1998.

Interventions Participants received conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d, in 1 tablet (n=8506) or placebo (n=8102).

Main Outcomes Measures The primary outcome was coronary heart disease (CHD) (nonfatal myocardial infarction and CHD death), with invasive breast cancer as the primary adverse outcome. A global index summarizing the balance of risks and benefits included the 2 primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

Results On May 31, 2002, after a mean of 5.2 years of follow-up, the data and safety monitoring board recommended stopping the trial of estrogen plus progestin vs placebo because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits. This report includes data on the major clinical outcomes through April 30, 2002. Estimated hazard ratios (HRs) (nominal 95% confidence intervals [CIs]) were as follows: CHD, 1.29 (1.02-1.63) with 286 cases; breast cancer, 1.26 (1.00-1.59) with 290 cases; stroke, 1.41 (1.07-1.85) with 212 cases; PE, 2.13 (1.39-3.25) with 101 cases; colorectal cancer, 0.63 (0.43-0.92) with 112 cases; endometrial cancer, 0.83 (0.47-1.47) with 47 cases; hip fracture, 0.66 (0.45-0.98) with 106 cases; and death due to other causes, 0.92 (0.74-1.14) with 331 cases. Corresponding HRs (nominal 95% CIs) for composite outcomes were 1.22 (1.09-1.36) for total cardiovascular disease (arterial and venous disease), 1.03 (0.90-1.17) for total cancer, 0.76 (0.69-0.85) for combined fractures, 0.98 (0.82-1.18) for total mortality, and 1.15 (1.03-1.28) for the global index. Absolute excess risks per 10 000 person-years attributable to estrogen plus progestin were 7 more CHD events, 8 more strokes, 8 more PEs, and 8 more invasive breast cancers, while absolute risk reductions per 10 000 person-years were 6 fewer colorectal cancers and 5 fewer hip fractures. The absolute excess risk of events included in the global index was 19 per 10 000 person-years.

Conclusions Overall health risks exceeded benefits from use of combined estrogen plus progestin for an average 5.2-year follow-up among healthy postmenopausal US women. All-cause mortality was not affected during the trial. The risk-benefit profile found in this trial is not consistent with the requirements for a viable intervention for primary prevention of chronic diseases, and the results indicate that this regimen should not be initiated or continued for primary prevention of CHD.

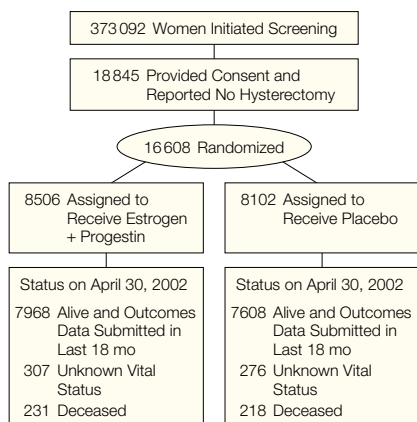
JAMA. 2002;288:321-333

www.jama.com

For editorial comment see p 366.

Author Information and Financial Disclosures appear at the end of this article.

Figure 1. Profile of the Estrogen Plus Progestin Component of the Women's Health Initiative



supportive data on lipid levels in intermediate outcome clinical trials, trials in nonhuman primates, and a large body of observational studies suggesting a 40% to 50% reduction in risk among users of either estrogen alone or, less frequently, combined estrogen and progestin.²⁻⁵ Hip fracture was designated as a secondary outcome, supported by observational data as well as clinical trials showing benefit for bone mineral density.^{6,7} Invasive breast cancer was designated as a primary adverse outcome based on observational data.^{3,8} Additional clinical outcomes chosen as secondary outcomes that may plausibly be affected by hormone therapy include other cardiovascular diseases; endometrial, colorectal, and other cancers; and other fractures.^{3,6,9}

The effect of hormones on overall health was an important consideration in the design and conduct of the WHI clinical trial. In an attempt to summarize important aspects of health benefits vs risks, a global index was defined as the earliest occurrence of CHD, invasive breast cancer, stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, or death due to other causes. Compared with total mortality, which may be too insensitive, this index assigns additional weight to the 7 listed diseases. Procedures for monitoring the trial involved semiannual comparisons of the

estrogen plus progestin and placebo groups with respect to each of the elements of the global index and to the overall global index.

This report pertains primarily to estrogen plus progestin use among healthy postmenopausal women, since only 7.7% of participating women reported having had prior cardiovascular disease. During the course of the WHI trial, the Heart and Estrogen/progestin Replacement Study (HERS) reported its principal results.¹⁰ HERS was another blinded, randomized controlled trial comparing the same regimen of estrogen plus progestin with placebo among women with a uterus; however, in HERS, all 2763 participating women had documented CHD prior to randomization. The HERS findings of no overall effect on CHD but an apparent increased risk in the first year after randomization seemed surprising given preceding observational studies of hormone use in women with CHD.³ Subsequent to HERS, some investigators reanalyzed their observational study data and were able to detect an early elevation in CHD risk among women with prior CHD¹¹⁻¹³ but not in ostensibly healthy women,¹⁴ prompting speculation that any early adverse effect of hormones on CHD incidence was confined to women who have experienced prior CHD events.

The WHI is the first randomized trial to directly address whether estrogen plus progestin has a favorable or unfavorable effect on CHD incidence and on overall risks and benefits in predominantly healthy women.

METHODS

Study Population

Detailed eligibility criteria and recruitment methods have been published.¹ Briefly, most women were recruited by population-based direct mailing campaigns to age-eligible women, in conjunction with media awareness programs. Eligibility was defined as age 50 to 79 years at initial screening, postmenopausal, likelihood of residence in the area for 3 years, and provision of written informed consent. A woman was

considered postmenopausal if she had experienced no vaginal bleeding for 6 months (12 months for 50- to 54-year-olds), had had a hysterectomy, or had ever used postmenopausal hormones. Major exclusions were related to competing risks (any medical condition likely to be associated with a predicted survival of <3 years), safety (eg, prior breast cancer, other prior cancer within the last 10 years except nonmelanoma skin cancer, low hematocrit or platelet counts), and adherence and retention concerns (eg, alcoholism, dementia).

A 3-month washout period was required before baseline evaluation of women using postmenopausal hormones at initial screening. Women with an intact uterus at initial screening were eligible for the trial of combined postmenopausal hormones, while women with a prior hysterectomy were eligible for the trial of unopposed estrogen. This report is limited to the 16 608 women with an intact uterus at baseline who were enrolled in the trial component of estrogen plus progestin vs placebo. The protocol and consent forms were approved by the institutional review boards for all participating institutions (see Acknowledgment).

Study Regimens, Randomization, and Blinding

Combined estrogen and progestin was provided in 1 daily tablet containing conjugated equine estrogen (CEE), 0.625 mg, and medroxyprogesterone acetate (MPA), 2.5 mg (Prempro, Wyeth Ayerst, Philadelphia, Pa). A matching placebo was provided to the control group. Eligible women were randomly assigned to receive estrogen plus progestin or placebo after eligibility was established and baseline assessments made (FIGURE 1). The randomization procedure was developed at the WHI Clinical Coordinating Center and implemented locally through a distributed study database, using a randomized permuted block algorithm, stratified by clinical center site and age group. All study medication bottles had a unique bottle number and bar code to allow for blinded dispensing.

Initially, the design allowed women with a uterus to be randomized to receive unopposed estrogen, estrogen plus progestin, or placebo. After the release of the Postmenopausal Estrogen/Progestin Intervention (PEPI) trial results¹⁵ indicating that long-term adherence to unopposed estrogen was not feasible in women with a uterus, the WHI protocol was changed to randomize women with a uterus to only estrogen plus progestin or placebo in equal proportions. The 331 women previously randomized to unopposed estrogen were unblinded and reassigned to estrogen plus progestin. These women are included in the estrogen plus progestin group in this report, resulting in 8506 participants in the estrogen plus progestin group vs 8102 in the placebo group. Analysis of the data excluding the women randomized before this protocol change did not affect the results. Considerable effort was made to maintain blinding of other participants and clinic staff. When required for safety or symptom management, an unblinding officer provided the clinic gynecologist, who was not involved with study outcomes activities, with the treatment assignment.

Follow-up

Study participants were contacted by telephone 6 weeks after randomization to assess symptoms and reinforce adherence. Follow-up for clinical events occurred every 6 months, with annual in-clinic visits required. At each semi-annual contact, a standardized interview collected information on designated symptoms and safety concerns, and initial reports of outcome events were obtained using a self-administered questionnaire. Adherence to study interventions was assessed by weighing of returned bottles. The study protocol required annual mammograms and clinical breast examinations; study medications were withheld if safety procedures were not performed, but these participants continued to be followed up. Electrocardiograms were collected at baseline and at follow-up years 3 and 6.

Data Collection, Management, and Quality Assurance

All data were collected on standardized study forms by certified staff according to documented study procedures. Study data were entered into a local clinical center database developed and maintained by the Clinical Coordinating Center and provided to each site in the form of a local area network connected to the Clinical Coordinating Center through a wide area network. Data quality was ensured through standard data entry mechanisms, routine reporting and database checks, random chart audits, and routine site visits.

Maintenance/Discontinuation of Study Medications

During the trial, some flexibility of the dosages of both estrogen and progestin was allowed to manage symptoms such as breast tenderness and vaginal bleeding. Vaginal bleeding was managed according to an algorithm that accounted for the time since randomization, severity of the bleeding, treatment assignment, and endometrial histology. Women who had a hysterectomy after randomization for indications other than cancer were switched to unopposed estrogen or the corresponding placebo without unblinding. These women are included in the original randomization group for analyses.

Permanent discontinuation of study medication was required by protocol for women who developed breast cancer, endometrial pathologic state (hyperplasia not responsive to treatment, atypia, or cancer), deep vein thrombosis (DVT) or PE, malignant melanoma, meningioma, triglyceride level greater than 1000 mg/dL (11.3 mmol/L), or prescription of estrogen, testosterone, or selective estrogen-receptor modulators by their personal physician. Medications were temporarily discontinued in participants who had acute myocardial infarction (MI), stroke, fracture, or major injury involving hospitalization, surgery involving use of anesthesia, any illness resulting in immobilization for

more than 1 week, or any other severe illness in which hormone use is temporarily inappropriate.

Outcome Ascertainment

Cardiovascular Disease. Coronary heart disease was defined as acute MI requiring overnight hospitalization, silent MI determined from serial electrocardiograms (ECGs), or CHD death. The diagnosis of acute MI was established according to an algorithm adapted from standardized criteria¹⁶ that included cardiac pain, cardiac enzyme and troponin levels, and ECG readings. The primary analyses included both definite and probable MIs as defined by the algorithm. Myocardial infarction occurring during surgery and aborted MIs were included. An aborted MI was defined as chest pain and ECG evidence of acute MI at presentation, an intervention (eg, thrombolysis) followed by resolution of ECG changes, and all cardiac enzyme levels within normal ranges. Silent MI was diagnosed by comparing baseline and follow-up ECGs at 3 and 6 years after randomization. Coronary death was defined as death consistent with CHD as underlying cause plus 1 or more of the following: preterminal hospitalization with MI within 28 days of death, previous angina or MI and no potentially lethal noncoronary disease, death resulting from a procedure related to coronary artery disease, or death certificate consistent with CHD as the underlying cause. Stroke diagnosis was based on rapid onset of a neurologic deficit lasting more than 24 hours, supported by imaging studies when available. Pulmonary embolism and DVT required clinical symptoms supported by relevant diagnostic studies.

Cancer. Breast, colorectal, endometrial, and other cancers were confirmed by pathological reports when available. Current data indicate that at least 98% of breast, colorectal, and endometrial cancers and 92% of other cancers were documented with pathological reports.

Fractures. Reports of hip, vertebral, and other osteoporotic fractures (including all fractures except those of

Table 1. Baseline Characteristics of the Women's Health Initiative Estrogen Plus Progestin Trial Participants (N = 16 608) by Randomization Assignment*

Characteristics	Estrogen + Progestin (n = 8506)	Placebo (n = 8102)	P Value†
Age at screening, mean (SD), y	63.2 (7.1)	63.3 (7.1)	.39
Age group at screening, y			
50-59	2839 (33.4)	2683 (33.1)	.80
60-69	3853 (45.3)	3657 (45.1)	
70-79	1814 (21.3)	1762 (21.7)	
Race/ethnicity			
White	7140 (83.9)	6805 (84.0)	.33
Black	549 (6.5)	575 (7.1)	
Hispanic	472 (5.5)	416 (5.1)	
American Indian	26 (0.3)	30 (0.4)	
Asian/Pacific Islander	194 (2.3)	169 (2.1)	
Unknown	125 (1.5)	107 (1.3)	
Hormone use			
Never	6280 (73.9)	6024 (74.4)	.49
Past	1674 (19.7)	1588 (19.6)	
Current‡	548 (6.4)	487 (6.0)	
Duration of prior hormone use, y			
<5	1538 (69.1)	1467 (70.6)	.25
5-10	426 (19.1)	357 (17.2)	
≥10	262 (11.8)	253 (12.2)	
Body mass index, mean (SD), kg/m ² §	28.5 (5.8)	28.5 (5.9)	.66
Body mass index, kg/m ²			
<25	2579 (30.4)	2479 (30.8)	.89
25-29	2992 (35.3)	2834 (35.2)	
≥30	2899 (34.2)	2737 (34.0)	
Systolic BP, mean (SD), mm Hg	127.6 (17.6)	127.8 (17.5)	.51
Diastolic BP, mean (SD), mm Hg	75.6 (9.1)	75.8 (9.1)	.31
Smoking			
Never	4178 (49.6)	3999 (50.0)	.85
Past	3362 (39.9)	3157 (39.5)	
Current	880 (10.5)	838 (10.5)	
Parity			
Never pregnant/no term pregnancy	856 (10.1)	832 (10.3)	.67
≥1 term pregnancy	7609 (89.9)	7233 (89.7)	
Age at first birth, y			
<20	1122 (16.4)	1114 (17.4)	.11
20-29	4985 (73.0)	4685 (73.0)	
≥30	723 (10.6)	621 (9.7)	
Treated for diabetes	374 (4.4)	360 (4.4)	.88
Treated for hypertension or BP ≥140/90 mm Hg	3039 (35.7)	2949 (36.4)	.37
Elevated cholesterol levels requiring medication	944 (12.5)	962 (12.9)	.50
Statin use at baseline¶	590 (6.9)	548 (6.8)	.66
Aspirin use (≥80 mg/d) at baseline	1623 (19.1)	1631 (20.1)	.09
History of myocardial infarction	139 (1.6)	157 (1.9)	.14
History of angina	238 (2.8)	234 (2.9)	.73
History of CABG/PTCA	95 (1.1)	120 (1.5)	.04
History of stroke	61 (0.7)	77 (1.0)	.10
History of DVT or PE	79 (0.9)	62 (0.8)	.25
Female relative had breast cancer	1286 (16.0)	1175 (15.3)	.28
Fracture at age ≥55 y	1031 (13.5)	1029 (13.6)	.87

continued

the ribs, chest/sternum, skull/face, fingers, toes, and cervical vertebrae) were routinely ascertained. All fracture outcomes were verified by radiology reports. Study radiographs were not obtained to ascertain subclinical vertebral fractures.

This report is based on outcomes adjudicated by clinical center physician adjudicators, as used for trial-monitoring purposes. Clinical center physician adjudicators were centrally trained and blinded to treatment assignment and participants' symptoms. Future communications will report results based on centrally adjudicated outcomes and will include a broader range of outcomes with more extensive explanatory analyses. Since this report is presented before the planned study closeout, outcome information is still being collected and adjudicated. Local adjudication is complete for approximately 96% of the designated self-reported events. To date, agreement rates between local and central adjudication are: MI, 84%; revascularization procedures, 97%; PE, 89%; DVT, 84%; stroke, 94%; invasive breast cancer, 98%; endometrial cancer, 96%; colorectal cancer, 98%; hip fracture, 95%; and specific cause of death, 82%. When related cardiovascular conditions are combined (eg, when unstable angina or congestive heart failure is grouped with MI), agreement rates exceed 94% for cardiovascular disease and 90% for specific cause of death.

Statistical Analyses

All primary analyses use time-to-event methods and are based on the intention-to-treat principle. For a given outcome, the time of event was defined as the number of days from randomization to the first postrandomization diagnosis, as determined by the local adjudicator. For silent MIs, the date of the follow-up ECG applied. Participants without a diagnosis were censored for that event at the time of last follow-up contact. Primary outcome comparisons are presented as hazard ratios (HRs) and 95% confidence intervals (CIs) from Cox proportional haz-

ards analyses,¹⁷ stratified by clinical center, age, prior disease, and randomization status in the low-fat diet trial.

Two forms of CIs are presented, nominal and adjusted. Nominal 95% CIs describe the variability in the estimates that would arise from a simple trial for a single outcome. Although traditional, these CIs do not account for the multiple statistical testing issues (across time and across outcome categories) that occurred in this trial, so the probability is greater than .05 that at least 1 of these CIs will exclude unity under an overall null hypothesis. The adjusted 95% CIs presented herein use group sequential methods to correct for multiple analyses over time. A Bonferroni correction for 7 outcomes as specified in the monitoring plan (described herein) was applied to all clinical outcomes other than CHD and breast cancer, the designated primary and primary adverse effect outcomes, and the global index. The adjusted CIs are closely related to the monitoring procedures and, as such, represent a conservative assessment of the evidence. This report focuses primarily on results using the unadjusted statistics and also relies on consistency across diagnostic categories, supportive data from other studies, and biologic plausibility for interpretation of the findings.

Data and Safety Monitoring

Trial monitoring guidelines for early stopping considerations were based on O'Brien-Fleming boundaries¹⁸ using asymmetric upper and lower boundaries: a 1-sided, .025-level upper boundary for benefit and 1-sided, .05-level lower boundaries for adverse effects. The adverse-effect boundaries were further adjusted with a Bonferroni correction for the 7 major outcomes other than breast cancer that were specifically monitored (CHD, stroke, PE, colorectal cancer, endometrial cancer, hip fracture, and death due to other causes). The global index of monitored outcomes played a supportive role as a summary measure of the overall balance of risks and benefits. Trial monitoring for early stopping consider-

Table 1. Baseline Characteristics of the Women's Health Initiative Estrogen Plus Progestin Trial Participants (N = 16 608) by Randomization Assignment* (cont)

Characteristics	Estrogen + Progestin (n = 8506)	Placebo (n = 8102)	P Value
Gail model 5-year risk of breast cancer, %			
<1	1290 (15.2)	1271 (15.7)	.64
1-<2	5384 (63.3)	5139 (63.4)	
2-<5	1751 (20.6)	1621 (20.0)	
≥5	81 (1.0)	71 (0.9)	
No. of falls in last 12 mo			
0	5168 (66.2)	5172 (67.5)	.18
1	1643 (21.0)	1545 (20.2)	
2	651 (8.3)	645 (8.4)	
≥3	349 (4.5)	303 (4.0)	

*Data are presented as number (percentage) of patients unless otherwise noted. BP indicates blood pressure; CABG/PTCA, coronary artery bypass graft/percutaneous transluminal coronary angioplasty; DVT, deep vein thrombosis; and PE, pulmonary embolism.

†Based on χ^2 tests (categorical variables) or *t* tests (continuous variables).

‡Required a 3-month washout prior to randomization.

§Total number of participants with data available was 8470 for estrogen plus progestin and 8050 for placebo.

||Among women who reported having a term pregnancy.

¶Statins are 3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors.

ations was conducted semiannually by an independent data and safety monitoring board (DSMB). Aspects of the monitoring plan have been published.¹⁹

RESULTS

Trial Monitoring and Early Stopping

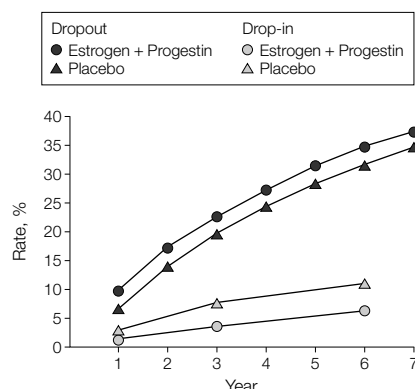
Formal monitoring began in the fall of 1997 with the expectation of final analysis in 2005 after an average of approximately 8.5 years of follow-up. Late in 1999, with 5 interim analyses completed, the DSMB observed small but consistent early adverse effects in cardiovascular outcomes and in the global index. None of the disease-specific boundaries had been crossed. In the spring of 2000 and again in the spring of 2001, at the direction of the DSMB, hormone trial participants were given information indicating that increases in MI, stroke, and PE/DVT had been observed and that the trial continued because the balance of risks and benefits remained uncertain.

In reviewing the data for the 10th interim analyses on May 31, 2002, the DSMB found that the adverse effects in cardiovascular diseases persisted, although these results were still within the monitoring boundaries. However, the design-specified weighted log-rank test statistic for breast cancer ($z = -3.19$)

crossed the designated boundary ($z = -2.32$) and the global index was supportive of a finding of overall harm ($z = -1.62$). Updated analyses including 2 months of additional data, available by the time of the meeting, did not appreciably change the overall results. On the basis of these data, the DSMB concluded that the evidence for breast cancer harm, along with evidence for some increase in CHD, stroke, and PE, outweighed the evidence of benefit for fractures and possible benefit for colon cancer over the average 5.2-year follow-up period. Therefore, the DSMB recommended early stopping of the estrogen plus progestin component of the trial. Because the balance of risks and benefits in the unopposed-estrogen component remains uncertain, the DSMB recommended continuation of that component of the WHI. Individual trial participants have been informed.

Baseline Characteristics

There were no substantive differences between study groups at baseline; 8506 women were randomized into the estrogen plus progestin group and 8102 into the placebo group (TABLE 1). The mean (SD) age was 63.3 (7.1) years. Two thirds of the women who reported prior or current hormone use had taken combined hormones and one third had used unopposed estrogen.

Figure 2. Cumulative Dropout and Drop-in Rates by Randomization Assignment and Follow-up Duration

Dropout refers to women who discontinued study medication; drop-in, women who discontinued study medication and received postmenopausal hormones through their own clinician.

Prevalence of prior cardiovascular disease was low and levels of cardiovascular risk factors were consistent with a generally healthy population of postmenopausal women. An assessment of commonly studied breast cancer risk factors, both individually and combined using the Gail model,²⁰ indicate that the cohort in general was not at increased risk of breast cancer.

Follow-up, Adherence, and Unblinding

Vital status is known for 16025 randomized participants (96.5%), including 449 (2.7%) known to be deceased. A total of 583 (3.5%) participants were lost to follow-up or stopped providing outcomes information for more than 18

months. The remaining 15576 (93.8%) provided recent outcome information (Figure 1).

At the time of this report, all women had been enrolled for at least 3.5 years, with an average follow-up of 5.2 years and a maximum of 8.5 years. A substantial number of women had stopped taking study drugs at some time (42% of estrogen plus progestin and 38% of placebo). Dropout rates over time (FIGURE 2) exceeded design projections, particularly early on, but compare favorably with community-based adherence to postmenopausal hormones.²¹ Some women in both groups initiated hormone use through their own clinician (6.2% in the estrogen plus progestin group and 10.7% in the placebo group cumulatively by the sixth

Table 2. Clinical Outcomes by Randomization Assignment*

Outcomes	No. of Patients (Annualized %)		Hazard Ratio	Nominal 95% CI	Adjusted 95% CI
	Estrogen + Progestin (n = 8506)	Placebo (n = 8102)			
Follow-up time, mean (SD), mo	62.2 (16.1)	61.2 (15.0)	NA	NA	NA
Cardiovascular disease†					
CHD	164 (0.37)	122 (0.30)	1.29	1.02-1.63	0.85-1.97
CHD death	33 (0.07)	26 (0.06)	1.18	0.70-1.97	0.47-2.98
Nonfatal MI	133 (0.30)	96 (0.23)	1.32	1.02-1.72	0.82-2.13
CABG/PTCA	183 (0.42)	171 (0.41)	1.04	0.84-1.28	0.71-1.51
Stroke	127 (0.29)	85 (0.21)	1.41	1.07-1.85	0.86-2.31
Fatal	16 (0.04)	13 (0.03)	1.20	0.58-2.50	0.32-4.49
Nonfatal	94 (0.21)	59 (0.14)	1.50	1.08-2.08	0.83-2.70
Venous thromboembolic disease	151 (0.34)	67 (0.16)	2.11	1.58-2.82	1.26-3.55
Deep vein thrombosis	115 (0.26)	52 (0.13)	2.07	1.49-2.87	1.14-3.74
Pulmonary embolism	70 (0.16)	31 (0.08)	2.13	1.39-3.25	0.99-4.56
Total cardiovascular disease	694 (1.57)	546 (1.32)	1.22	1.09-1.36	1.00-1.49
Cancer					
Invasive breast	166 (0.38)	124 (0.30)	1.26	1.00-1.59	0.83-1.92
Endometrial	22 (0.05)	25 (0.06)	0.83	0.47-1.47	0.29-2.32
Colorectal	45 (0.10)	67 (0.16)	0.63	0.43-0.92	0.32-1.24
Total	502 (1.14)	458 (1.11)	1.03	0.90-1.17	0.86-1.22
Fractures					
Hip	44 (0.10)	62 (0.15)	0.66	0.45-0.98	0.33-1.33
Vertebral	41 (0.09)	60 (0.15)	0.66	0.44-0.98	0.32-1.34
Other osteoporotic‡	579 (1.31)	701 (1.70)	0.77	0.69-0.86	0.63-0.94
Total	650 (1.47)	788 (1.91)	0.76	0.69-0.85	0.63-0.92
Death					
Due to other causes	165 (0.37)	166 (0.40)	0.92	0.74-1.14	0.62-1.35
Total	231 (0.52)	218 (0.53)	0.98	0.82-1.18	0.70-1.37
Global index§	751 (1.70)	623 (1.51)	1.15	1.03-1.28	0.95-1.39

*CI indicates confidence interval; NA, not applicable; CHD, coronary heart disease; MI, myocardial infarction; CABG, coronary artery bypass grafting; and PTCA, percutaneous transluminal coronary angioplasty.

†CHD includes acute MI requiring hospitalization, silent MI determined from serial electrocardiograms, and coronary death. There were 8 silent MIs. Total cardiovascular disease is limited to events during hospitalization except venous thromboembolic disease reported after January 1, 2000.

‡Other osteoporotic fractures include all fractures other than chest/sternum, skull/face, fingers, toes, and cervical vertebrae, as well as hip and vertebral fractures reported separately.

§The global index represents the first event for each participant from among the following types: CHD, stroke, pulmonary embolism, breast cancer, endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

year). These “drop-in” rates were also greater than expected.

At the time of this report, clinic gynecologists had been unblinded to treatment assignment for 3444 women in the estrogen plus progestin group and 548 women in the placebo group, primarily to manage persistent vaginal bleeding. During the trial, 248 women in the estrogen plus progestin group and 183 in the placebo group had a hysterectomy.

Intermediate Cardiovascular Disease End Points

Blood lipid levels, assessed in an 8.6% subsample of fasting blood specimens collected from women at baseline and year 1, showed greater reductions in low-density lipoprotein cholesterol (−12.7%) and increases in high-density lipoprotein cholesterol (7.3%) and triglycerides (6.9%) with estrogen plus progestin relative to placebo (data not shown), consistent with HERS and PEPI.^{10,22} Systolic blood pressure was, on average, 1.0 mm Hg higher in women taking estrogen plus progestin at 1 year, rising to 1.5 mm Hg at 2 years and beyond (data not shown). Diastolic blood pressures did not differ.

Clinical Outcomes

Cardiovascular Disease. Overall CHD rates were low (TABLE 2). The rate of women experiencing CHD events was increased by 29% for women taking estrogen plus progestin relative to placebo (37 vs 30 per 10000 person-years), reaching nominal statistical significance (at the .05 level). Most of the excess was in nonfatal MI. No significant differences were observed in CHD deaths or revascularization procedures (coronary artery bypass grafting or percutaneous transluminal coronary angioplasty). Stroke rates were also higher in women receiving estrogen plus progestin (41% increase; 29 vs 21 per 10000 person-years), with most of the elevation occurring in nonfatal events. Women in the estrogen plus progestin group had 2-fold greater rates of venous thromboembolism (VTE), as well as DVT and PE individually, with almost all associated CIs excluding 1.

Table 3. Cause of Death by Randomization Assignment

	No. (Annualized %)	
	Estrogen + Progestin (n = 8506)	Placebo (n = 8102)
Total deaths	231 (0.52)	218 (0.53)
Adjudicated deaths	215 (0.49)	201 (0.49)
Cardiovascular	65 (0.15)	55 (0.13)
Breast cancer	3 (0.01)	2 (<0.01)
Other cancer	104 (0.24)	86 (0.21)
Other known cause	34 (0.08)	41 (0.10)
Unknown cause	9 (0.02)	17 (0.04)

Rates of VTE were 34 and 16 per 10000 person-years in the estrogen plus progestin and placebo groups, respectively. Total cardiovascular disease, including other events requiring hospitalization, was increased by 22% in the estrogen plus progestin group.

Cancer. The invasive breast cancer rates in the placebo group were consistent with design expectations. The 26% increase (38 vs 30 per 10000 person-years) observed in the estrogen plus progestin group almost reached nominal statistical significance and, as noted herein, the weighted test statistic used for monitoring was highly significant. No significant difference was observed for in situ breast cancers. Follow-up rates for mammography were comparable in the estrogen plus progestin and placebo groups. Colorectal cancer rates were reduced by 37% (10 vs 16 per 10000 person-years), also reaching nominal statistical significance. Endometrial cancer incidence was not affected, nor was lung cancer incidence (54 vs 50; HR, 1.04; 95% CI, 0.71-1.53) or total cancer incidence.

Fractures. This cohort experienced low hip fracture rates (10 per 10000 person-years in the estrogen plus progestin group vs 15 per 10000 person-years in the placebo group). Estrogen plus progestin reduced the observed hip and clinical vertebral fracture rates by one third compared with placebo, both nominally significantly. The reductions in other osteoporotic fractures (23%) and total fractures (24%) were statistically significant (all associated CIs exclude 1).

The global index showed a nominally significant 15% increase in the es-

trogen plus progestin group (170 vs 151 per 10000 person-years). There were no differences in mortality or cause of death between groups (TABLE 3).

Time Trends

The Kaplan Meier estimates of cumulative hazards (FIGURE 3) for CHD indicate that the difference between treatment groups began to develop soon after randomization. These curves provide little evidence of convergence through 6 years of follow-up. The cumulative hazards for stroke begin to diverge between 1 and 2 years after randomization, and this difference persists beyond the fifth year. For PE, the curves separate soon after randomization and show continuing adverse effects throughout the observation period. For breast cancer, the cumulative hazard functions are comparable through the first 4 years, at which point the curve for estrogen plus progestin begins to rise more rapidly than that for placebo. Curves for colorectal cancer show benefit beginning at 3 years, and curves for hip fracture show increasing cumulative benefit over time. The difference in hazard rates for the global index (FIGURE 4) suggests a gradual increase in adverse effects compared with benefits for estrogen plus progestin through year 5, with a possible narrowing of the difference by year 6; however, HR estimates tend to be unstable beyond 6 years after randomization. Total mortality rates are indistinguishable between estrogen plus progestin and placebo.

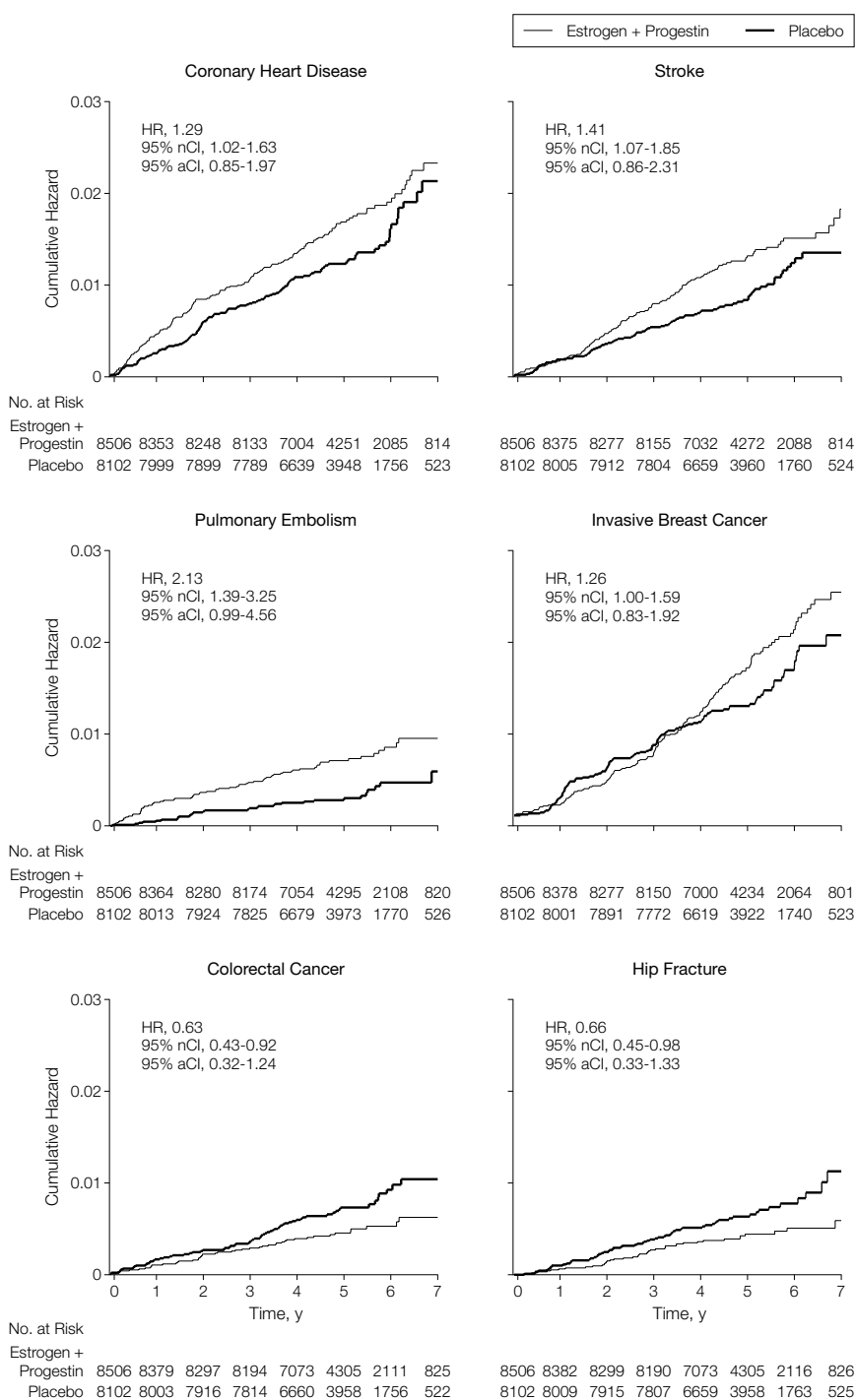
Tests for linear trends with time since randomization, based on a Cox proportional hazards model with a time-

dependent covariate, detected no trend with time for CHD, stroke, colorectal cancer, hip fracture, total mortality, or the global index (TABLE 4). There was

some evidence for an increasing risk of breast cancer over time with estrogen plus progestin ($z=2.56$ compared with a nominal z score for statistical signifi-

cance of 1.96) and a decreasing risk of VTE with time ($z=-2.45$). These results must be viewed cautiously because the number of events in each interval is modest, the data in later years are still incomplete, and later year comparisons are limited to women still at risk of their first event for that outcome.

Figure 3. Kaplan-Meier Estimates of Cumulative Hazards for Selected Clinical Outcomes



HR indicates hazard ratio; nCI, nominal confidence interval; and aCI, adjusted confidence interval.

Subgroup Analyses

Cardiovascular Disease. A small subset of women ($n=400$; average follow-up, 57.4 months) in WHI reported conditions at baseline that would have made them eligible for HERS, ie, prior MI or revascularization procedures. Among these women with established coronary disease, the HR for subsequent CHD for estrogen plus progestin relative to placebo was 1.28 (95% CI, 0.64-2.56) with 19 vs 16 events. The remaining women, those without prior CHD, had an identical HR for CHD (145 vs 106; HR, 1.28; 95% CI, 1.00-1.65). Few women with a history of VTE were enrolled, but these data suggest a possibility that these women may be at greater risk of future VTE events when taking estrogen plus progestin (7 vs 1; HR, 4.90; 95% CI, 0.58-41.06) than those without a history of VTE (144 vs 66; HR, 2.06; 95% CI, 1.54-2.76). For stroke, prior history did not confer additional risk (1 vs 5 in women with prior stroke; HR, 0.46; 95% CI, 0.05-4.51; 126 vs 80 with no prior stroke; HR, 1.47; 95% CI, 1.11-1.95). No noteworthy interactions with age, race/ethnicity, body mass index, prior hormone use, smoking status, blood pressure, diabetes, aspirin use, or statin use were found for the effect of estrogen plus progestin on CHD, stroke, or VTE.

Breast Cancer. Women reporting prior postmenopausal hormone use had higher HRs for breast cancer associated with estrogen plus progestin use than those who never used postmenopausal hormones (among never users, 114 vs 102; HR, 1.06; 95% CI, 0.81-1.38; for women with <5 years of prior use, 32 vs 15; HR, 2.13; 95% CI, 1.15-3.94; for women with 5-10

years of prior use, 11 vs 2; HR, 4.61; 95% CI, 1.01-21.02; and for women with ≥ 10 years of prior use, 9 vs 5; HR, 1.81; 95% CI, 0.60-5.43; test for trend, $z = 2.17$). No interactions between estrogen plus progestin and age, race/ethnicity, family history, parity, age at first birth, body mass index, or Gail-model risk score were observed for invasive breast cancer.

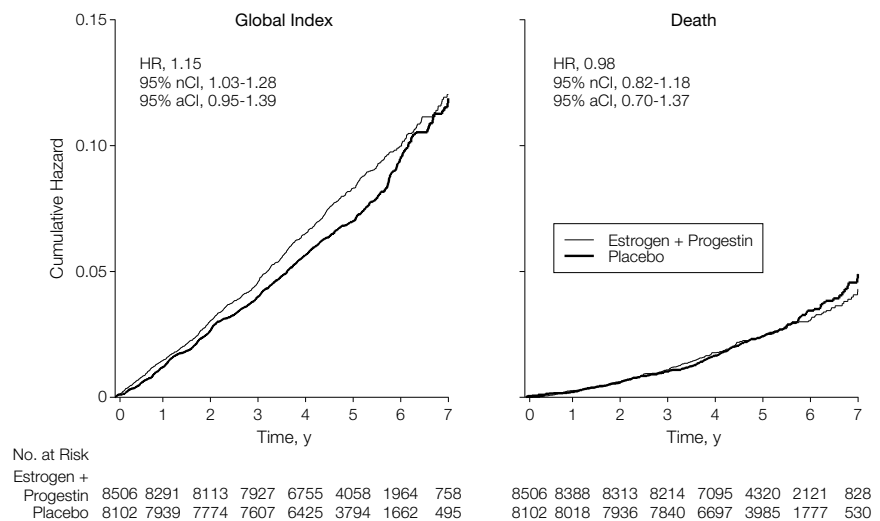
Further Analyses

Because a number of women stopped study medications during follow-up, several analyses were performed to examine the sensitivity of the principal HR estimates to actual use of study medications. Analyses that censored a woman's event history 6 months after becoming nonadherent (using $< 80\%$ of or stopping study drugs) produced the largest changes to estimated effect sizes. This approach increased HRs to 1.51 for CHD, to 1.49 for breast cancer, to 1.67 for stroke, and to 3.29 for VTE. Analyses attributing events to actual hor-

mone use ("as treated," allowing for a 6-month lag) produced more modest changes to these estimates. Analyses excluding women randomized during the

period when the unopposed-estrogen component was open to women with a uterus and analyses stratifying by enrollment period did not substantially

Figure 4. Kaplan-Meier Estimates of Cumulative Hazards for Global Index and Death



HR indicates hazard ratio; nCI, nominal confidence interval; and aCI, adjusted confidence interval.

Table 4. Selected Clinical Outcomes by Follow-up Year and Randomization Assignment*

Outcomes	Year 1			Year 2			Year 3			z Score for Trend†
	E + P	Placebo	Ratio	E + P	Placebo	Ratio	E + P	Placebo	Ratio	
No. of participant-years	8435	8050		8353	7980		8268	7888		
Coronary heart disease	43 (0.51)	23 (0.29)	1.78	36 (0.43)	30 (0.38)	1.15	20 (0.24)	18 (0.23)	1.06	
Stroke	17 (0.20)	17 (0.21)	0.95	27 (0.32)	15 (0.19)	1.72	30 (0.36)	16 (0.20)	1.79	
Venous thromboembolism	49 (0.58)	13 (0.16)	3.60	26 (0.31)	11 (0.14)	2.26	21 (0.25)	12 (0.15)	1.67	
Invasive breast cancer	11 (0.13)	17 (0.21)	0.62	26 (0.31)	30 (0.38)	0.83	28 (0.34)	23 (0.29)	1.16	
Endometrial cancer	2 (0.02)	2 (0.02)	0.95	4 (0.05)	4 (0.05)	0.96	4 (0.05)	5 (0.06)	0.76	
Colorectal cancer	10 (0.12)	15 (0.19)	0.64	11 (0.13)	9 (0.11)	1.17	6 (0.07)	8 (0.10)	0.72	
Hip fracture	6 (0.07)	9 (0.11)	0.64	8 (0.10)	13 (0.16)	0.59	11 (0.13)	12 (0.15)	0.87	
Total death	22 (0.26)	17 (0.21)	1.24	30 (0.36)	30 (0.38)	0.96	39 (0.47)	35 (0.44)	1.06	
Global index	123 (1.46)	96 (1.19)	1.22	134 (1.60)	117 (1.47)	1.09	127 (1.54)	107 (1.36)	1.13	
Outcomes	Year 4			Year 5			Year 6 and Later			z Score for Trend†
	E + P	Placebo	Ratio	E + P	Placebo	Ratio	E + P	Placebo	Ratio	
No. of participant-years	7926	7562		5964	5566		5129	4243		
Coronary heart disease	25 (0.32)	24 (0.32)	0.99	23 (0.39)	9 (0.16)	2.38	17 (0.33)	18 (0.42)	0.78	-1.19
Stroke	25 (0.32)	14 (0.19)	1.70	16 (0.27)	8 (0.14)	1.87	12 (0.23)	15 (0.35)	0.66	-0.51
Venous thromboembolism	27 (0.34)	14 (0.19)	1.84	16 (0.27)	6 (0.11)	2.49	12 (0.23)	11 (0.26)	0.90	-2.45
Invasive breast cancer	40 (0.50)	22 (0.29)	1.73	34 (0.57)	12 (0.22)	2.64	27 (0.53)	20 (0.47)	1.12	2.56
Endometrial cancer	10 (0.13)	5 (0.07)	1.91	1 (0.02)	4 (0.07)	0.23	1 (0.02)	5 (0.12)	0.17	-1.58
Colorectal cancer	9 (0.11)	20 (0.26)	0.43	4 (0.07)	8 (0.14)	0.47	5 (0.10)	7 (0.16)	0.59	-0.81
Hip fracture	8 (0.10)	11 (0.15)	0.69	5 (0.08)	8 (0.14)	0.58	6 (0.12)	9 (0.21)	0.55	0.25
Total death	55 (0.69)	48 (0.63)	1.09	41 (0.69)	44 (0.79)	0.87	44 (0.86)	44 (1.04)	0.83	-0.79
Global index	155 (1.96)	127 (1.68)	1.16	112 (1.88)	77 (1.38)	1.36	100 (1.95)	99 (2.33)	0.84	-0.87

*E + P indicates estrogen plus progestin. All outcome data are number of patients (annualized percentage).

†Tests for trends are based on Cox proportional hazards models with time-dependent treatment effects. The z scores shown indicate trends across all years.

affect the results. These analyses suggest that the intention-to-treat estimates of HRs may somewhat underestimate the effect sizes relative to what would be observed with full adherence to study medications.

COMMENT

The WHI provides evidence from a large randomized trial that addresses the important issue of whether most women with an intact uterus in the decades of life following menopause should consider hormone therapy to prevent chronic disease. The WHI enrolled a cohort of mostly healthy, ethnically diverse women, spanning a large age range (50-79 years at baseline). It is noteworthy that the increased risks for cardiovascular disease and invasive breast cancer were present across racial/ethnic and age strata and were not influenced by the antecedent risk status or prior disease. Hence, the results are likely to be generally applicable to healthy women in this age range. At the time the trial was stopped, the increases in numbers of invasive breast cancers, CHD, stroke, and PE made approximately equal contributions to harm in the estrogen plus progestin group compared with placebo, which were not counterbalanced by the smaller reductions in numbers of hip fractures and colorectal cancers.

Cardiovascular Disease

Even though the trial was stopped early for harm from breast cancer, a sufficient number of CHD events had occurred by 5.2 years of average follow-up to suggest that continuation to the planned end would have been unlikely to yield a favorable result for the primary outcome of CHD. Even if there were a reversal of direction toward benefit of a magnitude seen in the observational studies (ie, a risk reduction of 55%) during the remaining years, conditional power analyses indicate that less than 10% power remained for showing potential benefit if the trial continued.

The WHI finding that estrogen plus progestin does not confer benefit for

preventing CHD among women with a uterus concurs with HERS findings among women with clinically apparent CHD,¹⁰ with the Estrogen Replacement for Atherosclerosis trial, in which estrogen plus progestin did not inhibit progression,²³ and with a trial in women with unstable angina that did not observe a reduction in ischemic events.²⁴ The finding of an increased risk after initiation of treatment in WHI is similar to HERS. In HERS, after 4.1 and 6.8 years of follow-up, hormone therapy did not increase or decrease risk of cardiovascular events in women with CHD.²⁵ The WHI extends these findings to include a wider range of women, including younger women and those without clinically apparent CHD, and indicates that the risk may persist for some years.

Unlike CHD, the excess risk of stroke in the estrogen plus progestin group was not present in the first year but appeared during the second year and persisted through the fifth year. Preliminary analyses indicate that the modest difference in blood pressure between groups does not contribute much to an explanation of the increase in strokes (data not shown). The findings in WHI for stroke are consistent with but somewhat more extreme than those of HERS, which reported a nonsignificant 23% increase in the treatment group.²⁶ The results were also more extreme than those of the Women's Estrogen and Stroke Trial of estradiol (without progestin) in women with prior stroke, which found no effect of estrogen on recurrent strokes overall but some increase in the first 6 months.²⁷ Trials of the effect of estradiol on carotid intima-media thickness have yielded conflicting results.^{28,29} At least 1 observational study has suggested that that use of estrogen plus progestin is associated with higher risk of stroke than estrogen alone.¹⁴ In WHI, there was no indication that excess strokes due to estrogen plus progestin were more likely to occur in older women, in women with prior stroke history, by race/ethnicity, or in women with high blood pressure at baseline. Therefore, it appears that estrogen plus

progestin increases the risk of strokes in apparently healthy women.

Venous thromboembolism is an expected complication of postmenopausal hormones, and the pattern over time in WHI is consistent with the findings from HERS and several observational studies.^{30,31}

Cancer

The WHI is the first randomized controlled trial to confirm that combined estrogen plus progestin does increase the risk of incident breast cancer and to quantify the degree of risk. The WHI could not address the risk of death due to breast cancer because with the relatively short follow-up time, few women in the WHI have thus far died as a result of breast cancer (3 in the active treatment group and 2 in the placebo group). The risk of breast cancer emerged several years after randomization. After an average follow-up of about 5 years, the adverse effect on breast cancer had crossed the monitoring boundary. The 26% excess of breast cancer is consistent with estimates from pooled epidemiological data, which reported a 15% increase for estrogen plus progestin use for less than 5 years and a 53% increase for use for more than 5 years.³² It is also consistent with the (nonsignificant) 27% increase found after 6.8 years of follow-up in HERS.³³

With more common use of estrogen plus progestin, several epidemiological studies have reported that estrogen plus progestin appears to be associated with greater risk of breast cancer than estrogen alone.³⁴⁻³⁷ In the PEPI trial, women in the 3 estrogen plus progestin groups had much greater increases in mammographic density (a predictor of breast cancer) than women in the estrogen or placebo groups.³⁸ In WHI, the HR for estrogen plus progestin was not higher in women with a family history or other risk factors for breast cancer, except for reported prior use of postmenopausal hormones. This may suggest a cumulative effect of years of exposure to postmenopausal hormones.

Endometrial cancer rates were low and were not increased by 5 years of es-

trogen plus progestin exposure. Close monitoring for bleeding and treatment of hyperplasia may contribute to the absence of increased risk of endometrial cancer.

The reduction in colorectal cancer in the hormone group is consistent with observational studies, which have suggested fairly consistently that users of postmenopausal hormones may be at lower risk of colorectal cancer.³⁹ The mechanisms by which hormone use might reduce risk are unclear. Results from other trials of postmenopausal hormones will help resolve the effects of hormones on colorectal cancer.⁴⁰

Fractures

The reductions in clinical vertebral fractures, other osteoporotic fractures, and combined fractures supported the benefit for hip fractures found in this trial. These findings are consistent with the observational data and limited data from clinical trials⁴¹ and are also consistent with the known ability of estrogen (with or without progestin) to maintain bone mineral density.⁴² The WHI is the first trial with definitive data supporting the ability of postmenopausal hormones to prevent fractures at the hip, vertebrae, and other sites.

Overall Risks and Benefits

At the end of the trial, the global index indicated that there were more harmful than beneficial outcomes in the estrogen plus progestin group vs the placebo group. The monitored outcomes included in the global index were selected to represent diseases of serious import that estrogen plus progestin treatment might affect, and do not include a variety of other conditions and measures that may be affected in unfavorable or favorable ways (eg, gallbladder disease, diabetes, quality of life, and cognitive function). The data on these and other outcomes will be the subject of future publications. All-cause mortality was balanced between the groups; however, longer follow-up may be needed to assess the impact of the incident diseases on total mortality.

The absolute excess risk (or risk reduction) attributable to estrogen plus progestin was low. Over 1 year, 10000 women taking estrogen plus progestin compared with placebo might experience 7 more CHD events, 8 more strokes, 8 more PEs, 8 more invasive breast cancers, 6 fewer colorectal cancers, and 5 fewer hip fractures. Combining all the monitored outcomes, women taking estrogen plus progestin might expect 19 more events per year per 10000 women than women taking placebo. Over a longer period, more typical of the duration of treatment that would be needed to prevent chronic disease, the absolute numbers of excess outcomes would increase proportionately.

During the 5.2 years of this trial, the number of women experiencing a global index event was about 100 more per 10000 women taking estrogen plus progestin than taking placebo. If the current findings can be extrapolated to an even longer treatment duration, the absolute risks and benefits associated with estrogen plus progestin for each of these conditions could be substantial and on a population basis could account for tens of thousands of conditions caused, or prevented, by hormone use.

Limitations

This trial tested only 1 drug regimen, CEE, 0.625 mg/d, plus MPA, 2.5 mg/d, in postmenopausal women with an intact uterus. The results do not necessarily apply to lower dosages of these drugs, to other formulations of oral estrogens and progestins, or to estrogens and progestins administered through the transdermal route. It remains possible that transdermal estradiol with progesterone, which more closely mimics the normal physiology and metabolism of endogenous sex hormones, may provide a different risk-benefit profile. The WHI findings for CHD and VTE are supported by findings from HERS, but there is no other evidence from clinical trials for breast cancer and colorectal cancer, and only limited data from trials concerning fractures.

Importantly, this trial could not distinguish the effects of estrogen from

those of progestin. The effects of progestin may be important for breast cancer and atherosclerotic diseases, including CHD and stroke. Per protocol, in a separate and adequately powered trial, WHI is testing the hypothesis of whether oral estrogen will prevent CHD in 10739 women who have had a hysterectomy. The monitoring of this trial is similar to that for the trial of estrogen plus progestin. At an average follow-up of 5.2 years, the DSMB has recommended that this trial continue because the balance of overall risks and benefits remains uncertain. These results are expected to be available in 2005, at the planned termination.

The relatively high rates of discontinuation in the active treatment arm (42%) and crossover to active treatment in the placebo arm (10.7%) are a limitation of the study; however, the lack of adherence would tend to decrease the observed treatment effects. Thus, the results presented here may underestimate the magnitude of both adverse effects on cardiovascular disease and breast cancer and the beneficial effects on fractures and colorectal cancer among women who adhere to treatment.

The fact that the trial was stopped early decreases the precision of estimates of long-term treatment effects. A longer intervention period might have shown more pronounced benefit for fractures and might have yielded a more precise test of the hypothesis that treatment reduces colorectal cancer. Nonetheless, it appears unlikely that benefit for CHD would have emerged by continuing the trial to its planned termination. The trial results indicate that treatment for up to 5.2 years is not beneficial overall and that there is early harm for CHD, continuing harm for stroke and VTE, and increasing harm for breast cancer with increasing duration of treatment. This risk-benefit profile is not consistent with the requirements for a viable intervention for the primary prevention of chronic diseases.

Implications

The WHI trial results provide the first definitive data on which to base treat-

ment recommendations for healthy postmenopausal women with an intact uterus. This trial did not address the short-term risks and benefits of hormones given for the treatment of menopausal symptoms. On the basis of HERS and other secondary prevention trials, the American Heart Association recommended against initiating postmenopausal hormones for the secondary prevention of cardiovascular disease.⁴³ The American Heart Association made no firm recommendation for primary prevention while awaiting the results from randomized clinical trials such as WHI, and stated that continuation of the treatment should be considered on the basis of established noncoronary benefits and risks, possible coronary benefits and risks, and patient preference.

Results from WHI indicate that the combined postmenopausal hormones CEE, 0.625 mg/d, plus MPA, 2.5 mg/d, should not be initiated or continued for the primary prevention of CHD. In addition, the substantial risks for cardiovascular disease and breast cancer must be weighed against the benefit for fracture in selecting from the available agents to prevent osteoporosis.

Writing Group for the Women's Health Initiative Investigators: Jacques E. Rossouw, MBChB, MD, National Heart, Lung, and Blood Institute, Bethesda, Md; Garnet L. Anderson, PhD, Ross L. Prentice, PhD, Andrea Z. LaCroix, PhD, and Charles Kooperberg, PhD, Fred Hutchinson Cancer Research Center, Seattle, Wash; Marcia L. Stefanick, PhD, Stanford University Clinical Center, Stanford, Calif; Rebecca D. Jackson, MD, Ohio State University Clinical Center, Columbus; Shirley A. A. Beresford, PhD, Fred Hutchinson Cancer Research Center, Seattle, Wash; Barbara V. Howard, PhD, MedStar Research Institute, Washington, DC; Karen C. Johnson, MD, MPH, University of Tennessee, Memphis; Jane Morley Kotchen, MD, Medical College of Wisconsin, Milwaukee; Judith Ockene, PhD, University of Massachusetts Medical School, Worcester.

Financial Disclosures: Dr LaCroix is an investigator on 2 osteoporosis studies separately funded by Merck and Pfizer. Dr Jackson is an investigator on 1 osteoporosis study funded by Merck and 1 hormone study on libido in women funded by Procter & Gamble Pharmaceuticals.

For Correspondence: Jacques E. Rossouw, MBChB, MD, Division of Women's Health Initiative, National Heart, Lung, and Blood Institute, 6705 Rockledge Dr, One Rockledge Ctr, Suite 300, MS/7966, Bethesda, MD 20817 (e-mail: rossouw@nih.gov); Garnet L. Anderson, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, MP-1002, PO Box 19024, Seattle, WA 98109-1024 (e-mail: garnet@whi.org).

Reprints: WHI Clinical Coordinating Center, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, MP-1002, PO Box 19024, Seattle, WA 98109-1024.

Author Contributions: Dr Anderson, as co-principal investigator of the Women's Health Initiative Clinical Coordinating Center, had full access to the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analyses.

Study concept and design: Rossouw, Anderson, Prentice.

Acquisition of data: Anderson, Prentice, LaCroix, Kooperberg, Stefanick, Jackson, Beresford, Howard, Johnson, Kotchen, Ockene.

Analysis and interpretation of data: Rossouw, Anderson, Prentice, LaCroix, Kooperberg, Stefanick, Jackson, Beresford, Howard, Johnson, Kotchen, Ockene.

Drafting of the manuscript: Rossouw, Anderson, Prentice.

Critical revision of the manuscript for important intellectual content: Rossouw, Anderson, LaCroix, Kooperberg, Stefanick, Jackson, Beresford, Howard, Johnson, Kotchen, Ockene.

Statistical expertise: Anderson, Prentice, Kooperberg.

Obtained funding: Rossouw, Anderson, Prentice, Stefanick, Beresford, Howard, Kotchen, Ockene.

Administrative, technical, or material support: Rossouw, Anderson, Prentice, LaCroix, Kooperberg, Stefanick, Jackson, Beresford, Howard, Johnson, Kotchen, Ockene.

Study supervision: Rossouw, Anderson, LaCroix, Stefanick, Jackson, Beresford, Howard, Johnson, Kotchen, Ockene.

WHI Steering Committee: *Clinical Centers:* Marcia L. Stefanick (chair), Stanford Center for Research in Disease Prevention, Stanford University; Rebecca D. Jackson (vice chair), Ohio State University; Catherine I. Allen, University of Wisconsin, Madison; Annlouise R. Assaf, Brown University, Providence, RI; Tamsen Bassford, University of Arizona, Tucson/Phoenix; Shirley A. A. Beresford, Fred Hutchinson Cancer Research Center; Henry Black, Rush-Presbyterian-St Luke's Medical Center, Chicago, Ill; Robert Brunner, University of Nevada, Reno; Gregory L. Burke, Wake Forest University School of Medicine, Winston-Salem, NC; Bette Caan, Kaiser Permanente Division of Research, Oakland, Calif; Rowan T. Chlebowski, Harbor-UCLA Research and Education Institute, Torrance, Calif; David Curb, University of Hawaii, Honolulu; Margery Gass, University of Cincinnati, Cincinnati, Ohio; Jennifer Hays, Baylor College of Medicine, Houston, Tex; Gerardo Heiss, University of North Carolina, Chapel Hill; Susan Hendrix, Wayne State University School of Medicine/Hutzel Hospital, Detroit, Mich; Barbara V. Howard, MedStar Research Institute, Washington, DC; Judith Hsia, George Washington University, Washington, DC; F. Allan Hubbell, University of California, Irvine, Orange; Karen C. Johnson, University of Tennessee, Memphis; Howard Judd, University of California, Los Angeles; Jane Morley Kotchen, Medical College of Wisconsin, Milwaukee; Lewis Kuller, University of Pittsburgh, Pittsburgh, Pa; Dorothy Lane, State University of New York at Stony Brook; Robert D. Langer, University of California, San Diego, LaJolla/Chula Vista; Norman Lasser, University of Medicine and Dentistry of New Jersey, Newark; Cora E. Lewis, University of Alabama at Birmingham; Marian Limacher, University of Florida, Gainesville/Jacksonville; JoAnn Manson, Brigham and Women's Hospital, Harvard Medical School, Boston, Mass; Karen Margolis, University of Minnesota, Minneapolis; Judith Ockene, University of Massachusetts Medical School, Worcester; Mary Jo O'Sullivan, University of Miami, Miami, Fla; Lawrence Phillips, Emory University, Atlanta, Ga; Cheryl Ritenbaugh, Kaiser Permanente Center for Health Research, Portland, Ore; John Robbins, University of California, Davis, Sacramento; Robert Schenken, University of Texas Health Science Center, San Antonio; Sylvia Wassertheil-Smolter, Albert Einstein College of Medicine, Bronx, NY; Maurizio Trevisan, State University of New York at Buffalo; Linda Van Horn, Northwestern University, Chicago/Evanston, Ill; and Robert Wallace, University of Iowa, Iowa City/Davenport; *Program Office:* Jacques E. Rossouw, National Heart, Lung, and Blood Institute; *Clinical Coordinating Center:* Andrea Z. LaCroix, Ruth E. Patterson, and Ross L. Prentice, Fred Hutchinson Cancer Research Center.

Data and Safety Monitoring Board: Janet Wittes (chair), Eugene Braunwald, Margaret Chesney, Harvey Cohen, Elizabeth Barrett-Connor, David DeMets, Leo Dunn, Johanna Dwyer, Robert P. Heaney, Victor Vogel, LeRoy Walters, and Salim Yusuf.

Funding/Support: The National Heart, Lung, and Blood Institute funds the WHI program. Wyeth-Ayerst Research provided the study medication (active and placebo).

Acknowledgment: The WHI Steering Committee gratefully acknowledges the dedicated efforts of the WHI participants and of key WHI investigators and staff at the clinical centers and the Clinical Coordinating Center. A full listing of the WHI investigators can be found at <http://www.whi.org>.

REFERENCES

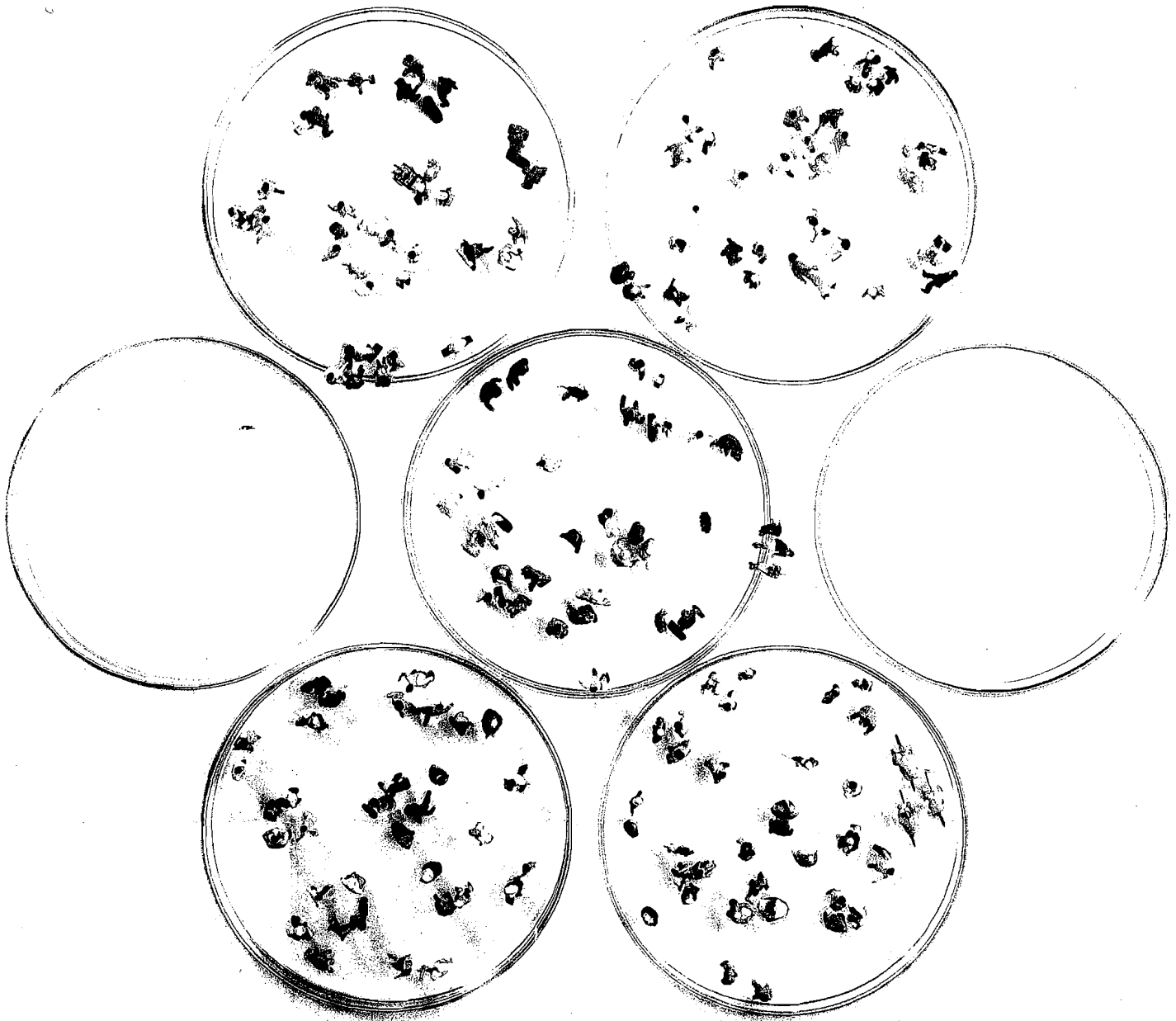
1. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998;19: 61-109.
2. Stampfer M, Colditz G. Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med*. 1991;20:47-63.
3. Grady D, Ruben SB, Pettit DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med*. 1992;117: 1016-1037.
4. Rijkema AH, van der Sanden AA, Ruijs AH. Effects of postmenopausal estrogen-progesterone therapy on serum lipids and lipoproteins: a review. *Maturitas*. 1990;12:259-285.
5. Adams MR, Kaplan JR, Manuck SB, et al. Inhibition of coronary artery atherosclerosis by 17-beta estradiol in ovariectomized monkeys: lack of an effect of added progesterone. *Arteriosclerosis*. 1990;10: 1051-1057.
6. Weiss NS, Ure CL, Ballard JH, Williams AR, Daling JR. Decreased risk of fractures of the hip and lower forearm with postmenopausal use of estrogen. *N Engl J Med*. 1980;303:1195-1198.
7. Genant HK, Baylink DJ, Gallagher JC, Harris ST, Steiger P, Herber M. Effect of estrone sulfate on postmenopausal bone loss. *Obstet Gynecol*. 1990;76: 579-584.
8. Steinberg KA, Thacker SB, Smith SJ, et al. A meta-analysis of the effect of estrogen replacement therapy on the risk of breast cancer. *JAMA*. 1991;265:1985-1990.
9. Gerhardsson de VM, London S. Reproductive factors, exogenous female hormones, and colorectal cancer by subsite. *Cancer Causes Control*. 1992;3: 355-360.
10. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA*. 1998;280:605-613.
11. Grodstein F, Manson JE, Stampfer MJ. Postmenopausal hormone use and secondary prevention of coronary events in the Nurses' Health Study. *Ann Intern Med*. 2001;135:1-8.
12. Alexander KP, Newby LK, Hellkamp AS, et al. Initiation of hormone replacement therapy after acute myocardial infarction is associated with more cardiac events during follow-up. *J Am Coll Cardiol*. 2001; 38:1-7.
13. Heckbert SR, Kaplan RC, Weiss NS, et al. Risk of recurrent coronary events in relation to use and recent initiation of postmenopausal hormone therapy. *Arch Intern Med*. 2001;161:1709-1713.

14. Grodstein F, Manson JE, Colditz GA, Willit WC, Speizer FE, Stampfer MJ. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med.* 2000;133:933-941.
15. The Writing Group for the PEPI Trial. Effects of hormone replacement therapy on endometrial histology in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA.* 1996;275:370-375.
16. Ives DG, Fitzpatrick AL, Bild DE, et al. Surveillance and ascertainment of cardiovascular events: the Cardiovascular Health Study. *Ann Epidemiol.* 1995;5:275-285.
17. Cox DR. Regression analysis and life tables. *J R Stat Soc.* 1972;34:187-220.
18. O'Brien PC, Fleming RT. A multiple testing procedure for clinical trials. *Biometrics.* 1979;35:549-556.
19. Freedman LS, Anderson G, Kipnis V, et al. Approaches to monitoring the results of long-term disease prevention trials: examples from the Women's Health Initiative. *Control Clin Trials.* 1996;17:509-525.
20. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst.* 1989;81:1879-1886.
21. Pilon D, Castilloux AM, Leloir J. Estrogen replacement therapy: determinants of persistence with treatment. *Obstet Gynecol.* 2001;97:97-100.
22. Writing Group for the PEPI Trial. Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women. *JAMA.* 1995;273:199-208.
23. Herrington DM, Reboussin DM, Brosnihan KB, et al. Effects of estrogen replacement on the progression of coronary artery atherosclerosis. *N Engl J Med.* 2000;343:522-529.
24. Schulman SP, Thiemann DR, Ouyang P, et al. Effects of acute hormone therapy on recurrent ischemia in postmenopausal women with unstable angina. *J Am Coll Cardiol.* 2002;39:231-237.
25. Grady D, Herrington D, Bittner V, et al, for the HERS Research Group. Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study Follow-up (HERS II). *JAMA.* 2002;288:49-57.
26. Simon JA, Hsia J, Cauley JA, et al. Postmenopausal hormone therapy and risk of stroke: the Heart and Estrogen/progestin Replacement Study (HERS). *Circulation.* 2001;103:638-642.
27. Viscoli CM, Brass LM, Kernan WN, Sarrel PM, Sussman S, Horwitz RJ. A clinical trial of estrogen-replacement therapy after ischemic stroke. *N Engl J Med.* 2001;345:1243-1249.
28. Hodis HN, Mack WJ, Lobo RA, et al. Estrogen in the prevention of atherosclerosis: a randomized, double-blind controlled trial. *Ann Intern Med.* 2001;135:939-953.
29. Angerer P, Stork S, Kothny W, Schmitt P, von Schacky C. Effect of oral postmenopausal hormone replacement on progression of atherosclerosis: a randomized, controlled trial. *Arterioscler Thromb Vasc Biol.* 2001;21:262-268.
30. Castellsague J, Perez Gutthann S, Garcia Rodriguez LA. Recent epidemiological studies of the association between hormone replacement therapy and venous thromboembolism: a review. *Drug Saf.* 1998;18:117-123.
31. Grady D, Wenger NK, Herrington D, et al. Postmenopausal hormone therapy increases risk for venous thromboembolic disease: the Heart and Estrogen/progestin Replacement Study. *Ann Intern Med.* 2000;132:689-696.
32. 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.
33. Hulley S, Furberg C, Barrett-Connor E, et al, for the HERS Research Group. Noncardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study Follow-up (HERS II). *JAMA.* 2002;288:58-66.
34. Schairer C, Lubin J, Troisi R, Sturgeon S, Brinton L, Hoover R. Menopausal estrogen and estrogen-progestin replacement therapy and breast cancer risk. *JAMA.* 2000;283:485-491.
35. Ross RK, Paganini-Hill A, Wan PC, Pike MC. Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. *J Natl Cancer Inst.* 2000;92:328-332.
36. Colditz GA, Hankinson SE, Hunter DJ, et al. The use of estrogens and progestins and the risk of breast cancer in postmenopausal women. *N Engl J Med.* 1995;332:1589-1593.
37. Magnusson C, Baron JA, Correia N, Bergstrom R, Adami H-O, Persson I. Breast-cancer risk following long-term oestrogen and oestrogen-progestin-replacement therapy. *Int J Cancer.* 1999;81:339-344.
38. Greendale GA, Reboussin BA, Sie A, et al. Effects of estrogen and estrogen-progestin on mammographic parenchymal density. *Ann Intern Med.* 1999;130:262-269.
39. Grodstein F, Newcomb PA, Stampfer MJ. Postmenopausal hormone therapy and the risk of colorectal cancer: a review and meta-analysis. *Am J Med.* 1999;106:574-582.
40. Vickers MR, Meade TW, Wilkes HC. Hormone replacement therapy and cardiovascular disease: the case for a randomized controlled trial. *Ciba Found Symp.* 1995;191:150-160.
41. Torgerson DJ, Bell-Seyer SE. Hormone replacement therapy and prevention of nonvertebral fractures: a meta-analysis of randomized trials. *JAMA.* 2001;285:2891-2897.
42. Writing Group for the PEPI Trial. Effects of hormone therapy on bone mineral density: results from the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA.* 1996;275:1389-1396.
43. Mosca L, Collins P, Herrington DM, et al. Hormone replacement therapy and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. *Circulation.* 2001;104:499-503.

"GHOST WORLD" AND "ART SCHOOL CONFIDENTIAL") THAT JUST MIGHT BE THE STRANGEST STORY OF A BLIND DATE EVER ... WELL, EVER TOLD BY A GRAPHIC NOVELIST.

The New York Times Magazine

SEPTEMBER 16, 2007 / SECTION 6



Unhealthy Science

LYNN HIRSCHBERG: RACHEL ZOE, UNDRESSED ZEV CHAFETS: HOW GOOD FOR THE JEWS IS LEV LEVIEV?

D

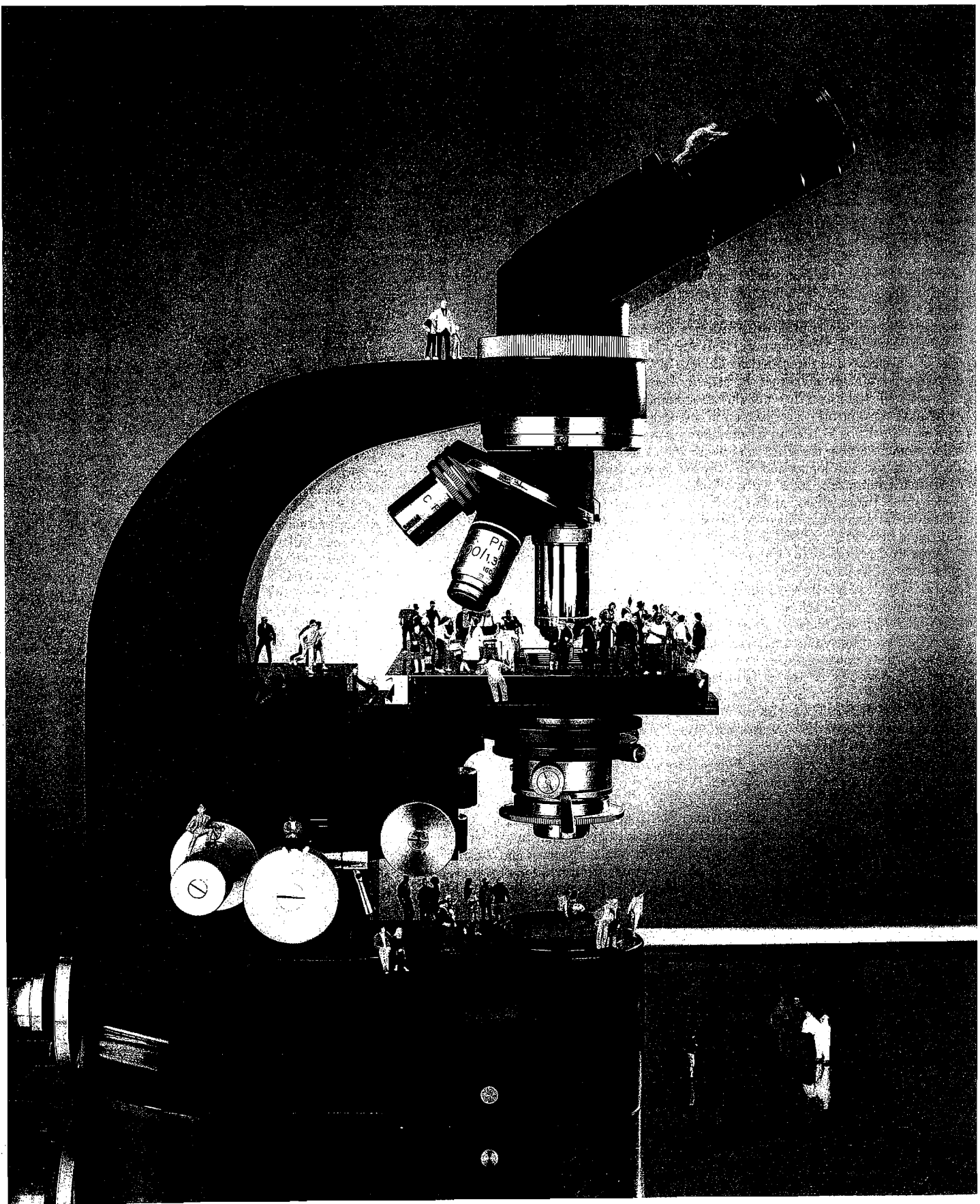
Do We Really
Know What Makes
Us Healthy?

M

uch of what we're told
about diet, lifestyle and disease
is based on epidemiologic
studies.

What if it is just bad science? **By Gary Taubes**

Photographs by Reinhard Hunger



Once

upon a time, women took estrogen only to relieve the hot flashes, sweating, vaginal dryness and the other discomforting symptoms of menopause. In the late 1960s, thanks in part to the efforts of Robert Wilson, a Brooklyn gynecologist, and his 1966 best seller, "Feminine Forever," this began to change, and estrogen therapy evolved into a long-term remedy for the chronic ills of aging. Menopause, Wilson argued, was not a natural age-related condition; it was an illness, akin to diabetes or kidney failure, and one that could be treated by taking estrogen to replace the hormones that a woman's ovaries secreted in ever diminishing amounts. With this argument estrogen evolved into hormone-replacement therapy, or H.R.T., as it came to be called, and became one of the most popular prescription drug treatments in America.

By the mid-1990s, the American Heart Association, the American College of Physicians and the American College of Obstetricians and Gynecologists had all concluded that the beneficial effects of H.R.T. were sufficiently well established that it could be recommended to older women as a means of warding off heart disease and osteoporosis. By 2001, 15 million women were filling H.R.T. prescriptions annually; perhaps 5 million were older women, taking the drug solely with the expectation that it would allow them to lead a longer and healthier life. A year later, the tide would turn. In the summer of 2002, estrogen therapy was exposed as a hazard to health rather than a benefit, and its story became what Jerry Avorn, a Harvard epidemiologist, has called the "estrogen debacle" and a "case study waiting to be written" on the elusive search for truth in medicine.

Many explanations have been offered to make sense of the here-today-gone-tomorrow nature of medical wisdom — what we are advised with confidence one year is reversed the next — but the simplest one is that it is the natural rhythm of science. An observation leads to a hypothesis. The hypothesis (last year's advice) is tested, and it fails this year's test, which is always the most likely outcome in any scientific endeavor. There are, after all, an infinite number of wrong hypotheses for every right one, and so the odds are always against any particular hypothesis being true, no matter how obvious or vitally important it might seem.

In the case of H.R.T., as with most issues of diet, lifestyle and disease, the hypotheses begin their transformation into public-health recommendations only after they've received the requisite support from a field of research known as epidemiology. This science evolved over the last 250 years to make sense of epidemics — hence the name — and infectious diseases. Since the 1950s, it has been used to identify, or at least to try to identify, the causes of the common chronic diseases that befall us, particularly heart disease and cancer. In the process, the perception of what epidemiologic research can legitimately accomplish — by the public, the press and perhaps by many epidemiologists themselves — may have run far ahead of the reality. The case of hormone-replacement therapy for post-menopausal women is just one of the cautionary tales in the annals of epidemiology. It's a particularly glaring example of the difficulties of trying to establish reliable knowledge in any scientific field with research tools that themselves may be unreliable.

What was considered true about estrogen therapy in the 1960s and is still the case today is that it is an effective treatment for menopausal symptoms. Take H.R.T. for a few menopausal years and it's extremely unlikely that any harm will come from it. The uncertainty involves the lifelong risks and benefits should a woman choose to continue taking H.R.T. long past menopause. In 1985, the Nurses' Health Study run out of the Harvard Medical School and the Harvard School of Public Health reported that women taking estrogen had only a third as many heart attacks as women who had

never taken the drug. This appeared to confirm the belief that women were protected from heart attacks until they passed through menopause and that it was estrogen that bestowed that protection, and this became the basis of the therapeutic wisdom for the next 17 years.

Faith in the protective powers of estrogen began to erode in 1998, when a clinical trial called HERS, for Heart and Estrogen-progestin Replacement Study, concluded that estrogen therapy increased, rather than decreased, the likelihood that women who already had heart disease would suffer a heart attack. It evaporated entirely in July 2002, when a second trial, the Women's Health Initiative, or W.H.I., concluded that H.R.T. constituted a potential health risk for all postmenopausal women. While it might protect them against osteoporosis and perhaps colorectal cancer, these benefits would be outweighed by increased risks of heart disease, stroke, blood clots, breast cancer and perhaps even dementia. And that was the final word. Or at least it was until the June 21 issue of *The New England Journal of Medicine*. Now the idea is that hormone-replacement therapy may indeed protect women against heart disease if they begin taking it during menopause, but it is still decidedly deleterious for those women who begin later in life.

This latest variation does come with a caveat, however, which could have been made at any point in this history. While it is easy to find authority figures in medicine and public health who will argue that today's version of H.R.T. wisdom is assuredly the correct one, it's equally easy to find authorities who will say that surely we don't know. The one thing on which they will all agree is that the kind of experimental trial necessary to determine the truth would be excessively expensive and time-consuming and so will almost assuredly never happen. Meanwhile, the question of how many women may have died prematurely or suffered strokes or breast cancer because they were taking a pill that their physicians had prescribed to protect them against heart disease lingers unanswered. A reasonable estimate would be tens of thousands.

The Flip-Flop Rhythm of Science

At the center of the H.R.T. story is the science of epidemiology itself and, in particular, a kind of study known as a prospective or cohort study, of which the Nurses' Health Study is among the most renowned. In these studies, the investigators monitor disease rates and lifestyle factors (diet, physical activity, prescription drug use, exposure to pollutants, etc.) in or between large populations (the 122,000 nurses of the Nurses' study, for example). They then try to infer conclusions — i.e., hypotheses — about what caused the disease variations observed. Because these studies can generate an enormous number of speculations about the causes or prevention of chronic diseases, they provide the fodder for much of the health news that appears in the media — from the potential benefits of fish oil, fruits and vegetables to the supposed dangers of sedentary lives, trans fats and electromagnetic fields. Because these studies often provide the only available evidence outside the laboratory on critical issues of our well-being, they have come to play a significant role in generating public-health recommendations as well.

The dangerous game being played here, as David Sackett, a retired Oxford University epidemiologist, has observed, is in the presumption of preventive medicine. The goal of the endeavor is to tell those of us who are otherwise in fine health how to remain healthy longer. But this advice comes with the expectation that any prescription given — whether diet or drug or a change in lifestyle — will indeed prevent disease rather than be the agent of our disability or untimely death. With that presumption, how unambiguous does the evidence have to be before any advice is offered?

The catch with observational studies like the Nurses' Health Study, no matter how well designed and how many tens of thousands of subjects they might include, is that they have a fundamental limitation. They can distinguish asso-

Gary Taubes is the author of the forthcoming book "Good Calories, Bad Calories: Challenging the Conventional Wisdom on Diet, Weight Control and Disease."



ciations between two events — that women who take H.R.T. have less heart disease, for instance, than women who don't. But they cannot inherently determine causation — the conclusion that one event causes the other; that H.R.T. protects against heart disease. As a result, observational studies can only provide what researchers call hypothesis-generating evidence — what a defense attorney would call circumstantial evidence.

Testing these hypotheses in any definitive way requires a randomized-controlled trial — an experiment, not an observational study — and these clinical trials typically provide the flop to the flip-flop rhythm of medical wisdom. Until August 1998, the faith that H.R.T. prevented heart disease was based primarily on observational evidence, from the Nurses' Health Study most prominently. Since then, the conventional wisdom has been based on clinical trials — first HERS, which tested H.R.T. against a placebo in 2,700 women with heart disease, and then the Women's Health Initiative, which tested the therapy against a placebo in 16,500 healthy women. When the Women's Health Initiative concluded in 2002 that H.R.T. caused far more harm than good, the lesson to be learned, wrote Sackett in *The Canadian Medical Association Journal*, was about the "disastrous inadequacy of lesser evidence" for shaping medical and public-health policy. The contentious wisdom circa mid-2007 — that estrogen benefits women who begin taking it around the time of menopause but not women who begin substantially later — is an attempt to reconcile the discordance between the observational studies and the experimental ones. And it may be right. It may not. The only way to tell for sure would be to do yet another randomized trial, one that now focused exclusively on women given H.R.T. when they begin their menopause.

A Poor Track Record of Prevention

No one questions the value of these epidemiologic studies when they're used to identify the unexpected side effects of prescription drugs or to study the progression of diseases or their distribution between and within populations. One reason researchers believe that heart disease and many cancers can be prevented is because of observational evidence that the incidence of these diseases differ greatly in different populations and in the same populations over time. Breast cancer is not the scourge among Japanese women that it is

among American women, but it takes only two generations in the United States before Japanese-Americans have the same breast cancer rates as any other ethnic group. This tells us that something about the American lifestyle or diet is a cause of breast cancer. Over the last 20 years, some two dozen large studies, the Nurses' Health Study included, have so far failed to identify what that factor is. They may be inherently incapable of doing so. Nonetheless, we know that such a carcinogenic factor of diet or lifestyle exists, waiting to be identified.

These studies have also been invaluable for identifying predictors of disease — risk factors — and this information can then guide physicians in weighing the risks and benefits of putting a particular patient on a particular drug. The studies have repeatedly confirmed that high blood pressure is associated with an increased risk of heart disease and that obesity is associated with an increased risk of most of our common chronic diseases, but they have not told us what it is that raises blood pressure or causes obesity. Indeed, if you ask

the more skeptical epidemiologists in the field what diet and lifestyle factors have been convincingly established as causes of common chronic diseases based on observational studies without clinical trials, you'll get a very short list: smoking as a cause of lung cancer and cardiovascular disease, sun exposure for skin cancer, sexual activity to spread the papilloma virus that causes cervical cancer and perhaps alcohol for a few different cancers as well.

Richard Peto, professor of medical statistics and epidemiology at Oxford University, phrases the nature of the conflict this way: "Epidemiology is so beautiful and provides such an important perspective on human life and death, but an incredible amount of rubbish is published," by which he means the results of observational studies that appear daily in the news media and often become the basis of public-health recommendations about what we should or should not do to promote our continued good health.

In January 2001, the British epidemiologists George Davey Smith and Shah Ebrahim, co-editors of *The International Journal of Epidemiology*, discussed this issue in an editorial titled "Epidemiology — Is It Time to Call It a Day?" They noted that those few times that a randomized trial had been financed to test a hypothesis supported by results from these large observational studies, the hypothesis either failed the test or, at the very least, the test failed to confirm the hypothesis: antioxidants like vitamins E and C and beta carotene did not prevent heart disease, nor did eating copious fiber protect against colon cancer.

The Nurses' Health Study is the most influential of these cohort studies, and in the six years since the Davey Smith and Ebrahim editorial, a series of new trials have chipped away at its credibility. The Women's Health Initiative hormone-therapy trial failed to confirm the proposition that H.R.T. prevented heart disease; a W.H.I. diet trial with 49,000 women failed to confirm the notion that fruits and vegetables protected against heart disease; a 40,000-woman trial failed to confirm that a daily regimen of low-dose aspirin prevented colorectal cancer and heart attacks in women under 65. And this June, yet another clinical trial — this one of 1,000 men and women with a high risk of colon cancer — contradicted the inference from the Nurses' study that folic acid supplements reduced the risk of colon cancer. Rather, if anything, they appear to increase risk.

The implication of this track record seems hard to avoid. "Even the

Observational studies can only provide what researchers call hypothesis-generating evidence or what a defense attorney would call circumstantial evidence.

Nurses' Health Study, one of the biggest and best of these studies, cannot be used to reliably test small-to-moderate risks or benefits," says Charles Hennekens, a principal investigator with the Nurses' study from 1976 to 2001. "None of them can."

Proponents of the value of these studies for telling us how to prevent common diseases — including the epidemiologists who do them, and physicians, nutritionists and public-health authorities who use their findings to argue for or against the health benefits of a particular regimen — will argue that they are never relying on any single study. Instead, they base their ultimate judgments on the "totality of the data," which in theory includes all the observational evidence, any existing clinical trials and any laboratory work that might provide a biological mechanism to explain the observations.

This in turn leads to the argument that the fault is with the press, not the epidemiology. "The problem is not in the research but in the way it is interpreted for the public," as Jerome Kassirer and Marcia Angell, then the editors of *The New England Journal of Medicine*, explained in a 1994 editorial titled "What Should the Public Believe?" Each study, they explained, is just a "piece of a puzzle" and so the media had to do a better job of communicating the many limitations of any single study and the caveats involved — the foremost, of course, being that "an association between two events is not the same as a cause and effect."

Stephen Pauker, a professor of medicine at Tufts University and a pioneer in the field of clinical decision making, says, "Epidemiologic studies, like diagnostic tests, are probabilistic statements." They don't tell us what the truth is, he says, but they allow both physicians and patients to "estimate the truth" so they can make informed decisions. The question the skeptics will ask, however, is how can anyone judge the value of these studies without taking into account their track record? And if they take into account the track record, suggests Sander Greenland, an epidemiologist at the University of California, Los Angeles, and an author of the textbook "Modern Epidemiology," then wouldn't they do just as well if they simply tossed a coin?

As John Bailer, an epidemiologist who is now at the National Academy of Science, once memorably phrased it, "The appropriate question is not whether there are uncertainties about epidemiologic data, rather, it is whether the uncertainties are so great that one cannot draw useful conclusions from the data."

Science vs. the Public Health

Understanding how we got into this situation is the simple part of the story. The randomized-controlled trials needed to ascertain reliable knowledge about long-term risks and benefits of a drug, lifestyle factor or aspect of our diet are inordinately expensive and time consuming. By randomly assigning research subjects into an intervention group (who take a particular pill or eat a particular diet) or a placebo group, these trials "control" for all other possible variables, both known and unknown, that might effect the outcome: the relative health or wealth of the subjects, for instance. This is why randomized trials, particularly those known as placebo-controlled, double-blind trials, are typically considered the gold standard for establishing reliable knowledge about whether a drug, surgical intervention or diet is really safe and effective.

But clinical trials also have limitations beyond their exorbitant costs and the years or decades it takes them to provide meaningful results. They can

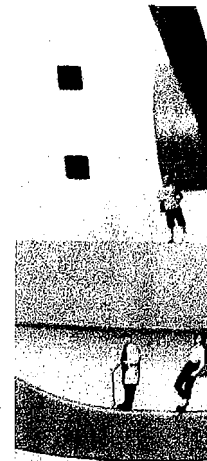
rarely be used, for instance, to study suspected harmful effects. Randomly subjecting thousands of individuals to secondhand tobacco smoke, pollutants or potentially noxious trans fats presents obvious ethical dilemmas. And even when these trials are done to study the benefits of a particular intervention, it's rarely clear how the results apply to the public at large or to any specific patient. Clinical trials invariably enroll subjects who are relatively healthy, who are motivated to volunteer and will show up regularly for treatments and checkups. As a result, randomized trials "are very good for showing that a drug does what the pharmaceutical company says it does," David Atkins, a preventive-medicine specialist at the Agency for Healthcare Research and Quality, says, "but not very good for telling you how big the benefit really is and what are the harms in typical people. Because they don't enroll typical people."

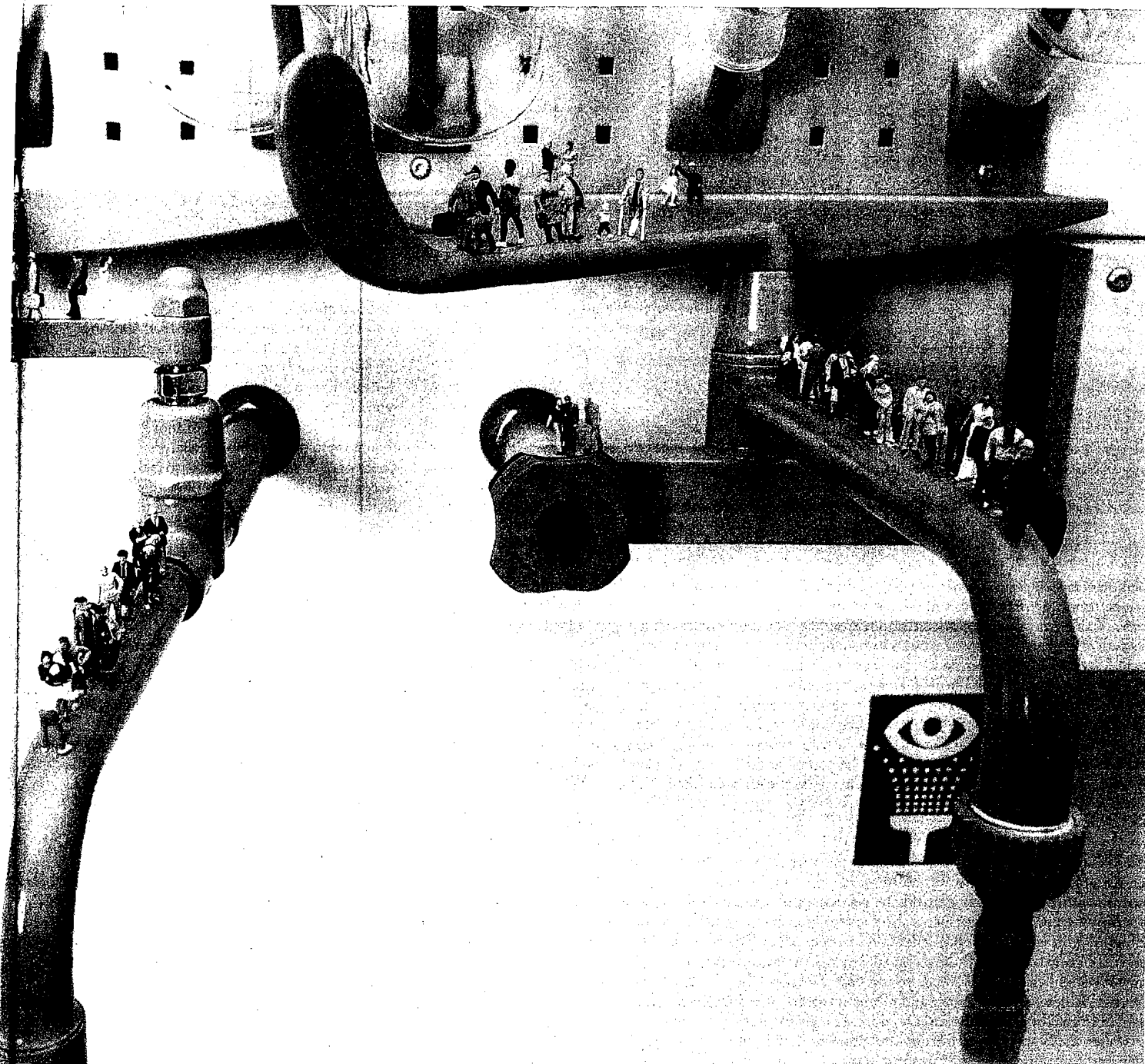
These limitations mean that the job of establishing the long-term and relatively rare risks of drug therapies has fallen to observational studies, as has the job of determining the risks and benefits of virtually all factors of diet and lifestyle that might be related to chronic diseases. The former has been a fruitful field of research; many side effects of drugs have been discovered by these observational studies. The latter is the primary point of contention.

While the tools of epidemiology — comparisons of populations with and without a disease — have proved effective over the centuries in establishing that a disease like cholera is caused by contaminated water, as the British physician John Snow demonstrated in the 1850s, it's a much more complicated endeavor when those same tools are employed to elucidate the more subtle causes of chronic disease.

And even the success stories taught in epidemiology classes to demonstrate the historical richness and potential of the field — that pellagra, a disease that can lead to dementia and death, is caused by a nutrient-deficient diet, for instance, as Joseph Goldberger demonstrated in the 1910s — are only known to be successes because the initial hypotheses were subjected to rigorous tests and happened to survive them. Goldberger tested the competing hypothesis, which posited that the disease was caused by an infectious agent, by holding what he called "filth parties," injecting himself and seven volunteers, his wife among them, with the blood of pellagra victims. They remained healthy, thus doing a compelling, if somewhat revolting, job of refuting the alternative hypothesis.

Smoking and lung cancer is the emblematic success story of chronic-disease epidemiology. But lung cancer was a rare disease before cigarettes became widespread, and the association between smoking and



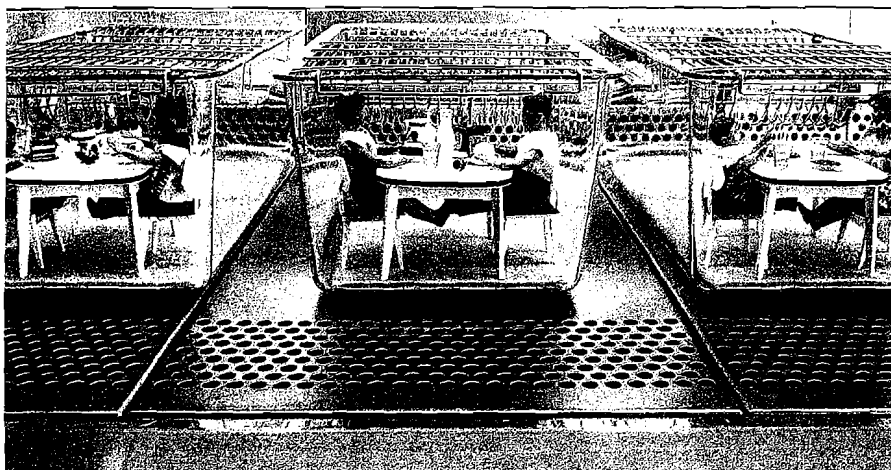


lung cancer was striking: heavy smokers had 2,000 to 3,000 percent the risk of those who had never smoked. This made smoking a “turkey shoot,” says Greenland of U.C.L.A., compared with the associations epidemiologists have struggled with ever since, which fall into the tens of a percent range. The good news is that such small associations, even if causal, can be considered relatively meaningless for a single individual. If a 50-year-old woman with a small risk of breast cancer takes H.R.T. and increases her risk by 30 percent, it remains a small risk.

The compelling motivation for identifying these small effects is that their impact on the public health can be enormous if they’re aggregated

over an entire nation: if tens of millions of women decrease their breast cancer risk by 30 percent, tens of thousands of such cancers will be prevented each year. In fact, between 2002 and 2004, breast cancer incidence in the United States dropped by 12 percent, an effect that may have been caused by the coincident decline in the use of H.R.T. (And it may not have been. The coincident reduction in breast cancer incidence and H.R.T. use is only an association.)

Saving tens of thousands of lives each year constitutes a powerful reason to lower the standard of evidence needed to suggest a cause-and-effect relationship — to take a leap of faith. This is the crux of the issue. From a scientific



T
he wonder and horror

perspective, epidemiologic studies may be incapable of distinguishing a small effect from no effect at all, and so caution dictates that the scientist refrain from making any claims in that situation. From the public-health perspective, a small effect can be a very dangerous or beneficial thing, at least when aggregated over an entire nation, and so caution dictates that action be taken, even if that small effect might not be real. Hence the public-health logic that it's better to err on the side of prudence even if it means persuading us all to engage in an activity, eat a food or take a pill that does nothing for us and ignoring, for the moment, the possibility that such an action could have unforeseen harmful consequences. As Greenland says, "The combination of data, statistical methodology and motivation seems a potent anesthetic for skepticism."

The Bias of Healthy Users

The Nurses' Health Study was founded at Harvard in 1976 by Frank Speizer, an epidemiologist who wanted to study the long-term effects of oral contraceptive use. It was expanded to include postmenopausal estrogen therapy because both treatments involved long-term hormone use by millions of women, and nobody knew the consequences. Speizer's assistants in this endeavor, who would go on to become the most influential epidemiologists in the country, were young physicians — Charles Hennekens, Walter Willett, Meir Stampfer and Graham Colditz — all interested in the laudable goal of preventing disease more than curing it after the fact.

When the Nurses' Health Study first published its observations on estrogen and heart disease in 1985, it showed that women taking estrogen therapy had only a third the risk of having a heart attack as had women who had never taken it; the association seemed compelling evidence for a cause and effect. Only 90 heart attacks had been reported among the 32,000 postmenopausal nurses in the study, and Stampfer, who had done the bulk of the analysis, and his colleagues "considered the possibility that the apparent protective effect of estrogen could be attributed to some other factor associated with its use." They decided, though, as they have ever since, that this was unlikely. The paper's ultimate conclusion was that "further work is needed to define the optimal type, dose and duration of postmenopausal hormone use" for maximizing the protective benefit.

Only after Stampfer and his colleagues published their initial report on estrogen therapy did other investigators begin to understand the nature of the other factors that might explain the association. In 1987, Diana Petitti,

an epidemiologist now at the University of Southern California, reported that she, too, had detected a reduced risk of heart-disease deaths among women taking H.R.T. in the Walnut Creek Study, a population of 16,500 women. When Petitti looked at all the data, however, she "found an even more dramatic reduction in death from homicide, suicide and accidents." With little reason to believe that estrogen would ward off homicides or accidents, Petitti concluded that something else appeared to be "confounding" the association she had observed. "The same thing causing this obvious spurious association might also be contributing to the lower risk of coronary heart disease," Petitti says today.

That mysterious something is encapsulated in what epidemiologists call the healthy-user bias, and some of the most fascinating research in observational epidemiology is now aimed at understanding this phenomenon in all its insidious subtlety. Only then can epidemiologists learn how to filter out the effect of this healthy-user bias from what might otherwise appear in their studies to be real causal relationships. One complication is that it encompasses a host of different and complex issues, many or most of which might be impossible to quantify. As Jerry Avorn of Harvard puts it, the effect of healthy-user bias has the potential for "big mischief" throughout these large epidemiologic studies.

At its simplest, the problem is that people who faithfully engage in activities that are good for them — taking a drug as prescribed, for instance, or eating what they believe is a healthy diet — are fundamentally different from those who don't. One thing epidemiologists have established with certainty, for example, is that women who take H.R.T. differ from those who don't in many ways, virtually all of which associate with lower heart-disease risk: they're thinner; they have fewer risk factors for heart disease to begin with; they tend to be more educated and wealthier; to exercise more; and to be generally more health conscious.

Considering all these factors, is it possible to isolate one factor — hormone-replacement therapy — as the legitimate cause of the small association observed or even part of it? In one large population studied by Elizabeth Barrett-Connor, an epidemiologist at the University of California, San Diego, having gone to college was associated with a 50 percent lower risk of heart disease. So if women who take H.R.T. tend to be more educated than women who don't, this confounds the association between hormone therapy and heart disease. It can give the appearance of cause and effect where none exists.

Another thing that epidemiologic studies have established convincingly is that wealth associates with less heart disease and better health, at least in developed countries. The studies have been unable to establish why this is so, but this, too, is part of the healthy-user problem and a possible confounder of the hormone-therapy story and many of the other associations these epidemiologists try to study. George Davey Smith, who began his career studying how socioeconomic status associates with health, says one thing this research teaches is that misfortunes "cluster" together. Poverty is a misfortune, and the poor are less edu-

epidemiology,' one researcher says, 'is that it's not enough to just measure one thing very accurately. To get the right answer, you may have to measure a great many things very accurately.'

cated than the wealthy; they smoke more and weigh more; they're more likely to have hypertension and other heart-disease risk factors, to eat what's affordable rather than what the experts tell them is healthful, to have poor medical care and to live in environments with more pollutants, noise and stress. Ideally, epidemiologists will carefully measure the wealth and education of their subjects and then use statistical methods to adjust for the effect of these influences — multiple regression analysis, for instance, as one such method is called — but, as Avorn says, it "doesn't always work as well as we'd like it to."

The Nurses' investigators have argued that differences in socioeconomic status cannot explain the associations they observe with H.R.T. because all their subjects are registered nurses and so this "controls" for variations in wealth and education. The skeptics respond that even if all registered nurses had identical educations and income, which isn't necessarily the case, then their socioeconomic status will be determined by whether they're married, how many children they have and their husbands' income. "All you have to do is look at nurses," Petitti says. "Some are married to C.E.O.'s of corporations and some are not married and still living with their parents. It cannot be true that there is no socioeconomic distribution among nurses." Stampfer says that since the Women's Health Initiative results came out in 2002, the Nurses' Health Study investigators went back into their data to examine socioeconomic status "to the extent that we could" — looking at measures that might indirectly reflect wealth and social class. "It doesn't seem plausible" that socioeconomic status can explain the association they observed, he says. But the Nurses' investigators never published that analysis, and so the skeptics have remained unconvinced.

The Bias of Compliance

A still more subtle component of healthy-user bias has to be confronted. This is the compliance or adherer effect. Quite simply, people who comply with their doctors' orders when given a prescription are different and healthier than people who don't. This difference may be ultimately unquantifiable. The compliance effect is another plausible explanation for many of the beneficial associations that epidemiologists commonly report, which means this alone is a reason to wonder if much of what we hear about what constitutes a healthful diet and lifestyle is misconceived.

The lesson comes from an ambitious clinical trial called the Coronary Drug Project that set out in the 1970s to test whether any of five different drugs might prevent heart attacks. The subjects were some 8,500 middle-aged men with established heart problems. Two-thirds of them were randomly assigned to take one of the five drugs and the other third a placebo. Because one of the drugs, clofibrate, lowered cholesterol levels, the researchers had high hopes that it would ward off heart disease. But when the results were tabulated after five years, clofibrate showed no beneficial effect. The researchers then considered the possibility that clofibrate appeared to fail only because the subjects failed to faithfully take their prescriptions.

As it turned out, those men who said they took more than 80 percent of the pills prescribed fared substantially better than those who didn't. Only

15 percent of these faithful "adherers" died, compared with almost 25 percent of what the project researchers called "poor adherers." This might have been taken as reason to believe that clofibrate actually did cut heart-disease deaths almost by half, but then the researchers looked at those men who faithfully took their placebos. And those men, too, seemed to benefit from adhering closely to their prescription: only 15 percent of them died compared with 28 percent who were less conscientious. "So faithfully taking the placebo cuts the death rate by a factor of two," says David Freedman, a professor of statistics at the University of California, Berkeley. "How can this be? Well, people who take their placebo regularly are just different than the others. The rest is a little speculative. Maybe they take better care of themselves in general. But this compliance effect is quite a big effect."

The moral of the story, says Freedman, is that whenever epidemiologists compare people who faithfully engage in some activity with those who don't — whether taking prescription pills or vitamins or exercising regularly or eating what they consider a healthful diet — the researchers need to account for this compliance effect or they will most likely infer the wrong answer. They'll conclude that this behavior, whatever it is, prevents disease and saves lives, when all they're really doing is comparing two different types of people who are, in effect, incomparable.

This phenomenon is a particularly compelling explanation for why the Nurses' Health Study and other cohort studies saw a benefit of H.R.T. in current users of the drugs, but not necessarily in past users. By distinguishing among women who never used H.R.T., those who used it but then stopped and current users (who were the only ones for which a consistent benefit appeared), these observational studies may have inadvertently focused their attention specifically on, as Jerry Avorn says, the "Girl Scouts in the group, the compliant ongoing users, who are probably doing a lot of other preventive things as well."


How Doctors Confound the Science

Another complication to what may already appear (for good reason) to be a hopelessly confusing story is what might be called the prescriber effect. The reasons a physician will prescribe one medication to one patient and another or none at all to a different patient are complex and subtle. "Doctors go through a lot of different filters when they're thinking about what kind of drug to give to what kind of person," says Avorn, whose group at Harvard has spent much of the last decade studying this effect. "Maybe they give the drug to their sickest patients; maybe they give it to the people for whom nothing else works."

It's this prescriber effect, combined with what Avorn calls the eager-patient effect, that is one likely explanation for why people who take cholesterol-lowering drugs called statins appear to have a greatly reduced risk of dementia and death from all causes compared with people who don't take statins. The medication itself is unlikely to be the primary cause in either case, says Avorn, because the observed associations are "so much larger than the effects that have been seen in randomized-clinical trials."

If we think like physicians, Avorn explains, then we get a plausible explanation: "A physician is not going to take somebody either dying of metastatic cancer or in a persistent vegetative state or with end-stage neurologic disease and say, 'Let's get that cholesterol'"

Continued on Page 74

 **STUDY GUIDE** Gary Taubes answers your questions about how to make sense of health news and research. nytimes.com/magazine

down, Mrs. Jones.' The consequence of that, multiplied over tens of thousands of physicians, is that many people who end up on statins are a lot healthier than the people to whom these doctors do not give statins. Then add into that the people who come to the doctor and say, 'My brother-in-law is on this drug,' or, 'I saw it in a commercial,' or, 'I want to do everything I can to prevent heart disease, can I now have a statin, please?' Those kinds of patients are very different from the patients who don't come in. The coup de grâce then comes from the patients who consistently take their medications on an ongoing basis, and who are still taking them two or three years later. Those people are special and unusual and, as we know from clinical trials, even if they're taking a sugar pill they will have better outcomes."

The trick to successfully understanding what any association might really mean, Avorn adds, is "being clever." "The whole point of science is self-doubt," he says, "and asking could there be another explanation for what we're seeing."

H.R.T. and the Plausibility Problem

Until the HERS and W.H.I. trials tested and refuted the hypothesis that hormone-replacement therapy protected women against heart disease, Stampfer, Willett and their colleagues argued that these alternative explanations could not account for what they observed. They had gathered so much information about their nurses, they said, that it allowed them to compare nurses who took H.R.T. and engaged in health-conscious behaviors against women who didn't take H.R.T. and appeared to be equally health-conscious. Because this kind of comparison didn't substantially change the size of the association observed, it seemed reasonable to conclude that the association reflected the causal effect of H.R.T. After the W.H.I. results were published, says Stampfer, their faith was shaken, but only temporarily. Clinical trials, after all, also have limitations, and so the refutation of what was originally a simple hypothesis — that H.R.T. wards off heart disease — spurred new hypotheses, not quite so simple, to explain it.

At the moment, at least three plausible explanations exist for the discrepancy between the clinical trial results and those of the Nurses' Health Study and other observational studies. One is that the associations perceived by the epidemiologic studies were due to healthy-user and prescriber effects and not H.R.T. itself. Women who took H.R.T. had less heart disease than women who didn't, because women who took H.R.T. are different from women who didn't take H.R.T. And maybe their physicians are also different. In this case, the trials got the right answer; the observational studies got the wrong answer.

A second explanation is that the observational studies got the wrong answer, but only partly. Here, healthy-user and prescriber effects are

viewed as minor issues; the question is whether observational studies can accurately determine if women were really taking H.R.T. before their heart attacks. This is a measurement problem, and one conspicuous limitation of all epidemiology is the difficulty of reliably assessing whatever it is the investigators are studying: not only determining whether or not subjects have really taken a medication or consumed the diet that they reported, but whether their subsequent diseases were correctly diagnosed. "The wonder and horror of epidemiology," Avorn says, "is that it's not enough to just measure one thing very accurately. To get the right answer, you may have to measure a great many things very accurately."

The most meaningful associations are those in which all the relevant factors can be ascertained reliably. Smoking and lung cancer, for instance. Lung cancer is an easy diagnosis to make, at least compared with heart disease. And "people sort of know whether they smoke a full pack a day or half or what have you," says Graham Colditz, who recently left the Nurses' study and is now at Washington University School of Medicine in St. Louis. "That's one of the easier measures you can get." Epidemiologists will also say they believe in the associations between LDL cholesterol, blood pressure and heart disease, because these biological variables are measured directly. The measurements don't require that the study subjects fill out a questionnaire or accurately recall what their doctors may have told them.

Even the way epidemiologists frame the questions they ask can bias a measurement and produce an association that may be particularly misleading. If researchers believe that physical activity protects against chronic disease and they ask their subjects how much leisure-time physical activity they do each week, those who do more will tend to be wealthier and healthier, and so the result the researchers get will support their preconceptions. If the questionnaire asks how much physical activity a subject's job entails, the researchers might discover that the poor tend to be more physically active, because their jobs entail more manual labor, and they tend to have more chronic diseases. That would appear to refute the hypothesis.

The simpler the question or the more objective the measurement the more likely it is that an association may stand in the causal pathway, as these researchers put it. This is why the question of whether hormone-replacement therapy effects heart-disease risk, for instance, should be significantly easier to nail down than whether any aspect of diet does. For a measurement "as easy as this," says Jamie Robins, a Harvard epidemiologist, "where maybe the confounding is not horrible, maybe you can get it right." It's simply easier to imagine that women who have taken estrogen therapy will remember and report that correctly — it's yes or no, after all — than that they will recall and report accurately what they ate and how much of it over the last week or the last year.

But as the H.R.T. experience demonstrates, even the timing of a yes-or-no question can

introduce problems. The subjects of the Nurses' Health Study were asked if they were taking H.R.T. every two years, which is how often the nurses were mailed new questionnaires about their diets, prescription drug use and whatever other factors the investigators deemed potentially relevant to health. If a nurse fills out her questionnaire a few months before she begins taking H.R.T., as Colditz explains, and she then has a heart attack, say, six months later, the Nurses' study will classify that nurse as "not using" H.R.T. when she had the heart attack.

As it turns out, 40 percent of women who try H.R.T. stay on it for less than a year, and most of the heart attacks recorded in the W.H.I. and HERS trials occurred during the first few years that the women were prescribed the therapy. So it's a reasonable possibility that the Nurses' Health Study and other observational studies misclassified many of the heart attacks that occurred among users of hormone therapy as occurring among nonusers. This is the second plausible explanation for why these epidemiologic studies may have erroneously perceived a beneficial association of hormone use with heart disease and the clinical trials did not.

In the third explanation, the clinical trials and the observational studies both got the right answer, but they asked different questions. Here the relevant facts are that the women who took H.R.T. in the observational studies were mostly younger women going through menopause. Most of the women enrolled in the clinical trials were far beyond menopause. The average age of the women in the W.H.I. trial was 63 and in HERS it was 67. The primary goal of these clinical trials was to test the hypothesis that H.R.T. prevented heart disease. Older women have a higher risk of heart disease, and so by enrolling women in their 60s and 70s, the researchers didn't have to wait nearly as long to see if estrogen protected against heart disease as they would have if they only enrolled women in their 50s.

This means the clinical trials were asking what happens when older women were given H.R.T. years after menopause. The observational studies asked whether H.R.T. prevented heart disease when taken by younger women near the onset of menopause. A different question. The answer, according to Stampfer, Willett and their colleagues, is that estrogen protects those younger women — perhaps because their arteries are still healthy — while it induces heart attacks in the older women whose arteries are not. "It does seem clear now," Willett says, "that the observational studies got it all right. The W.H.I. also got it right for the question they asked: what happens if you start taking hormones many years after menopause? But that is not the question that most women have cared about."

This last explanation is now known as the "timing" hypothesis, and it certainly seems plausible. It has received some support from analyses of small subsets of the women enrolled in the W.H.I. trial, *Continued on Page 78*



COLDWELL BANKER

PREVIEWS

INTERNATIONAL®



RIVERSIDE, CT \$12,750,000
Beautiful water views, tranquility, a perfect 1.7-acre English garden and a private beach at street's end characterize this desirable Riverside setting for a lovely vintage stone Cotswold. Classic details blend with today's amenities.

Linda Collins

(203) 622-1100



WESTPORT, CT \$4,999,999
Walk to Compo Beach and Westport train station from this remarkable new 6,600 sq. ft. Hampton single-style home. A luxurious classic, the 5-bedroom residence impeccably blends timeless warmth and flowing, grand design.

Judy Michaelis

(203) 227-8424



GUILFORD, CT \$4,900,000
A uniquely dramatic private retreat set within almost 2 acres of rare seaside solitude in Sachems Head, this 6,200 sq. ft. 6-bedroom celebration of contemporary design enjoys magnificent sunsets, glorious gardens and a pool.

Joe Piscitelli

(203) 878-7424



DOBBS FERRY, NY \$4,495,000
Built in 1917, this 9-bedroom, 7,000 sq. ft. Georgian includes a carriage house, pool, and a meandering stream that rest within 2.4 garden acres overlooking country club grounds.

T. Fitzpatrick/H. Troubetsky (914) 693-5476



IRVINGTON, NY \$3,550,000
Breathtaking 21.58-acre estate offers charm in a gracious 1753 landmark colonial. Glorious grounds include a pond, granite pool, barn, artist's studio, two-bedroom apartment, rustic cabin and extraordinary perennial gardens.

Therese Militana Vahron (914) 693-5476



NEW CANAAN, CT \$3,400,000
Embracing beautiful panoramic vistas from a scenic 401-acre hilltop, this superbly detailed modern Craftsman residence encompasses 5,800 sq. ft. of unparalleled architectural drama with balconies, terraces, pool and studio.

Ruth LeBlanc-Jones (203) 966-8994

For further information regarding the Previews International Luxury Homes Program, Please call: 800-767-0078

Coldwell Banker Residential Brokerage

ColdwellBankerMoves.com

© 2007 Coldwell Banker Real Estate LLC. Coldwell Banker® is a registered trademark licensed to Coldwell Banker Real Estate LLC. An Equal Opportunity Company. Equal Housing Opportunity. Owned and Operated by NRT Incorporated. We are pledged to the letter and spirit of U.S. policy for the achievement of equal housing opportunity throughout the Nation. We encourage and support an affirmative advertising and marketing program in which there are no barriers to obtaining housing because of race, color, religion, sex, handicap, familial status or national origin.



Country Living In Virginia



Round Hill Farm

119 acres less than 10 minutes from town with frontage on the Rivanna Reservoir and beautiful landscaped grounds. The spacious home features beautiful moldings, parquet floor and gorgeous mountain views as well as a pool and farm buildings.



Emerald Hill

Private park-like setting of 133 acres just west of Charlottesville with dramatic views and significant custom brick manor. Dependencies include a guest house, log home, office, pool and tennis court.



Airlie

509 acres just east of Charlottesville with a grand, meticulously restored manor, numerous farm buildings, guest houses, and board fencing for horses or cattle. This property is in one of Central Virginia's most desirable locations.



Dover Hall, Manakin Sabot

This magnificent 30,000 s.f. manor on 55 acres offers formal and informal spaces ideal for entertaining with grand ballroom and 9 bedrooms suites and an 8-car detached garage. Near Kinloch Golf Club and Hermitage Country Club.

FRANK HARDY, INC., REALTORS

(434) 296-0134

www.farmandestate.com

Equal Housing Opportunity

Continued from Page 74

like the study published in June in The New England Journal of Medicine. The dilemma at the moment is that the first two explanations are also plausible. If the compliance effect can explain why anyone faithfully following her doctor's orders will be 50 percent less likely to die over the next few years than someone who's not so inclined, then it's certainly possible that what the Nurses' Health Study and other observational studies did is observe a compliance effect and mistake it for a beneficial effect of H.R.T. itself. This would also explain why the Nurses' Health Study observed a 40 percent reduction in the yearly risk of death from all causes among women taking H.R.T. And it would explain why the Nurses' Health Study reported very similar seemingly beneficial effects for antioxidants, vitamins, low-dose aspirin and folic acid, and why these, too, were refuted by clinical trials. It's not necessarily true, but it certainly could be.

While Willert, Stampfer and their colleagues will argue confidently that they can reasonably rule out these other explanations based on everything they now know about their nurses — that they can correct or adjust for compliance and prescriber effects and still see a substantial effect of H.R.T. on heart disease — the skeptics argue that such confidence can never be justified without a clinical trial, at least not when the associations being studied are so small. "You can correct for what you can measure," says Rory Collins, an epidemiologist at Oxford University, "but you can't measure these things with precision so you will tend to under-correct for them. And you can't correct for things that you can't measure."

The investigators for the Nurses' Health Study "tend to believe everything they find," says Barrett-Connor of the University of California, San Diego. Barrett-Connor also studied hormone use and heart disease among a large group of women and observed and published the same association that the Nurses' Health Study did. She simply does not find the causal explanation as easy to accept, considering the plausibility of the alternatives. The latest variation on the therapeutic wisdom on

hormones is plausible, she says, but it remains untested. "Now we're back to the place where we're stuck with observational epidemiology," she adds. "I'm back to the place where I doubt everything."

What to Believe?

So how should we respond the next time we're asked to believe that an association implies a cause and effect, that some medication or some facet of our diet or lifestyle is either killing us or making us healthier? We can fall back on several guiding principles, these skeptical epidemiologists say. One is to assume that the first report of an association is incorrect or meaningless, no matter how big that association might be. After all, it's the first claim in any scientific endeavor that is most likely to be wrong. Only after that report is made public will the authors have the opportunity to be informed by their peers of all the many ways that they might have simply misinterpreted what they saw. The regrettable reality, of course, is that it's this first report that is most newsworthy. So be skeptical.

If the association appears consistently in study after study, population after population, but is small — in the range of tens of percent — then doubt it. For the individual, such small associations, even if real, will have only minor effects or no effect on overall health or risk of disease. They can have enormous public-health implications, but they're also small enough to be treated with suspicion until a clinical trial demonstrates their validity.

If the association involves some aspect of human behavior, which is, of course, the case with the great majority of the epidemiology that attracts our attention, then question its validity. If taking a pill, eating a diet or living in proximity to some potentially noxious aspect of the environment is associated with a particular risk of disease, then other factors of socioeconomic status, education, medical care and the whole gamut of healthy-user effects are as well. These will make the association, for all practical purposes, impossible to interpret reliably.

The exception to this rule is unexpected harm, what Avorn calls "bolt from the blue events," that no one, not the epidemiologists, the subjects or their physi-



CHARLESTON, SOUTH CAROLINA

Located within the City of Charleston, Daniel Island is a 4,000-acre island town featuring traditional neighborhoods, world-class sports and a vibrant downtown that is home to schools, churches, medical offices, boutiques and restaurants. Access to hundreds of acres of parks, 23 miles of rivers and creeks, leisure trails, and a private country club, with 2 golf courses designed by Rees Jones and Tom Fazio, provide a myriad of recreational options. This lakefront residence has stunning views and luxurious appointments.

2025 Purcell Lane | \$2,297,700



danielisland.com | 800-958-5635

Daniel Island Real Estate, LLC
Equal Housing Opportunity



1920's Gentleman's Estate on 20,000 private acres in Carmel, CA

Santa Lucia Preserve, a rare combination of expansive natural beauty and an intimate community of 300 families. World class amenities include a private Fazio golf course and Clubhouse.

The Hacienda for lodging and dining, 100 miles of private riding and hiking trails, An Equestrian Center and Sports Center all next to Carmel and Pebble Beach.

Estate sites are offered from \$1,800,000.

877-626-8200

www.santaluciapreserve.com

EQUAL HOUSING OPPORTUNITIES



Bedford, N.Y.
FOUR LANTERNS Over four private acres in the Buxton Road estate area. Long drive to hilltop setting. Gracious Shingle Colonial. Wide board oak flrs, French drs and front & rear porches. Entrance Hall with French drs to Living Rm with Fpl. Formal Dining Rm with French drs & raised paneling. Country Kit with honed granite, subway tile and stainless appliances. Open Family Rm. Den. Four Bedrooms + Au Pair Suite. Free-form Pool. Just listed! \$1,650,000
(914) 234-9234 **Ginnel** Equal Housing Opportunity



WILLIAMSBURG, VA.—GATED BEACHFRONT ESTATES
Virginia's Most Exclusive Community. Each estate at Shores of York has a huge beach, pier & 3 mile views of the York River. Only 9 beachfront estates on 70 manicured acres. \$1.5 to \$1.7 million. 804-798-0000.

VIDEO TOUR WWW.SHORESOFYORK.COM

cians, could possibly have seen coming — higher rates of vaginal cancer, for example, among the children of women taking the drug DES to prevent miscarriage, or mesothelioma among workers exposed to asbestos. If the subjects are exposing themselves to a particular pill or a vitamin or eating a diet with the goal of promoting health, and, lo and behold, it has no effect or a negative effect — it's associated with an increased risk of some disorder, rather than a decreased risk — then that's a bad sign and worthy of our consideration, if not some anxiety. Since healthy-user effects in these cases work toward reducing the association with disease, their failure to do so implies something unexpected is at work.

All of this suggests that the best advice is to keep in mind the law of unintended consequences. The reason clinicians test drugs with randomized trials is to establish whether the hoped-for benefits are real and, if so, whether there are unforeseen side effects that may outweigh the benefits. If the implication of an epidemiologist's study is that some drug or diet will bring us improved prosperity and health, then wonder about the unforeseen consequences. In these cases, it's never a bad idea to remain skeptical until somebody spends the time and the money to do a randomized trial and, contrary to much of the history of the endeavor to date, fails to refute it. ■

ANSWERS TO PUZZLES

OF SEPTEMBER 9, 2007

PROCESS OF ELIMINATION

Q	A	N	D	A	S	P	A	S	M	D	A	T	U	M	M	E	G
E	Q	U	S	C	A	N	T	O	I	R	A	T	E	A	D	O	
D	U	N	S	T	H	I	T	U	P	S	O	R	E	L	O	S	E
A	N	T	I	P	E	R	S	P	I	R	A	N	T	S	L	A	N
R	A	M	Y	E	S	E	S			R	A	D	I	I			
A	S	T	A	R	F	H	A	T	C	H	E	T	F	A	C	E	
S	H	A	D	S	T	A	S	H	D	E	L	X	I	E			
H	A	R	D	H	E	A	R	T	E	D	R	E	S	T	L	I	M
K	A	T	I	E	R	O	Y	A	L	O	S	O	L	E	M	I	O
S	E	N	O	R	S	A	L	B	O	Z	E	R	O	D	I	N	G
H	A	T	I	F	T	H	E	S	H	O	E	F	I	T	S	T	O
A	L	U	M	N	A	E	S	E	A	R	E	N	S	P	A	R	S
D	E	L	U	G	I	N	G										
E	G	A	D	S	C	A	N										
W	A	I	T	O	N												
S	H	E	S	A	L	L	T	H	A	T							
T	O	N	E	D													
I	N	T	L														
N	O	O	F	F	E	N	S	E									
G	U	M															
O	R	B															

A	L	T	O	A	M	E	S										
S	T	R	E	E	T	S	C	E	N	E							
S	E	P	T	E	M	B	E	R	S	O	N	G					
B	A	N	K	E	R												
Q	U	I	D														
H	U	R	L														
R	I	B		L	I	S	A										
S	T	O	N	E	S	O	U	P									
S	T	E	V	E	N	S	P	I	E	L	B	E	R	G			
S	T	I	E	S													
N	E	A	T														
S	K	I															
O	I	L															
S	A	L	S	O	D	A											
S	U	M	M	E	R	S	A	U	S	A	G	E					
S	O	O	N	E	R	S	T	A	T	E							
E	G	O	S														

SHAKER HEIGHTS

Cleveland's most prestigious suburb... Ohio



Exquisite, affordable homes on tree-lined streets.

- Cosmopolitan living
- Great value for your dollar
- 3,000-sq. feet from \$200,000

Visit

shakeronline.com NYT icon

SHOPPING HOME

Custom Table Pads

9-Day Delivery Available

Guaranteed quality and service from America's premiere table pad company. Free measuring service available in most areas. Satisfaction guaranteed with our return privilege. Backed by Sentry's 30-year limited warranty.

1-800-328-7237 ext.525
www.sentrytablepad.com

Breuer Classic. Natural, honey, teak, walnut, black, or white with blond cane. One piece chrome frame. Floor glides. Side Chair \$129 Back \$39

Arm Chair \$139

Breuer replacement seat. 16 3/8" x 18 3/8" w. brochure available

homeward

National Ordering 800-616-3667
Fax 202-526-5679 • 10-5 M-F (ET) • VISA • MC • AE
www.homewardfurniture.com *shipping

Great trips
start here.

Our suggested randomised design to evaluate drugs courts has limitations. Offenders who enter no or a delayed plea (about 20%) are not eligible. How judges are chosen for the drugs court also matters when extrapolating from the randomised evidence-base: fewer, and perhaps more senior, judges may preside in a drugs court than in conventional courts, where either a judge or magistrates sit.

Medicines of proven efficacy are only provided to patients in the UK's National Health Service if the drugs are cost effective. UK justice needs equivalent appraisal in its use of public funds. Speedier and more effective sentencing in drugs courts—in terms of reduced recidivism—might offset greater organisational and judicial costs. Additionally, recovery from drug dependency may save injectors' lives or reduce claims for welfare benefits. The a-priori case for affordability, and hence for evaluation, would need to be determined and be explicit in the study protocol.

Finding out about effectiveness for recidivism would mean randomising 700–900 offenders to have 80% power to discern a reduction in 2-year recidivism from 70% to 60%, and ten times as many to discern even a one-third reduction in 2-year mortality from 3% to 2%. Assuming only 180 eligible randomised clients in each of the four jurisdictions where drugs courts are planned, one-third of them assigned to the drugs court, should give answers on 2-year recidivism well within 4 years; answers on mortality would take much longer.

Drugs-court evaluations need the discipline of a well-written protocol. Ministers cannot duck the mathematics of numbers needed to neutralise the play of chance. Criminal justice should stop playing at evaluation, and recognise evidential rigour.¹²

*Sheila M Bird, Elizabeth L C Merrall

MRC Biostatistics Unit, Cambridge CB2 0SR, UK
sheila.bird@mrc-bsu.cam.ac.uk

SMB and ELCM are funded by Medical Research Council programme WBS number U.1052.00.002.00001.01. SMB chairs the Surveys, Design and Statistics Subcommittee of the Home Office's Scientific Advisory Committee, chaired the Royal Statistical Society's Working Party on Performance Monitoring in the Public Services (which called for greater attention to formal experiments, including randomisation), and served on the Appraisal Committee of the National Institute for Health and Clinical Excellence (1999–2005). ELCM declares that she has no conflicts of interest.

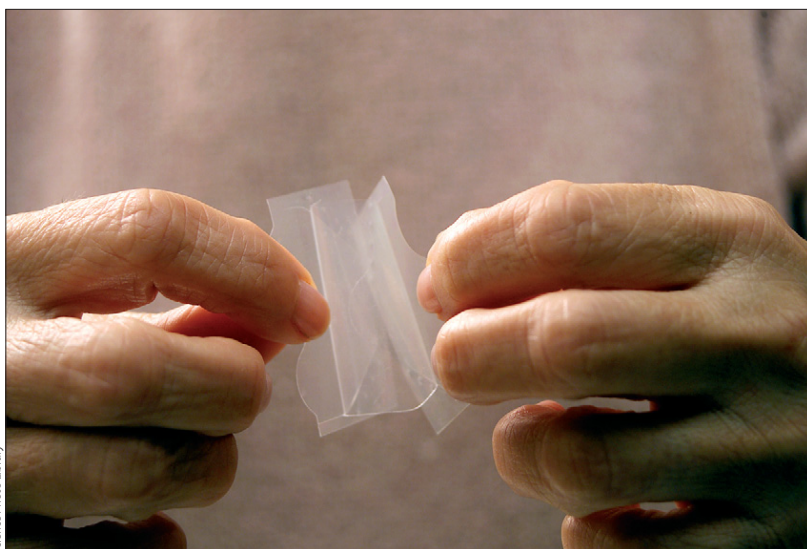
- 1 United Nations Office on Drugs and Crime. Drug treatment courts work! http://www.unodc.org/pdf/drug_treatment_courts_flyer.pdf (accessed Nov 18, 2008).
- 2 McIvor G, Barnsdale L, Malloch M, Eley S, Yates R. The Operation and Effectiveness of the Scottish Drugs Court Pilots. 2006. <http://www.scotland.gov.uk/Resource/Doc/100021/0024203.pdf> (accessed Nov 18, 2008).
- 3 Ministry of Justice. Dedicated drug court pilots: a process report. 2008. <http://www.justice.gov.uk/docs/dedicated-drug-courts.pdf> (accessed Nov 18, 2008).
- 4 Wilson DB, Mitchell O, MacKenzie DL. A systematic review of drug court effects on recidivism. *J Exp Criminol* 2007; **2**: 459–87.
- 5 Bird SM. Prescribing sentence: time for evidence-based justice. *Lancet* 2004; **364**: 1457–59.
- 6 King R, Bird SM, Brooks SP, Hutchinson SJ, Hay G. Prior information in behavioural capture-recapture methods: demographic influences on drug injectors' propensity to be listed in data sources and their drug-related mortality. *Am J Epidemiol* 2005; **162**: 694–703.
- 7 Farrington DP, Welsh BC. A half century of randomized experiments on crime and justice. *Crime Justice* 2006; **34**: 55–132.
- 8 Perera R, Heneghan C, Yudkin P. Graphical method for depicting randomized trials of complex interventions. *BMJ* 2007; **334**: 127–29.
- 9 Deschenes EP, Turner S, Greenwood PW. Drug court or probation: an experimental evaluation of Maricopa County Drug Court. *Justice System J* 1995; **18**: 55–73.
- 10 Surveys, Design and Statistics Subcommittee of Home Office's Scientific Advisory Committee. 21st century drugs and statistical science. Dec 17, 2008. <http://www.homeoffice.gov.uk/documents/science-advisory-committee/21st-century-drugs-stats?view=Binary> (accessed Dec 17, 2008).
- 11 Ministry of Justice. Time intervals for criminal proceedings in magistrates' courts. Aug 21, 2008. <http://www.justice.gov.uk/docs/time-intervals-criminal-proceedings-june-2008.pdf> (accessed Nov 18, 2008).
- 12 Shapland J, Atkinson A, Atkinson H, et al. Does restorative justice affect reconviction? The fourth report from the evaluation of three schemes. June 2008. http://www.justice.gov.uk/restorative-justice-report_06-08.pdf (accessed Nov 18, 2008).

The HRT controversy: observational studies and RCTs fall in line

For several years, we witnessed a disarraying debate about the conflicting messages between observational studies and randomised trials on the effect of hormone replacement therapy (HRT) on coronary heart disease and breast cancer. HRT seemed protective for coronary heart disease in observational studies, but randomised trials found an increase of coronary heart disease in the first years of use.¹ For breast cancer, combined oestrogen-progestin showed a lesser risk

in the large Women's Health Initiative randomised trial than in observational studies such as the Million Women Study.^{2,3} Unopposed oestrogens had a smaller breast cancer risk than combined preparations in observational studies, but carried no risk in the trial.⁴ Observational research suffered a credibility crisis.

Recent reanalyses have brought the results from observational and randomised studies into line. The results



are surprising. Neither design held superior truth. The reasons for the discrepancies were rooted in the timing of HRT and not in differences in study design.

For coronary heart disease, the results of observational data and trials fell in line, mainly by analysing the data according to time since start of HRT.^{1,5,6} For randomised trials, this is the natural analysis because therapy starts at randomisation. In the Women's Health Initiative and other trials, the first years of hormone replacement by combined oestrogen-progestin did increase coronary heart disease, which then waned. The analysis of the observational studies, however, had mostly been a contrast between current users at the time of enrolment to never users. Most current users were past the window wherein coronary heart disease risk was increased and were in a phase of decreased incidence. When cohort data from the observational part of the Women's Health Initiative were reanalysed according to time since start of therapy, the same pattern emerged of an initial increase in risk, followed by a decrease.⁵ Thus nothing was intrinsically wrong with the observational data; what went wrong was an analysis that had not taken into account that the effect of HRT might be different over time. The piece of evidence that closes the case is the recent reanalysis of the Nurses Health Study on combined oestrogen-progestin and coronary heart disease, which finds the same pattern of an initial increase in risk by contrast with the original analysis which showed overall protection.¹ An array of comments followed.⁷⁻¹² Whether the decrease in coronary heart disease on continued use

is due to deletion of susceptible individuals or a causal effect cannot be learned from these analyses.

For breast cancer, women in the randomised trials had on average been in menopause longer; in the observational study, the women had started HRT closer to menopause. Adjustment for previous use of hormones already increased the estimates in the trials, but the findings of observational and randomised studies fell in line when the reanalyses of the randomised trial data adjusted for the gap between menopause and treatment, showing a clear increase in risk for combined preparations and a slight increase for unopposed oestrogens.^{2,4} The observational studies had picked up a true signal for the women closer to menopause. In the randomised trial, that signal was diluted because fewer women close to menopause were enrolled. The signal is important for daily practice, because HRT is usually started close to menopause. Again, the discrepancies were not due to differences in study design, but to the timing of start of treatment relative to menopause.

The randomised trials had it right for coronary heart disease but failed to sufficiently focus on women close to menopause for breast cancer. The main reasons for the discrepancies were changes of the effects of HRT over different times: time from start of therapy and time since menopause. In the reanalyses, adjustments for standard risk factors had some additional effects, but did not clinch the analyses as much as the two principal interactions with time. A lesser effect of time since menopause was also seen for coronary heart disease: longer time since menopause heightened the risk. For coronary heart disease the effects also differed for oestrogen alone and combined preparations.⁶

The results put an end to years of debate about HRT, coronary heart disease, and breast cancer, but also clarify the debate on the merits of randomised versus observational studies. They show that "observational-randomised discrepancies cannot be automatically attributed to randomisation itself".¹ Still, randomised trials will almost always be necessary to show whether the hoped-for benefit of a medical intervention exists. Our knowledge about HRT and coronary heart disease would be different, were it not for the randomised trials, even if on reanalysis the observational data carried the same message.⁹ By contrast, observational research will often suffice to investigate adverse effects.¹³ Rarely,

the same adverse effect for the same treatment can be investigated by observational research and in very large randomised trials,¹⁴ as happened with breast cancer and HRT. These comparisons support the notion that observational studies may better reflect the true harm in real-life prescribing than selected populations enrolled in randomised trials.¹⁴

The resolution of the discrepancies between randomised and observational evidence is not just important for our insight into the merits of both types of research. It directly enlightens our knowledge about HRT by confirming that the cardiovascular risk is real, and slightly stronger in older women, while the breast cancer risk is equally real, and is stronger in women closer to menopause.⁹ It was a long and difficult debate, but we owe a tribute to the persons who inspired and have led these reanalyses.

Jan P Vandenbroucke

Department of Clinical Epidemiology, Leiden University Medical School, 2300 RC Leiden, Netherlands
j.p.vandenbroucke@lumc.nl

I thank Miguel Hernán and Frans Helmerhorst for comments on an earlier version. I declare that I have no conflicts of interest.

- 1 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**: 766–79.

- 2 Prentice RL, Chlebowski RT, Stefanick ML, et al. Estrogen plus progestin therapy and breast cancer in recently postmenopausal women. *Am J Epidemiol* 2008; **167**: 1207–16.
- 3 Reeves GK, Beral V, Green J, Gathani T, Bull D, and for the Million Women Study Collaborators. Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *Lancet Oncol* 2006; **7**: 910–18.
- 4 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**: 1407–15.
- 5 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**: 404–14.
- 6 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**: 589–99.
- 7 Hoover RN. The sound and the fury: was it all worth it? *Epidemiology* 2008; **19**: 780–82.
- 8 Stampfer MJ. ITT for observational data: worst of both worlds? *Epidemiology* 2008; **19**: 783–84.
- 9 Prentice RL. Data analysis methods and the reliability of analytic epidemiologic research. *Epidemiology* 2008; **19**: 785–88.
- 10 Hernán MA, Robins JM. Authors' response: part I. Observational studies analyzed like randomized experiments: best of both worlds. *Epidemiology* 2008; **19**: 789–92.
- 11 Willett WC, Manson JE, Grodstein F. Author's response: part II. *Epidemiology* 2008; **19**: 793.
- 12 Wilcox A, Wacholder S. Observational data and clinical trials: narrowing the gap? *Epidemiology* 2008; **19**: 765.
- 13 Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet*. 2004; **363**: 1728–31.
- 14 Vandenbroucke JP, Psaty BM. Benefits and risks of drug treatments: how to combine the best evidence on benefits with the best data about adverse effects. *JAMA* 2008; **300**: 2417–19.

Sirolimus to replace calcineurin inhibitors? Too early yet

Replacement of ciclosporin and mycophenolate mofetil with sirolimus has been associated with regression of Kaposi's sarcoma in renal transplant recipients on chronic immunosuppression. Disease development parallels reactivation of latent human herpesvirus 8 (HHV-8) or donor-to-recipient transfer of HHV-8-infected progenitor cells. Patrizia Barozzi and colleagues¹ recently reported nine patients with post-transplant Kaposi's sarcoma associated with a lack of HHV-8-specific T cells. In two patients who were switching from calcineurin inhibitors to sirolimus, disease recovery was paralleled by normalisation of the T-cell repertoire and recovery of both HHV-8 specific effector and memory T lymphocytes. Thus sirolimus might achieve remission of Kaposi's sarcoma by restoring a specific immune response against the tumour-associated virus.

Sirolimus is a macrolide with potent immuno-suppressive and antiproliferative activity.² This drug

suppresses interleukin-driven T-cell proliferation by blocking signal-transduction pathways required for the progression of cytokine-stimulated T cells from G₁ to S phase.² Early studies in animals showed that sirolimus, unlike calcineurin inhibitors, was devoid of intrinsic nephrotoxicity.³ Consistently, renal transplant patients on 2-year sirolimus therapy had significantly lower concentrations of serum creatinine than controls on ciclosporin.⁴ This attracted special attention to the use of this powerful immunosuppressant to replace ciclosporin and avoid the nephrotoxicity of chronic calcineurin inhibition.²

Enthusiasm faded, however, when the US Multicenter Trial showed that sirolimus-treated renal transplant recipients had significantly higher serum creatinine than ciclosporin-treated recipients, despite having fewer rejections.⁵ Subsequent studies consistently showed that this effect, first attributed to exacerbation

Observational Studies Analyzed Like Randomized Experiments

An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease

Miguel A. Hernán,^{a,b} Alvaro Alonso,^c Roger Logan,^a Francine Grodstein,^{a,d} Karin B. Michels,^{a,d,e} Walter C. Willett,^{a,d,f} JoAnn E. Manson,^{a,d,g} and James M. Robins^{a,h}

Background: The Women's Health Initiative randomized trial found greater coronary heart disease (CHD) risk in women assigned to estrogen/progestin therapy than in those assigned to placebo. Observational studies had previously suggested reduced CHD risk in hormone users.

Methods: Using data from the observational Nurses' Health Study, we emulated the design and intention-to-treat (ITT) analysis of the randomized trial. The observational study was conceptualized as a sequence of "trials," in which eligible women were classified as initiators or noninitiators of estrogen/progestin therapy.

Results: The ITT hazard ratios (HRs) (95% confidence intervals) of CHD for initiators versus noninitiators were 1.42 (0.92–2.20) for the first 2 years, and 0.96 (0.78–1.18) for the entire follow-up. The ITT HRs were 0.84 (0.61–1.14) in women within 10 years of menopause, and 1.12 (0.84–1.48) in the others (*P* value for interaction = 0.08). These ITT estimates are similar to those from the Women's Health Initiative. Because the ITT approach causes severe treatment misclassification, we also estimated adherence-adjusted effects by inverse probability weighting. The HRs were 1.61 (0.97–2.66) for the first 2 years, and 0.98 (0.66–1.49) for the entire follow-up. The HRs were 0.54 (0.19–1.51) in women within 10 years after menopause, and 1.20 (0.78–1.84) in others (*P* value for interaction = 0.01). We

also present comparisons between these estimates and previously reported Nurses' Health Study estimates.

Conclusions: Our findings suggest that the discrepancies between the Women's Health Initiative and Nurses' Health Study ITT estimates could be largely explained by differences in the distribution of time since menopause and length of follow-up.

(*Epidemiology* 2008;19: 766–779)

Causal inferences are drawn from both randomized experiments and observational studies. When estimates from both types of studies are available, it is reassuring to find that they are often similar.^{1–3} On the other hand, when randomized and observational estimates disagree, it is tempting to attribute the differences to the lack of random treatment assignment in observational studies.

This lack of randomization makes observational effect estimates vulnerable to confounding bias due to the different prognosis of individuals between treatment groups. The potential for confounding may diminish the enthusiasm for other desirable features of observational studies compared with randomized experiments—greater timeliness, less restrictive eligibility criteria, longer follow-up, and lower cost. However, even though randomization is the defining difference between randomized experiments and observational studies, further differences in both design and analysis are commonplace. As a consequence, observational-randomized discrepancies cannot be automatically attributed to randomization itself.

In this paper we assess the extent to which differences other than randomization contribute to discrepant observational versus randomized effect estimates in the well-known example of postmenopausal estrogen plus progestin therapy and the risk of coronary heart disease (CHD). Specifically, we explore discrepancies attributable to different distributions of time since menopause, length of follow-up, and analytic approach.

The published findings on this topic can be briefly summarized as follows. Large observational studies suggested a reduced risk of CHD among postmenopausal hormone users.

Submitted 11 January 2008; accepted 30 May 2008.

From the ^aDepartment of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; ^bHarvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts; ^cDivision of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota; ^dChanning Laboratory, Department of Medicine and ^eObstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ^fDepartment of Nutrition, Harvard School of Public Health, Boston, Massachusetts; ^gDivision of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; and ^hDepartment of Biostatistics, Harvard School of Public Health, Boston, Massachusetts.

This work was supported by National Institutes of Health grants HL080644 and CA87969.

Editors' note: Commentaries on this article appear on pages xxx, xxx, xxx. Correspondence: Miguel Hernán, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA. E-mail: miguel_hernan@post.harvard.edu.

Copyright © 2008 by Lippincott Williams & Wilkins

ISSN: 1044-3983/08/1906-0766

DOI: 10.1097/EDE.0b013e3181875e61

Two of the largest observational studies were based on the Nurses' Health Study (NHS)^{4,5} in the United States and on the General Practice Research Database⁶ in the United Kingdom. More recently, the Women's Health Initiative (WHI) randomized trial⁷ found a greater incidence of coronary heart disease among postmenopausal women in the estrogen plus progestin arm than in the placebo arm (68% greater in the first 2 years after initiation, 24% greater after an average of 5.6 years).^{8,9}

The present paper does not address the complex clinical and public health issues related to hormone therapy, including risk-benefit considerations. Rather, we focus on methodologic issues in the analysis of observational cohort studies. Specifically, we reanalyze the NHS observational data to yield effect estimates of hormone therapy that are directly comparable with those of the randomized WHI trial except for the fact that hormone therapy was not randomly assigned in the NHS. We do this by mimicking the design of the randomized trial as closely as possible in the NHS. As explained below, our approach requires conceptualizing the observational NHS cohort as if it were a sequence of nonrandomized trials. Because the randomized trial data were analyzed under the intention-to-treat (ITT) principle, we analyze our NHS trials using an observational analog of ITT (see below).

A recent reanalysis of the General Practice Research Database using this strategy could not adjust for lifestyle factors and it yielded wide confidence intervals (CI).¹⁰ Further, the estrogen used by women in that study was not the conjugated equine estrogen used by the women in the NHS and WHI studies. Our analysis of the NHS data incorporates lifestyle factors and includes women using the same type of estrogen as in the WHI randomized trial.

METHODS

The Observational Cohort as a Nonrandomized "Trial"

The NHS cohort was established in 1976 and comprised 121,700 female registered nurses from 11 US states, aged 30 to 55 years. Participants have received biennial questionnaires to update information on use, duration (1–4, 5–9, 10–14, 15–19, 20–24 months), and type of hormone therapy during the 2-year interval. Common use of oral estrogen plus progestin therapy among NHS participants began in the period between the 1982 and the 1984 questionnaires. The questionnaires also record information on potential risk factors for and occurrence of major medical events, including CHD (nonfatal myocardial infarction or fatal coronary disease). The process for confirming CHD endpoints has been described in detail elsewhere.⁴

We mimicked the WHI trial by restricting the study population to postmenopausal women who in the 1982 questionnaire had reported no use of any hormone therapy during the prior 2-year period ("washout" period), and in the 1984 questionnaire reported either use of oral estrogen plus progestin

therapy ("initiators") or no use of any hormone therapy ("non-initiators") during the prior 2-year period. Thus, as in the WHI, the initiator group includes both first-time users of hormone therapy and reinitiators (who stopped hormone therapy in 1980 or earlier and then reinitiated use in the period 1982–1984).

Women were followed from the start of follow-up to diagnosis of CHD, death, loss to follow-up, or June 2000, whichever occurred first. Unlike in the randomized WHI and the observational General Practice Research Database, the time of therapy initiation—and thus the most appropriate time of start of follow-up for initiators—was not known with precision in the NHS, and so we needed to estimate it. For women who reported hormone therapy initiation during the 2-year period before the 1984 questionnaire and were still using it at the time they completed this questionnaire, the start of follow-up was estimated as the month of return of the baseline questionnaire minus the duration of hormone therapy use (duration is reported as an interval, eg, 20–24 months; we used the upper limit of the interval, eg, 24 months). For women who reported starting hormone therapy during the same 2-year period but had stopped using it by the time they returned the 1984 questionnaire, the start of follow-up was estimated as the first month of the 2-year period (the earliest possible month of initiation). The start of follow-up for noninitiators was estimated as the average month of start of follow-up among initiators (stratified by age and past use of hormone therapy). Alternative methods to estimate the start of follow-up had little effect on our estimates (Appendix A1).

To further mimic the WHI, we restricted the study population to women who, before the start of follow-up, had a uterus, no past diagnosis of cancer (except nonmelanoma skin cancer) or acute myocardial infarction, and no diagnosis of stroke since the return of the previous questionnaire. To enable adjustment for dietary factors, we restricted the population to women who had reported plausible energy intakes (2510–14,640 kJ/d) and had left fewer than 10 of 61 food items blank on the most recent food frequency questionnaire before the 1984 questionnaire.

The NHS cohort study can now be viewed as a nonrandomized, nonblinded "trial" that mimics the eligibility criteria, definition of start of follow-up, and treatment arms (initiators vs. noninitiators) of the WHI randomized trial, but with a different distribution of baseline risk factors (eg, lower age and shorter time since menopause in the NHS compared with the WHI). We analyzed the NHS nonrandomized "trial" by comparing the CHD risk of initiators and noninitiators regardless of whether these women subsequently stopped or initiated therapy. Thus our analytic approach is the observational equivalent of the ITT principle that guided the main analysis of the WHI trial. Specifically, we estimated the average hazard (rate) ratio (HR) of CHD in initiators versus noninitiators, and its 95% CI, by fitting a Cox proportional hazards model, with "time since beginning of follow-up" as the time variable,

that included a non time-varying indicator for hormone therapy initiation. The Cox model was stratified on age (in 5-year intervals) and history of use of hormone therapy (yes, no).

To obtain valid effect estimates in a nonrandomized trial, all baseline confounders have to be appropriately measured and adjusted for in the analysis. We proceeded as if this condition was at least approximately true in the NHS nonrandomized “trial” once we added the following covariates to the Cox model: parental history of myocardial infarction before age 60 (yes, no), education (graduate degree: yes, no), husband’s education (less than high school, high school graduate, college, graduate school), ethnicity (non-Hispanic white, other), age at menopause (<50, 50–53, >53), calendar time, high cholesterol (yes, no), high blood pressure (yes, no), diabetes (yes, no), angina (yes, no), stroke (yes, no), coronary revascularization (yes, no), osteoporosis (yes, no), body mass index (<23, 23–<25, 25–<30, ≥30), cigarette smoking (never, past, current 1–14 cigarettes per day, current 15–24 cigarettes per day, current ≥25 cigarettes per day), aspirin use (nonuse, 1–4 years, 5–10 years, >10 years), alcohol intake (0, >0–<5, 5–<10, 10–<15, ≥15 g/d), physical activity (6 categories), diet score (quintiles),¹¹ multivitamin use (yes, no), and fruit and vegetable intake (<3, 3–<5, 5–<10, ≥10 servings/d). When available, we simultaneously adjusted for the reported value of each variable on both the 1982 and 1980 questionnaires.

The Observational Cohort as a Sequence of Nonrandomized Nested “Trials”

The approach described above would produce very imprecise ITT estimates if (as was the case) few women were initiators during the 1982–1984 period. However, our choice of this period was arbitrary. The approach described above can produce an additional NHS nonrandomized “trial” when applied to each of the 8 2-year periods between 1982–1984 and 1996–1998. Thus, as a strategy to increase the efficiency of our ITT estimate, we conducted 7 additional nonrandomized “trials” each subsequent questionnaire (1986, 1988, . . . 1998), and pooled all 8 “trials” into a single analysis. Because some women participated in more than one of these NHS “trials” (up to a maximum of 8), we used a robust variance estimator to account for within-person correlation. We assessed the potential heterogeneity of the ITT effect estimates across “trials” by 2 Wald tests: first, we estimated a separate parameter for therapy initiation in each “trial” and tested for heterogeneity of the parameters (χ^2 ; 6 *df*), and then we calculated a product term (for the indicators of “trial” and therapy initiation), testing for whether the product term was different from 0 (χ^2 ; 1 *df*).

In each “trial,” we used the corresponding questionnaire information to apply the eligibility criteria at the start of follow-up, and to define initiators and noninitiators. We then estimated the CHD average HR in initiators versus noninitiators (adjusted for the values of covariates reported in the 2

previous questionnaires), regardless of whether these women subsequently stopped or initiated therapy. To allow for the possibility that the HR varied with time since baseline, we added product terms between time of follow-up (linear and quadratic terms) and initiation status to a pooled logistic model that approximated our previous Cox model. We then used the fitted model to estimate CHD-free survival curves for initiators and noninitiators.

The subset of women considered for eligibility in each “trial” is approximately nested in the subset of women who were considered for eligibility in the prior “trial.” Our conceptualization of an observational study with a time-varying treatment as a sequence of nested “trials,” each with nontime-varying treatment, is a special case of g-estimation of nested structural models.¹²

Several lines of evidence suggest a modification of the effect of hormone therapy by time of initiation.¹³ We therefore conducted stratified analyses by time since menopause (<10, ≥10 years) and age (<60, ≥60 years). We computed *P* values for “interaction” between hormone therapy and years since menopause by adding a single product term (indicator for hormone therapy times indicator for <10 years since menopause) to the model for the overall HR, and then testing the hypothesis that its coefficient was equal to zero. A less powerful alternative strategy, testing for heterogeneity of the HR estimated from separate models for women <10 years and for women >10 years since menopause, resulted in *P* > 0.15 in all analyses.

Adherence-Adjusted Effect Estimates

Because the primary analysis of the WHI randomized trial was conducted under the ITT principle, we analyzed our NHS “trials” using an observational analog of ITT to compare the NHS with the WHI estimates. However, ITT estimates are problematic because the magnitude of the ITT effect varies with the proportion of subjects who adhere to the assigned treatment, and thus ITT comparisons can underestimate the effect that would have been observed if everyone had adhered to the assigned treatment. Thus, ITT effect estimates may be unsatisfactory when studying the efficacy, and inappropriate when studying the safety, of an active treatment compared with no treatment. An alternative to the ITT effect is the effect that would have been observed if everyone had remained on her initial treatment throughout the follow-up, which we refer to as an adherence-adjusted effect. Under additional assumptions, consistent adherence-adjusted effect estimates can be obtained in both randomized experiments and observational studies by using g-estimation^{14,15} or inverse probability weighting.

We used inverse probability weighting to estimate the adherence-adjusted HR of CHD. In each NHS “trial” we censored women when they discontinued their baseline treatment (either hormone therapy or no hormone therapy), and then weighted the uncensored women months by the inverse

of their estimated probability of remaining uncensored until that month.¹⁶ To estimate “trial”-specific probabilities for each woman, we fit a pooled logistic model for the probability of remaining on the baseline treatment through a given month. The model included the baseline covariates used in the “trial”-specific Cox models described previously, and the most recent postbaseline values of the same covariates. Inclusion of time-dependent covariates is necessary to adjust for any dependence between noncompliance and CHD within levels of baseline covariates. We fit separate models for initiators and noninitiators. In each “trial,” each woman contributed as many observations to the model as the number of months she was on her baseline therapy.

To stabilize the inverse probability weights, we multiplied the weights by the probability of censoring given the trial-specific baseline values of the covariates. Weight stabilization improves precision by helping to reduce random variability. If the true adherence-adjusted HR is constant over time, this method produces valid estimates provided that discontinuing the baseline treatment is unrelated to unmeasured risk factors for CHD incidence within levels of the covariates, and that the logistic model used to estimate the inverse probability weights is correctly specified. When the adherence-adjusted HR changes with time since baseline, this method estimates a weighted average adherence-adjusted HR with time-specific weights proportional to the number of uncensored CHD events occurring at each time. Thus, with heavy censoring due to lack of adherence, the early years of follow-up contribute relatively more weight than would be the case without censoring. To more appropriately adjust for a time-varying HR, we also fit an inverse probability weighted Cox model (approximated through a weighted pooled logistic model) that included product terms between time of follow-up (linear and quadratic terms) and initiation status. We then used the weighted model to estimate adherence-adjusted CHD-free survival curves for initiators and noninitiators.

We also present additional subsidiary analyses to explain the relation between our estimates and previously reported NHS estimates, which can be regarded as estimates of the adherence-adjusted HR using an alternative to our inverse probability weighting approach.

RESULTS

The NHS Nonrandomized “Trials”

Of the 101,819 NHS participants alive and without a history of cancer, heart disease, or stroke in 1984, 81,073 had diet information and, of these, 77,794 were postmenopausal at some time during the follow-up. We excluded 14,764 women who received a form of hormone therapy other than oral estrogen plus progestin in all of the NHS “trials,” or did not provide information on the type of hormone therapy in any of the “trials.” Of the remaining 63,030 women, we excluded 17,146 who received hormone therapy in the 2

years before the baseline of all the “trials.” Of the remaining 45,884 women, we excluded 11,309 who did not have an intact uterus in 1984. Thus 34,575 women met our eligibility criteria for at least one NHS “trial.” Of these women, 1035 had a CHD event, 2596 died of other causes or were lost to follow-up, and 30,944 reached June 2000 free of CHD. Figure 1 shows the distribution of women by number of “trials” in which they participated. Table 1 shows the number of participants, initiators, and CHD events per “trial.” Table 2 shows the distribution of baseline characteristics in initiators and noninitiators.

ITT Estimates of the Effect of Hormone Therapy on CHD

The estimated average HR of CHD for initiators versus noninitiators was 0.96 (95% CI = 0.78–1.18) when the entire follow-up time was included in the analysis (Table 3). The HR was 1.83 (1.05–3.17) when the analysis was restricted to the first year of follow-up, 1.42 (0.92–2.20) for the first 2 years, 1.11 (0.84–1.47) for the first 5 years, and 1.00 (0.78–1.28) for the first 8 years. Equivalently, the HR was 0.96 (0.66–1.39) during years 2–5, 0.81 (0.51–1.28) during years 5–8, and 0.87 (0.58–1.30) after year 8. We did not find a strong indication of heterogeneity across trials (Wald tests *P* values 0.24 and 0.15 for the overall HR). Figure 2A shows that the estimated proportion of women free of CHD during the first 5 years of follow-up was lower in initiators of estrogen plus progestin therapy than in noninitiators of hormone therapy. By year 8, however, this proportion was greater in initiators.

We next examined effect modification, stratifying our ITT estimates by age and time since menopause (Table 3). The HR was 0.84 (CI = 0.61–1.14) in women within 10 years of menopause at baseline, and 1.12 (0.84–1.48) in the others (86% of initiators in this latter group initiated therapy 10 to 20 years after menopause). Similarly, the HRs were 0.86 (0.65–1.14) in women under age 60 at baseline, and 1.15 (0.85–1.57) in the others. Figure 2B, C shows the estimated proportion of women free of CHD by initiator status and time since menopause. The *P* value from a log-rank test for the equality of the survival curves was 0.70 for the entire population, 0.27 for women within 10 years of menopause, and 0.43 for the others.

When we repeated the analyses with no past use of hormone therapy as an additional eligibility criterion (26,797

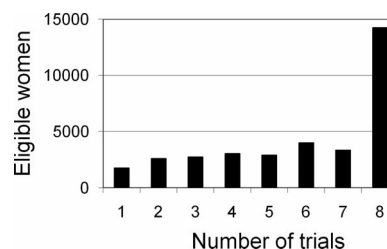


FIGURE 1. Distribution of eligible women by number of Nurses' Health Study “trials” of hormone therapy initiation in which they participated.

TABLE 1. Number of Participants, Therapy Initiators, and CHD Events in Each NHS “Trial” to Estimate the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy

Trial	Questionnaire	Yrs After Menopause	Participants	Initiators	CHD Events	
					Among All Participants	Among Initiators
1	1984	All	16,190	224	772	10
		<10	11,654	201	524	8
		≥10	4536	23	248	2
2	1986	All	17,147	518	671	9
		<10	10,246	416	322	5
		≥10	6901	102	349	4
3	1988	All	18,620	979	610	17
		<10	9550	745	226	8
		≥10	9070	234	384	9
4	1990	All	19,002	1082	528	14
		<10	8014	727	144	5
		≥10	10,988	355	384	9
5	1992	All	19,494	1152	441	13
		<10	7161	714	93	5
		≥10	12,333	438	348	8
6	1994	All	19,954	1344	354	19
		<10	6456	799	69	8
		≥10	13,498	545	285	11
7	1996	All	19,661	1188	228	11
		<10	5508	631	34	3
		≥10	14,153	557	194	8
8	1998	All	18,192	771	100	5
		<10	4287	338	13	2
		≥10	13,905	433	87	3

Based on 34,575 distinct women and 1035 distinct CHD events.

eligible women, 767 CHD events), the HR was 0.79 (CI = 0.60–1.03) for the entire follow-up and 1.49 (0.88–2.54) in the first 2 years (Table 4). The HR was 0.66 (0.44–0.98) in women within 10 years of menopause at baseline, and 1.02 (0.70–1.50) in the others. The appendix includes additional analyses to document the generally small sensitivity of the results regarding the assignment of the month of therapy initiation (Appendix A1), the inclusion of women under age 50 (Appendix A2), the exclusion of women who died between the start of follow-up and the return of the next questionnaire (Appendix A3), the adjustment for confounding by covariates in the proportional hazards model rather than by propensity score methods (Appendix A4), and the assumption of possible unmeasured confounding for therapy discontinuation (Appendix A5).

Adherence-Adjusted Effect Estimates

Figure 3 shows the adherence through year 8 in initiators and noninitiators. The estimated inverse probability weights had mean 1.02 (range = 0.02–30.7) in initiators, and 1.00 (0.17–19.3) in noninitiators. The inverse probability weighted HRs were 0.98 (CI = 0.66–1.49) for the entire follow-up, 1.53 (0.80–2.95) for the first year, 1.61 (0.97–

2.66) for the first 2 years, 1.14 (0.74–1.76) for the first 5 years, and 0.99 (0.66–1.50) for the first 8 years. The HR was 0.65 (0.30–1.38) during years 2 to 5, 0.47 (0.14–1.58), during years 5 to 8, and 0.85 (0.22–3.19) after year 8. The large standard errors that increase with time reflect the fact that few women continued on hormone therapy for long periods. We also examined the effect modification by age and time since menopause (Table 5). Figure 4 shows the estimated adherence-adjusted proportions of women free of CHD. The *P* value from a log-rank test for the equality of the survival curves was 0.91 for the entire population, 0.24 for women within 10 years after menopause, and 0.40 for the others.

Comparison of ITT Estimates With Previous NHS Estimates

The HR estimate of 0.96 from our ITT analysis is not directly comparable with the HR estimate of 0.68 (0.55–0.83) for current users versus never users of estrogen plus progestin reported in the most recent NHS publication.¹⁷ The 0.68 estimate can be interpreted as an adherence-adjusted effect estimate, in which incomplete adherence has been adjusted not by inverse probability weighting but by a comparison of

TABLE 2. Baseline Characteristics of Initiators and Noninitiators of Estrogen/Progestin Therapy in the NHS “Trials”

	Initiators (n = 7258)	Noninitiators (n = 141,002)
Age (y); mean (SD)	59.0 (5.8)	61.9 (6.0)
Age at menopause (y)		
<50	39.2	40.0
50–53	46.6	46.7
>53	14.3	13.4
Years since menopause; mean (SD)	9.0 (5.8)	12.2 (6.3)
<10 y since menopause	63.0	41.4
Past use of hormone therapy	32.0	22.6
Non-Hispanic white	94.6	94.0
Graduate degree	11.7	7.5
Husband's education		
Less than high school	4.8	8.5
High school graduate	36.4	42.8
College	29.9	27.3
Graduate school	29.0	21.4
Parental history of MI before age 60	18.5	17.1
High cholesterol	35.6	32.1
High blood pressure	27.3	33.3
Diabetes	3.5	5.3
Angina	2.1	2.6
Stroke	0.3	0.7
Coronary revascularization	0.7	0.7
Osteoporosis	6.8	6.2
Multivitamin use	44.7	41.8
Cigarette smoking		
Never	41.8	42.0
Past	42.0	36.1
Current	16.2	22.0
Alcohol intake (g/d)		
0	31.7	38.1
0.1–4.9	33.7	31.0
5.0–9.9	12.0	9.9
10.0–14.9	9.9	9.3
≥15	12.7	11.7
Diet score in the 2 upper quintiles	41.6	35.3
Fruit and vegetable intake (servings/d)		
<3	57.9	57.8
3–4.9	29.0	29.0
5–9.9	12.5	12.3
≥10	0.5	0.8
Body mass index (kg/m ²)		
<25	56.2	49.3
25–29.9	29.9	32.1
≥30	13.8	18.7
Aspirin use		
Nonuse	21.9	26.7
1–4 y	18.1	18.9
5–10 y	19.5	15.9
>10 y	40.6	38.5
Physical activity (h/wk)		
<1	43.2	47.1
1–1.9	15.4	15.3
2–3.9	19.8	18.9
4–5.9	10.6	9.4
6–9.9	7.4	6.6
≥10	3.6	3.0

Results are expressed as percentages unless otherwise indicated.

TABLE 3. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events in the NHS “Trials”

	All	Follow-Up Period	
		0–24 Mo	>24 Mo
Initiators			
Total no.	7258	7258	7221
No. CHD events	98	22	76
Noninitiators			
Total no.	141,002	141,002	139,599
No. CHD events	3606	512	3094
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.96 (0.78–1.18)	1.42 (0.92–2.20)	0.88 (0.69–1.12)
By time after menopause (y)			
<10	0.84 (0.61–1.14)	1.33 (0.66–2.64)	0.77 (0.54–1.09)
≥10	1.12 (0.84–1.48)	1.48 (0.83–2.64)	1.05 (0.77–1.43)
P for interaction	0.08	0.90	0.07
By age (y)			
<60	0.86 (0.65–1.14)	1.36 (0.73–2.52)	0.78 (0.57–1.07)
≥60	1.15 (0.85–1.57)	1.49 (0.79–2.80)	1.08 (0.76–1.54)
P for interaction	0.05	0.72	0.06

Adjusted for the following baseline variables: age, parental history of myocardial infarction before age 60, education, husband's education, ethnicity, age at menopause, calendar month, high cholesterol, high blood pressure, diabetes, angina, stroke, coronary revascularization, osteoporosis, body mass index, cigarette smoking, aspirin use, alcohol intake, physical activity, diet score, multivitamin use, fruits and vegetables intake, and previous use of hormone therapy. The last column is restricted to women who were still at risk after the first 2 years of follow-up of the corresponding trial.

current versus never users. This approach is used in many large observational cohorts, including the NHS (see “Discussion” for details). Table 6 shows the cumulative steps that link our estimates in Table 3 with the previously reported NHS estimate. These steps involve changes in the start of follow-up, the definition of the exposed and unexposed group, the covariates used for adjustment, and eligibility criteria.

Column i of Table 6 shows the estimates when (as in previous NHS analyses) the start of follow-up, and thus the “baseline,” of each trial was redefined as the date of return of the questionnaire. When “baseline” is modified in this way, the selected group of initiators differs from the initiator group in Table 3 because it does not include women who, during the 2-year interval before “baseline,” either initiated and stopped hormone therapy or survived a CHD event occurring after initiation. As in Table 3, we provide separate HR estimates for the entire follow-up (0.84), the first 2 years of follow-up (0.98), and the period after the first 2 years (0.80).

Second, we varied the definition of the user and non-user groups in 3 steps as shown in the next 3 columns of Table 6. In column ii we eliminated our “trial”-specific criterion of no therapy in the 2 years before “baseline” for initiators; that is, we compared current users with noninitia-

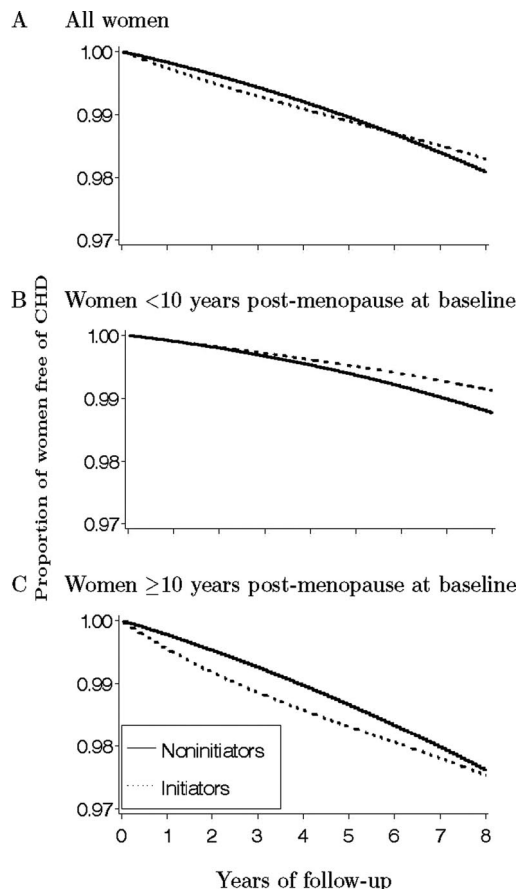


FIGURE 2. Proportion of women free of CHD by baseline treatment group in the Nurses' Health Study "trials."

tors. In column iii we eliminated our "trial"-specific criterion of no therapy in the 2 years before "baseline" for all women; that is, we compared current users with current nonusers. In column iv we used as the comparison group the subset of nonusers with no history of hormone therapy use; that is, we compared current users with never users as in previous NHS analyses. The HR estimates for columns ii, iii, iv were, respectively, 0.84, 0.86, 0.85 for the entire follow-up, 0.77, 0.77, 0.74 for 0 to 24 months, and 0.87, 0.90, 0.90 for >24 months.

To explain why the number of exposed cases ($n = 319$) in columns ii to iv far exceeds the number ($n = 66$) in column i, consider a woman who is continuously on hormone therapy from 1982–1984 until she dies of CHD just before the end of follow-up in 2000. In the analysis of column i, this woman participates as an exposed CHD case in the first (1984) "trial" only. In contrast, in the analyses of columns ii to iv, the same woman participates as an exposed CHD case in each of the 8 "trials" 1984–1998. Furthermore, in the analysis of column i, the woman would contribute 0 to the 0- to 24-month exposed case stratum and 1 to the >24-month exposed case stratum. In contrast, the same woman in the analyses of columns ii to

TABLE 4. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events Among Women With No History of Hormone Use in the NHS "Trials"

		Follow-Up Period	
	All	0–24 Mo	>24 Mo
Initiators			
Total no.	4939	4939	4914
No. CHD events	55	15	40
Noninitiators			
Total	109,205	109,205	108,108
No. CHD events	2723	379	2344
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.79 (0.60–1.03)	1.49 (0.88–2.54)	0.68 (0.49–0.93)
By time after menopause (y)			
<10	0.66 (0.44–0.98)	1.32 (0.58–3.03)	0.58 (0.37–0.90)
≥10	1.02 (0.70–1.50)	1.71 (0.85–3.45)	0.88 (0.56–1.38)
<i>P</i> for interaction	0.09	0.51	0.15
By age (y)			
<60	0.68 (0.48–0.97)	1.38 (0.66–2.88)	0.59 (0.39–0.88)
≥60	1.06 (0.69–1.64)	1.64 (0.73–3.69)	0.93 (0.56–1.56)
<i>P</i> for interaction	0.10	0.69	0.13
Adjusted for same baseline variables as in Table 3.			

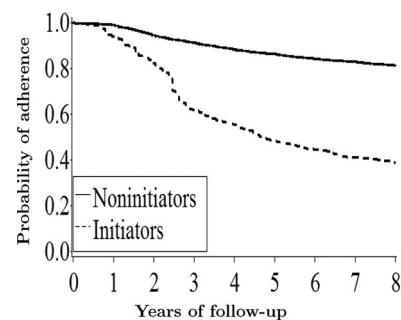


FIGURE 3. Proportion of women who adhered to their baseline treatment in the Nurses' Health Study "trials."

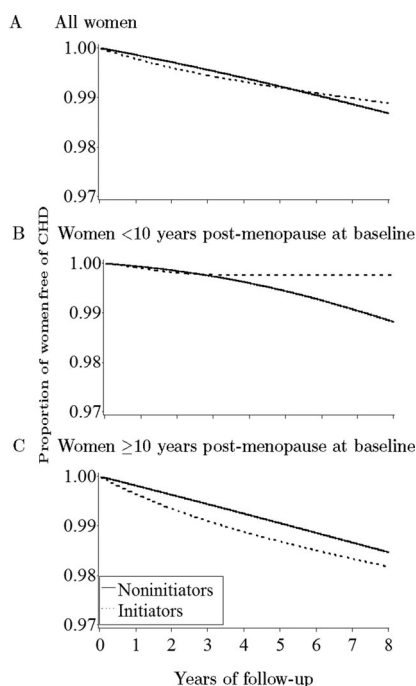
iv would contribute 1 to the 0- to 24-month exposed case stratum (corresponding to the 1998 "trial") and 7 to the >24-month exposed case stratum (corresponding to each of the other 7 "trials").

Third, we repeated the analysis in column iv after adjusting for the set of covariate values used in the most recent NHS publication. Thus, the estimates in column v—0.81 for the entire follow-up, 0.71 for 0 to 24 months, and 0.85 for >24 months—were adjusted for the most recent values available at the time of return of the "baseline" questionnaire, rather than the most recent values available at the 2 previous questionnaires.

TABLE 5. Estimates of the (Adherence-Adjusted) Effect of Continuous Estrogen/Progestin Therapy Versus No Hormone Therapy on the Incidence of CHD Events in the NHS "Trials"

	All HR (95% CI)	Follow-Up Period	
		0–24 Mo HR (95% CI)	>24 Mo HR (95% CI)
All women	0.98 (0.66–1.45)	1.61 (0.97–2.66)	0.64 (0.35–1.15)
By time after menopause (y)			
<10	0.54 (0.19–1.51)	1.21 (0.40–3.61)	0.14 (0.02–1.16)
≥10	1.20 (0.78–1.84)	1.92 (1.09–3.39)	0.84 (0.45–1.56)
<i>P</i> for interaction	0.01	0.18	0.11
By age (y)			
<60	0.78 (0.44–1.40)	1.65 (0.81–3.37)	0.45 (0.19–1.09)
≥60	1.36 (0.81–2.29)	1.69 (0.86–3.32)	1.08 (0.50–2.36)
<i>P</i> for interaction	0.06	0.74	0.09

Adjusted for same baseline variables as in Table 3. In each "trial," women were censored when they discontinued their baseline treatment (either hormone therapy or no hormone therapy), and the uncensored women months were weighted by the inverse of their estimated probability of remaining uncensored until that month.

**FIGURE 4.** Proportion of women free of CHD under full adherence with the baseline treatment in the Nurses' Health Study "trials."

Fourth, we repeated the analysis in column v after dropping the requirement of an intact uterus, which was not used in previous NHS analyses. The estimates in column vi were 0.82 for the entire follow-up, 0.67 for 0 to 24 months, and 0.87 for >24 months. The estimate 0.67 in the row 0 to 24 months corresponds almost exactly to the analytic approach used in the most recent NHS publication,¹⁷ which

estimated the HR over the 2-year period after the reclassification (ie, updating) of treatment status at the return of each questionnaire.

DISCUSSION

We used the NHS observational data to emulate the design and analysis of the WHI randomized trial. The ITT HRs of CHD for therapy initiation were 1.42 (95% CI = 0.92–2.20) in the NHS vs. 1.68 (95% CI = 1.15–2.45) in the WHI⁹ during the first 2 years, and 1.00 (0.78–1.28) in the NHS versus approximately 1.24 (0.97–1.60) in the WHI⁸ during the first 8 years. However, much of the apparent WHI-NHS difference disappeared after stratification by time since menopause at hormone therapy initiation. The ITT HRs were 0.84 (0.61–1.14) in the NHS versus 0.88 (0.54–1.43) in the WHI^{8,18} for women within 10 years after menopause, and approximately 1.12 (0.84–1.48) in the NHS versus 1.23 (0.85–1.77) in the WHI^{8,18} for women between 10 and 20 years after menopause.

These findings provide additional support to the hypothesis that hormone therapy may increase the long-term CHD risk only in women who were 10 or more years after menopause at initiation,^{17,19} and to the rationale for an ongoing randomized clinical trial to determine the effect of estrogen plus progestin on coronary calcification in younger women.²⁰ When the analyses were limited to women with no history of hormone use, the ITT HR was 0.79 (0.60–1.03) for the entire follow-up and 0.66 (0.44–0.98) for women who initiated hormone use within 10 years of menopause.

We computed average ITT HRs in the NHS for comparison with the main result of the WHI. Our ITT estimates suggest that any remaining differences between NHS and WHI estimates are not explained by unmeasured joint risk factors for CHD and therapy discontinuation. However, the average ITT HR is not the ideal effect measure because the survival curves crossed during the follow-up in both the WHI trial and the NHS trials, and also because ITT estimates like the ones shown here are generally attenuated toward the null due to misclassification of actual treatment. We addressed the first problem by estimating survival curves to first CHD event, and the second problem by estimating these curves under full adherence (via inverse probability weighting). Therefore the adherence-adjusted survival curves of Figure 4 provide the most appropriate summary of our results. It will be of interest to compare these results with adherence-adjusted curves (via inverse probability weighting) from the WHI when they become available. The curves suggest that continuous hormone therapy causes a net reduction in CHD among women starting therapy within 10 years of menopause, and a net increase among those starting later. However, either of these effects could be due to sampling variability.

Previously published NHS estimates¹⁷ compared the hazards of current versus never users over the 2-year period

TABLE 6. Comparison of Several Alternative Hazard Ratio Estimates With the Previously Reported Estimate From the NHS (Column vi, Row 0–24 Mo)

					Current Users vs. Never Users		
	Initiators vs. Noninitiators ^a	Selected ^b Initiators vs. Noninitiators (i)	Current Users vs. Noninitiators (ii)	Current Users vs. Nonusers (iii)	(iv)	Covariates of Previous NHS Analyses (v)	Not Requiring Presence of Uterus (vi)
Entire Follow-up							
Users							
Total no.	7258	6400	41,441	41,441	41,441	41,441	45,793
No. CHD events	98	66	319	319	319	319	398
Nonusers							
Total no.	141,002	141,316	141,316	173,094	126,235	126,235	147,045
No. CHD events	3606	3271	3271	3764	2778	2778	3404
All women	0.96 (0.78–1.18)	0.84 (0.64–1.09)	0.84 (0.67–1.06)	0.86 (0.70–1.06)	0.85 (0.68–1.07)	0.81 (0.65–1.01)	0.82 (0.68–0.99)
Time from menopause							
<10 y	0.84 (0.61–1.14)	0.66 (0.45–0.98)	0.76 (0.57–1.02)	0.79 (0.60–1.03)	0.76 (0.57–1.01)	0.74 (0.56–0.98)	0.77 (0.59–1.00)
≥10 y	1.12 (0.84–1.48)	1.05 (0.75–1.47)	0.95 (0.72–1.27)	0.95 (0.72–1.25)	0.92 (0.68–1.25)	0.90 (0.68–1.20)	0.89 (0.70–1.13)
P interaction	0.08	0.03	0.05	0.09	0.08	0.08	0.12
Age							
<60 y	0.86 (0.61–1.14)	0.67 (0.47–0.97)	0.80 (0.61–1.05)	0.82 (0.64–1.06)	0.79 (0.60–1.03)	0.79 (0.60–1.05)	0.79 (0.62–1.01)
≥60 y	1.15 (0.85–1.57)	1.14 (0.80–1.63)	0.92 (0.67–1.26)	0.94 (0.69–1.27)	0.93 (0.67–1.29)	0.83 (0.63–1.10)	0.86 (0.68–1.10)
P interaction	0.05	0.01	0.14	0.20	0.17	0.49	0.24
0–24 mo							
CHD events							
No. users	22	17	80	80	80	80	90
No. nonusers	512	660	660	755	542	542	666
HR (95% CI)	1.42 (0.92–2.20)	0.98 (0.60–1.60)	0.77 (0.60–0.99)	0.77 (0.60–0.98)	0.74 (0.57–0.95)	0.71 (0.55–0.91)	0.67 (0.54–0.85)
>24 mo							
CHD events							
No. users	76	49	239	239	239	239	308
No. nonusers	3094	2611	2611	3008	2236	2236	2738
HR (95% CI)	0.88 (0.69–1.12)	0.80 (0.60–1.08)	0.87 (0.68–1.12)	0.90 (0.72–1.14)	0.90 (0.70–1.15)	0.85 (0.67–1.08)	0.87 (0.71–1.07)

^aFrom Table 3. Follow-up starts at time of therapy initiation. In all other columns follow starts at time of questionnaire return.

^bWomen who initiated and stopped therapy, or who survived a CHD event, between the time of therapy initiation and the time of questionnaire return are excluded. See main text for a description of each estimate.

after the updating of treatment status at the return of each questionnaire, and could thus be viewed as a form of adherence adjustment. In Table 6 we described the steps from our 2-year ITT estimate to the previously published adherence-adjusted estimate. Below we discuss the 2 key steps: the change of start of follow-up (time of therapy initiation vs. time of questionnaire return), and the change of the exposed group (selected initiators vs. current users).

The 2-year HR estimate changed from 1.42 (Table 3) to 0.98 (Table 6, column i) during the first 2 years, and from 0.96 (Table 3) to 0.84 (Table 6, column i) for the entire follow-up when the definition of start of follow-up was changed from the estimated time of therapy initiation to the time of return of the next questionnaire (the latter definition is commonly used in observational studies that collect treatment information at regular intervals). This latter definition excludes women who initiated treatment and then suffered a nonfatal myocardial infarction during the interval between

treatment initiation and treatment ascertainment (up to 2 years in the NHS). If hormone therapy increases the short-term risk of CHD, this exclusion will result in an underestimation of the early increase in risk and may result in selection bias,¹⁶ which may explain part of the change from 1.42 to 0.98. The impact of this exclusion bias, however, will be diluted over the entire follow-up, as previously suggested in a sensitivity analysis,¹⁷ which may explain the smaller change from 0.96 to 0.84. This exclusion bias may be quantified through simulations,²¹ reduced by stratification of the analysis on duration of therapy at baseline,²¹ and eliminated by making the start of follow-up coincident with the time of treatment initiation, as discussed by Robins^{22,23} and Ray.²⁴ The approach we present here and elsewhere^{10,25} generalizes Ray's "new-users design" to the case of time-varying treatments.

The point estimate further changed from 0.98 (Table 6, column i) to 0.77 (column ii) when the definition of exposure

changed from selected initiators to current users. These are estimates for different contrasts. The estimate in column i is based on the exposed person-time during the 2-year period immediately after the return of the questionnaire in which therapy initiation was reported, and thus can be viewed as a flawed attempt to estimate the early effect of therapy initiation (see previous paragraph). The estimate in column ii, however, is based on the exposed person-time pooled over all 2-year periods after the return of any questionnaire, and thus can be interpreted as an attempt to estimate the effect of therapy use during any 2-year period (that excludes the interval between therapy initiation and return of the next questionnaire, as discussed in the previous paragraph). More specifically, the approach in column ii can be understood as an attempt to estimate adherence-adjusted effects by entering the current value of exposure and the joint predictors of adherence and CHD as time-varying covariates in the model for CHD risk. Unlike inverse probability weighting, this approach to adherence adjustment requires that the time-dependent covariates not be strongly affected by prior treatment. This may be a reasonable assumption in the NHS. Thus the estimates in column ii may be more usefully compared with a weighted average of our interval-specific adherence adjusted estimates of 1.61 (0–2 years), 0.65 (2–5 years), 0.47 (5–8 years), and 0.85 (>8 years) than to the estimate in column i.

In summary, our findings suggest that the discrepancies between the WHI and NHS ITT estimates could be largely explained by differences in the distribution of time since menopause and length of follow-up. Residual confounding for the effect of therapy initiation in the NHS seems to play little role.

ACKNOWLEDGMENTS

We thank Murray Mittleman, Javier Nieto, Meir Stampfer, and Alexander Walker for their comments on an earlier version of the manuscript.

REFERENCES

- Ioannidis JP, Haidich AB, Pappa M, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286:821–830.
- Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878–1886.
- Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887–1892.
- Grodstein F, Stampfer M, Manson J, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease [Erratum in: *N Engl J Med*. 1996;335:1406]. *N Engl J Med*. 1996;335:453–461.
- Grodstein F, Manson JE, Colditz GA, et al. A prospective, observational study of postmenopausal hormone therapy and primary prevention of cardiovascular disease. *Ann Intern Med*. 2000;133:933–941.
- Varas-Lorenzo C, García-Rodríguez LA, Pérez-Gutthann S, et al. Hormone replacement therapy and incidence of acute myocardial infarction. *Circulation*. 2000;101:2572–2578.
- The Women's Health Initiative Study Group. Design of the women's health initiative clinical trial and observational study. *Control Clin Trials*. 1998;19:61–109.
- 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.
- Prentice RL, Pettinger M, Anderson GL. Statistical issues arising in the Women's Health Initiative. *Biometrics*. 2005;61:899–911.
- Hernán MA, Robins JM, García Rodríguez LA. In discussion of: Prentice RL, Pettinger M, Anderson GL. Statistical issues arising in the Women's Health Initiative. *Biometrics*. 2005;61:922–930.
- Stampfer MJ, Hu FB, Manson JE, et al. Primary prevention of coronary heart disease in women through diet and lifestyle. *N Engl J Med*. 2000;343:16–22.
- Robins JM. The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. In: Sechrest L, Freeman H, Mulley A, eds. *Health Services Research Methodology: A Focus on AIDS: NCHRS, US*. Washington, DC: Public Health Service;1989:113–159.
- Mendelsohn ME, Karas RH. Hormone replacement therapy and the young at heart. *N Engl J Med*. 2007;356:2639–2641.
- Mark SD, Robins JM. A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Control Clin Trials*. 1993;14:79–97.
- Cole SR, Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials*. 2005;26:300–310.
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
- 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;15:35–44.
- Manson JE, Bassuk SS. Invited commentary: hormone therapy and risk of coronary heart disease why renew the focus on the early years of menopause? *Am J Epidemiol*. 2007;166:511–517.
- Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med*. 2003;348:645–650.
- Harman SM, Brinton EA, Cedars M, et al. KEEPS: The Kronos Early Estrogen Prevention Study. *Climacteric*. 2005;8:3–12.
- Prentice RL, Langer RD, 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.
- Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period — Application to the healthy worker survivor effect [published errata appear in *Math Model*. 1987;14:917–921]. *Math Model*. 1986;7:1393–1512.
- Robins JM. Addendum to “A new approach to causal inference in mortality studies with a sustained exposure period—application to the healthy worker survivor effect” [published errata appear in *Comput Math Appl*. 1989;18:477]. *Comput Math Appl*. 1987;14:923–945.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158:915–920.
- Alonso A, García Rodríguez LA, Logroscino G, et al. Gout and risk of Parkinson's disease: a prospective study. *Neurology*. 2007;69:1696–1700.
- Connelly M, Richardson M, Platt R. Prevalence and duration of postmenopausal hormone replacement therapy use in a managed care organization. *J Gen Intern Med*. 2000;15:542–50.
- Robins JM, Rotnitzky A, Vansteelandt S. In discussion of: Frangakis CE, Rubin DB, An M, MacKenzie E. “Principal stratification designs to estimate input data missing due to death”. *Biometrics*. 2007;63:650–662.
- Robins JM. Structural nested failure time models. In: Armitage P, Colton T, eds. *Survival Analysis. The Encyclopedia of Biostatistics*. Chichester, UK: John Wiley and Sons; 1998:4372–4389.
- Hernán MA, Cole SR, Margolick J, et al. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf*. 2005;14:477–491.
- Robins JM, Blevins D, Ritter G, et al. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients [published errata appear in *Epidemiology*. 1993;4:189]. *Epidemiology*. 1992;3:319–336.

APPENDIX: SENSITIVITY TO OUR ANALYTIC CHOICES FOR THE NHS NONRANDOMIZED TRIALS

We now describe the estimates from sensitivity analyses that alter some of the decisions we made for the analyses shown in Table 3. The results from these sensitivity analyses indicate that these decisions had only a moderate influence on our estimates.

Appendix A1: The Determination of Month of Therapy Initiation

The duration of use of hormone therapy during a given 2-year period is ascertained as a categorical variable with 5 levels in the NHS questionnaires. Therefore any decisions regarding the exact month of therapy initiation will result in some error. We explored the sensitivity of our estimates to this error by conducting separate analyses in which we varied the decisions used to obtain the estimates in Table 3. In the analyses shown in Appendix Table 1, we used the latest possible month of initiation as the month of therapy initiation. For example, if a woman on hormone therapy reported 15–19 months of use during the 2-year period before the return of the baseline questionnaire, we calculated the month of initiation as the month of questionnaire return minus 19 in Table 3, and minus 15 in Appendix Table 1.

APPENDIX TABLE 1. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events When Using the Latest Possible Month of Therapy Initiation in the NHS “Trials”

		Follow-up Period	
	All	0–24 Mo	>24 Mo
Initiators			
Total no.	7245	7245	7165
No. CHD events	90	24	66
Noninitiators			
Total no.	140,881	140,881	139,331
No. CHD events	3533	545	2988
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.92 (0.74–1.15)	1.49 (0.97–2.27)	0.81 (0.63–1.05)
By time after menopause (y)			
<10	0.81 (0.59–1.12)	0.99 (0.44–2.20)	0.79 (0.55–1.12)
≥10	1.06 (0.79–1.43)	1.84 (1.11–3.05)	0.88 (0.62–1.26)
<i>P</i> for interaction	0.11	0.20	0.35
By age (y)			
<60	0.81 (0.60–1.09)	1.04 (0.51–2.10)	0.76 (0.55–1.07)
≥60	1.13 (0.82–1.56)	1.98 (1.16–3.40)	0.93 (0.63–1.38)
<i>P</i> for interaction	0.04	0.11	0.22

Adjusted for same baseline variables as in Table 3.

Appendix A2: The Inclusion of Women Over Age 50

The WHI trial excluded women younger than 50 years at baseline. Appendix Tables 2 and 3 show, respectively, the ITT and adherence-adjusted estimates when we added this exclusion criterion to the eligibility criteria of our NHS “trials.” The ITT HRs (95% CIs) of CHD for initiators versus non-initiators were 0.99 (0.80–1.22) for the entire follow-up, 1.80 (1.01–3.19) for the first year, 1.43 (0.92–2.23) for the first 2 years, 1.13 (0.85–1.50) for the first 5 years, and 1.05 (0.82–1.34) for the first 8 years. The adherence-adjusted HRs (95% CIs) were 1.30 (0.76–2.21) for the entire follow-up, 1.61 (0.84–3.08) for the first year, 1.71 (1.03–2.83) for the first 2 years, 1.22 (0.80–1.88) for the first 5 years, and 1.35 (0.78–2.35) for the first 8 years. The HR (95% CI) was 0.69 (0.32–1.48) during years 2–5, 1.73 (0.41–2.11) during years 5–8, and 0.91 (0.17–4.83) after year 8.

Appendix A3: The Exclusion of Women Who Died Between the Start of Follow-Up and the Return of the Next Questionnaire

There are 2 reasons why the initiators in our analysis were actually a selected group of all initiators. First, it is possible that some short-term users of hormone therapy were not detected in the NHS. Of note, the adherence of NHS women during the first year after initiation was higher than that previously found in other US²⁶ and UK¹⁰ women, which might reflect a truly greater adherence of NHS women or the

APPENDIX TABLE 2. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events Among Women Aged 50 or More at Baseline in the NHS “Trials”

		Follow-up Period	
	All	0–24 Mo	>24 Mo
Initiators			
Total no.	6602	6602	6566
No. CHD events	94	21	73
Noninitiators			
Total no.	135,877	135,887	134,491
No. CHD events	3503	503	3000
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.99 (0.80–1.22)	1.43 (0.92–2.23)	0.91 (0.72–1.16)
By time after menopause (y)			
<10	0.88 (0.63–1.21)	1.28 (0.62–2.64)	0.81 (0.56–1.17)
≥10	1.13 (0.85–1.49)	1.50 (0.84–2.68)	1.06 (0.77–1.44)
P for interaction	0.12	0.85	0.11
By age (y)			
<60	0.89 (0.67–1.19)	1.36 (0.71–2.57)	0.82 (0.59–1.14)
≥60	1.15 (0.85–1.57)	1.49 (0.79–2.80)	1.08 (0.77–1.53)
P for interaction	0.08	0.73	0.09

Adjusted for same baseline variables as in Table 3.

APPENDIX TABLE 3. Estimates of the (Adherence-Adjusted) Effect of Continuous Estrogen/Progestin Therapy Versus No Hormone Therapy on the Incidence of CHD Events Among Women Aged 50 or More at Baseline in the NHS "Trials"

	All HR (95% CI)	0–24 Mo HR (95% CI)	>24 Mo HR (95% CI)
All women	1.30 (0.76–2.21)	1.71 (1.03–2.83)	1.07 (0.44–2.63)
By time after menopause (y)			
<10 y	0.68 (0.24–1.91)	1.28 (0.43–3.86)	0.20 (0.03–1.54)
≥10 y	1.57 (0.86–2.85)	1.97 (1.11–3.47)	1.37 (0.54–3.45)
<i>P</i> for interaction	0.03	0.37	0.06
By age (y)			
<60	0.91 (0.49–1.69)	1.80 (0.83–3.87)	0.54 (0.20–1.49)
≥60	1.92 (0.90–4.10)	1.69 (0.87–3.32)	2.10 (0.68–6.50)
<i>P</i> for interaction	0.06	0.94	0.09

Adjusted for same baseline variables as in Table 3. In each "trial," women were censored when they discontinued their baseline treatment (either hormone therapy or no hormone therapy), and the uncensored women-months were weighted by the inverse of their estimated probability of remaining uncensored until that month.

questionnaires' inability to identify all short-term users. Second, both the initiators (and noninitiators) in our analysis did not include women who died before returning the questionnaire. The month of therapy initiation, if any, for women who died between the start of follow-up and the return of the next questionnaire is unknown. As a result, these women were not included in our analyses in Table 3, which might have resulted in selection bias if the women who had a CHD event and died before returning the questionnaire were more (or less) likely to have initiated therapy than those who did not die. As an aside, because the analyses presented in columns i–vi of Table 6 used the date of return of the questionnaire as the start of follow-up, the number of women excluded for this reason is lower in Table 6 than in Table 3. This explains why the number of CHD cases during the first 2 years of follow-up is 534 in Table 3 and 677 in column i of Table 6.

We used inverse probability weighting¹⁶ to adjust for the potential selection bias due to death before questionnaire return. Specifically, we estimated the conditional probability of surviving until the return of the questionnaire for every woman who, having had a CHD event during the 2-year interval prior to the baseline questionnaire, survived to return the questionnaire. We then upweighted these survivors by the inverse of their estimated conditional probability of survival. This approach implicitly assumes that there exists a hypothetical intervention to prevent death before returning the questionnaire among women who had a CHD event.

To estimate the probability of survival, we fit a logistic model among women who had a CHD event in the 2-year interval before the return of the questionnaire. The outcome of the model was the probability of survival until questionnaire return, and the covariates were those used in our Table 3 analyses to adjust for confounding. This approach adjusts only for the selection bias that can be explained by these

covariates. Appendix Table 4 shows the inverse probability weighted ITT HRs and their 95% CIs, which are similar to those in Table 3—although the HR for initiators versus noninitiators during the first 2 years of follow-up was closer to the null in Appendix Table 4 (1.30) than in Table 3 (1.48).

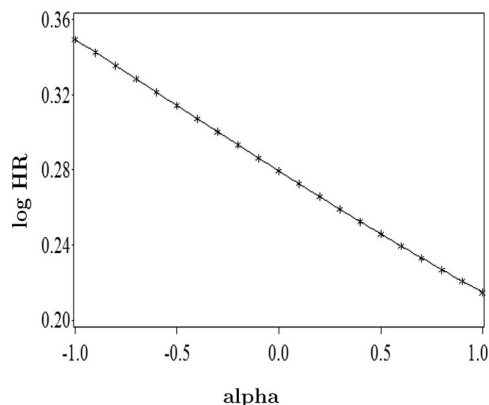
However, our inverse probability weighted analysis could not adjust for treatment status because it is unknown whether women who died before returning their questionnaire were initiators. Thus, if the probability of dying after or from a CHD event was affected by treatment, our inverse probability weighted analysis would not appropriately adjust for the selection bias. We conducted a sensitivity analysis to determine whether lack of adjustment for treatment status could explain the increased CHD incidence observed in initiators during the first 2 years of follow-up. The methodology for this sensitivity analysis has been recently described.²⁷ Appendix Figure 1 summarizes the results.

The ITT HR of CHD varies from 1.42 for $\alpha = -1$ to 1.24 for $\alpha = 1$, where α is the log odds ratio for the hypothesized association between treatment arm and death before returning the questionnaire, conditional on the other covariates. Our analysis in Appendix Table 4 corresponds to $\alpha = 0$. These results suggest that the potential selection bias due to lack of adjustment for treatment arm in the inverse probability-weighted analysis does not fully explain the increased CHD incidence rate during the first 2 years of follow-up in initiators versus noninitiators.

APPENDIX TABLE 4. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events in the NHS "Trials," After Adjustment for Exclusion of Women Who Died Between the Start of Follow-Up and the Return of the Next Questionnaire

	All	Follow-up Period	
		0–24 Mo	>24 Mo
Initiators			
Total no.	7258	7258	7221
No. CHD events	98	22	76
Noninitiators			
Total no.	141,002	141,002	139,599
No. CHD events	3606	512	3094
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.96 (0.77–1.18)	1.30 (0.83–2.05)	0.88 (0.69–1.12)
By time after menopause (y)			
<10	0.85 (0.62–1.16)	1.37 (0.69–2.73)	0.77 (0.54–1.09)
≥10	1.09 (0.82–1.46)	1.27 (0.69–2.34)	1.05 (0.77–1.43)
<i>P</i> for interaction	0.12	0.82	0.07
By age (y)			
< 60	0.86 (0.65–1.14)	1.33 (0.71–2.47)	0.78 (0.57–1.07)
≥60	1.14 (0.83–1.55)	1.31 (0.68–2.53)	1.08 (0.76–1.54)
<i>P</i> for interaction	0.07	0.93	0.06

Adjustment does not affect to estimates for >24 mo. HR adjusted for same baseline variables as in Table 3.



APPENDIX FIGURE 1. Sensitivity analysis for lack of adjustment for treatment arm in the inverse probability weighted analysis that adjusts for selection bias due to death between the start of follow-up and the return of questionnaire in the Nurses' Health Study "trials." The parameter alpha is the log odds ratio for the hypothesized association between treatment arm and death before returning the questionnaire. Log HR is the log HR of CHD for initiators versus noninitiators during the first 2 years of follow-up.

Appendix A4: The Use of Propensity Scores

To assess whether our results were affected by the choice of the effect measure (ie, HR) or by the method of adjustment for confounding, we also conducted the analyses by g-estimation of a nested, trial-specific, time-independent accelerated failure time model,^{10,28} which estimates the median survival time ratio of noninitiators versus initiators and adjusts for confounding by combining a model for the propensity score with a model for the effect of the covariates on time to CHD.²⁹ G-estimation of nested structural models is a particularly robust way of utilizing propensity scores as it is minimally affected by poor overlap in the propensity scores of the treated and untreated.^{29,30} The estimates, shown in Appendix Table 5, are qualitatively similar to those in Table 3, which suggests that our conclusions are not sensitive to the method used for confounding adjustment.

Appendix A5: The Assumption of No Unmeasured Confounding

To examine the amount of confounding by measured lifestyle and socioeconomic compared with other risk factors, we first repeated the analysis in Table 3 without adjusting for measured lifestyle factors (alcohol intake, physical activity, aspirin use, diet score, multivitamin use, fruit and vegetable intake). The HR was 0.94 (95% CI = 0.76–1.16). When we also omitted adjustment for our measures of socioeconomic status (education, ethnicity, husband's education), the HR was 0.92 (0.75–1.14). We repeated the analyses without adjusting for any of the potential confounders except age; the age-adjusted HR was 0.67 (0.54–0.83) for CHD. Finer stratification by age (in 2-year intervals) and adjustment for age as a continuous covariate did not materially affect the results.

APPENDIX TABLE 5. Estimates of the Intention-to-Treat Effect of Initiation of Estrogen/Progestin Therapy on the Incidence of CHD Events in the NHS "Trials" (Effect Measured as Median Survival Time Ratio)

Entire Follow-up	
Initiators	
Total no.	7258
No. CHD events	98
Noninitiators	
Total no.	141,002
No. CHD events	3,606
STR (95% CI)	0.87 (0.62–1.08)
By time after menopause (y)	
<10	0.71 (0.43–1.03)
≥10	1.04 (0.70–1.36)
By age (y)	
<60 y	0.66 (0.47–1.03)
≥60 y	1.11 (0.72–1.50)
0–24 mo	
CHD events no.	
Initiators	22
Noninitiators	512
STR (95% CI)	1.82 (0.50–3.70)
Survival time ratios (STRs) adjusted for same baseline variables as in Table 3.	

It is suspected that important confounders of the effect of hormone therapy on CHD risk also confound its effect on stroke risk. Thus we estimated the ITT effect of hormone therapy on stroke under the hypothesis that, in the presence of substantial unmeasured confounding for the effect on CHD risk, the effect estimates for stroke would also be biased. There were 574 cases of stroke among eligible women. Applying the same analytic strategy as in Table 3, the overall HR for stroke was 1.39 (CI = 1.09–1.77), which is similar to the estimate found in the WHI randomized trial.

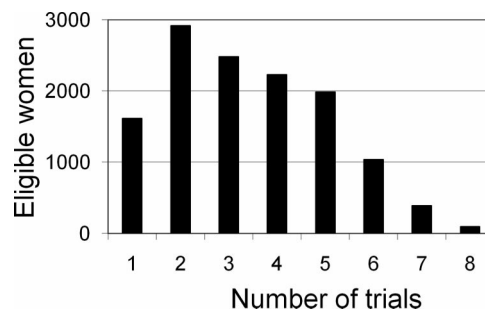
We also repeated the analysis in column vi of Table 6 without adjustment for measured lifestyle factors other than smoking (alcohol intake, physical activity, aspirin use, multivitamin use, vitamin E intake). The HR was 0.67 (CI = 0.53–0.85). When we also omitted adjustment for our measures of socioeconomic status (husband's education), the HR was 0.65 (0.52–0.82). We repeated the analyses without adjusting for any of the potential confounders except age; the age-adjusted HR was 0.48 (0.38–0.60).

To further evaluate whether our decision not to assume comparability on unmeasured factors between those continuing versus discontinuing therapy had an important effect on our adherence-adjusted estimates, we compared our estimated ITT effect of hormone initiation with an estimate of the ITT effect of discontinuation under the assumption of no unmeasured confounders for discontinuation. To calculate this latter effect we recreated a set of NHS "trials" with the same protocol and analytic approach described above except that we restricted participation in each "trial" to women who

reported use of hormone therapy in the questionnaire before baseline.

We implemented the ITT approach by considering the treatment variable to be either 1 or 0 depending on whether the woman reported herself to be off versus on hormone therapy at the baseline questionnaire (regardless of future hormone history), and fit the Cox models described above. Under the assumption of no unmeasured confounders for treatment discontinuation given the variables used in our analysis, the estimates of effect so obtained are comparable with those from a randomized trial among hormone users in which treatment discontinuation is assigned at random.

Our analyses included 12,739 women who met the eligibility criteria for at least 1 NHS estrogen/progestin discontinuation “trial.” Appendix Figure 2 shows the distribution of women by number of “trials” in which they participated. Of these, 131 had a CHD event, 49 died of other causes or were lost to follow-up, and 12,559 reached the administrative end of follow-up free of a diagnosis of CHD. Appendix Table 6 shows the number of participants, stoppers, and CHD events in each of the “trials,” which include fewer participants than those for hormone therapy initiation because they are restricted to the smaller group of hormone therapy users. The HR when we compared the 52 events in the 4617 stoppers with the 209 events in the 24,255 nonstoppers was 1.13 (CI = 0.82–1.56). The number of events was insufficient to conduct meaningful subgroup analyses.



APPENDIX FIGURE 2. Distribution of eligible women by number of NHS “trials” of hormone therapy discontinuation in which they participated.

APPENDIX TABLE 6. Number of Participants, Therapy Stoppers, and CHD Events in Each NHS “Trial” to Estimate the Intention-to-Treat Effect of Discontinuation of Estrogen/Progestin Therapy

Trial	Questionnaire Yr	Participants	Stoppers	CHD Events
1	1984	107	59	2
2	1986	438	172	14
3	1988	1311	327	26
4	1990	2917	619	42
5	1992	4303	670	47
6	1994	5736	867	54
7	1996	7446	917	48
8	1998	6614	986	28