

# Observational Studies Analyzed Like Randomized Experiments

## *An Application to Postmenopausal Hormone Therapy and Coronary Heart Disease*

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

**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.<sup>1–3</sup> 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 <sup>a</sup>Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts; <sup>b</sup>Harvard-MIT Division of Health Sciences and Technology, Boston, Massachusetts; <sup>c</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota; <sup>d</sup>Channing Laboratory, Department of Medicine and <sup>e</sup>Obstetrics and Gynecology Epidemiology Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; <sup>f</sup>Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts; <sup>g</sup>Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; and <sup>h</sup>Department 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)<sup>4,5</sup> in the United States and on the General Practice Research Database<sup>6</sup> in the United Kingdom. More recently, the Women's Health Initiative (WHI) randomized trial<sup>7</sup> 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).<sup>8,9</sup>

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).<sup>10</sup> 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.<sup>4</sup>

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),<sup>11</sup> 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 ( $\chi^2$ ; 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 ( $\chi^2$ ; 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.<sup>12</sup>

Several lines of evidence suggest a modification of the effect of hormone therapy by time of initiation.<sup>13</sup> 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<sup>14,15</sup> 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.<sup>16</sup> 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

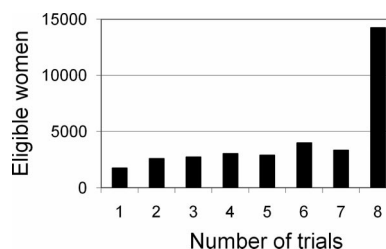
years before the baseline of all the “trials.” Of the remaining 45,884 women, we excluded 11,309 who did not have an intact uterus in 1984. Thus 34,575 women met our eligibility criteria for at least one NHS “trial.” Of these women, 1035 had a CHD event, 2596 died of other causes or were lost to follow-up, 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.<sup>17</sup> 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 <sup>2</sup> )		
<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”

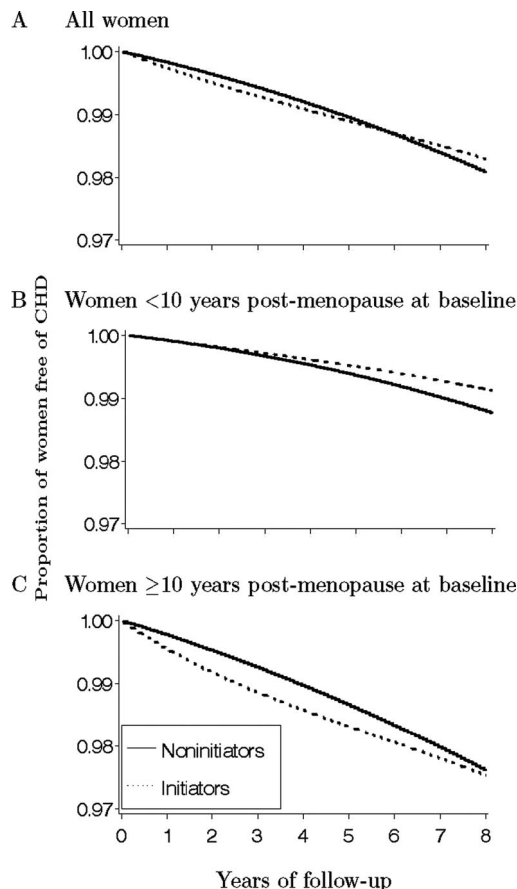
	All	Follow-Up Period	
		0–24 Mo	>24 Mo
Initiators			
Total no.	7258	7258	7221
No. CHD events	98	22	76
Noninitiators			
Total no.	141,002	141,002	139,599
No. CHD events	3606	512	3094
	HR (95% CI)	HR (95% CI)	HR (95% CI)
All women	0.96 (0.78–1.18)	1.42 (0.92–2.20)	0.88 (0.69–1.12)
By time after menopause (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."

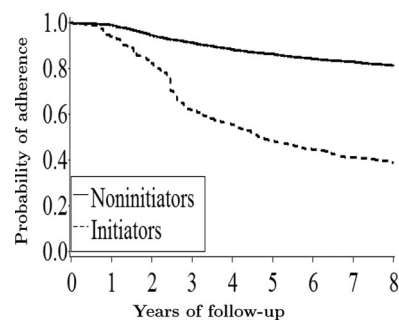
tors. In column iii we eliminated our "trial"-specific criterion of no therapy in the 2 years before "baseline" for all women; that is, we compared current users with current nonusers. In column iv we used as the comparison group the subset of nonusers with no history of hormone therapy use; that is, 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.			

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").

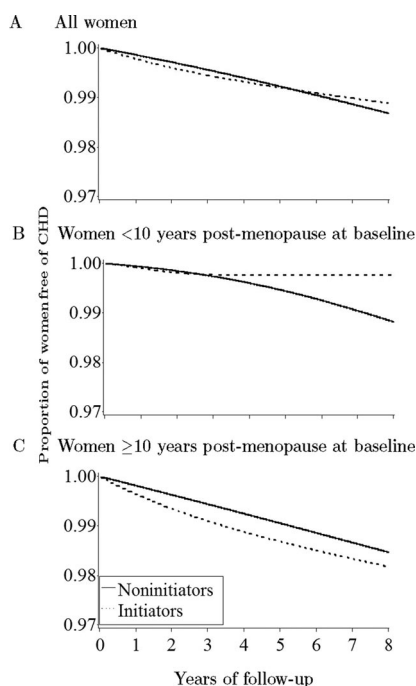
Third, we repeated the analysis in column iv after adjusting for the set of covariate values used in the most recent NHS publication. Thus, the estimates in column v—0.81 for the entire follow-up, 0.71 for 0 to 24 months, and 0.85 for >24 months—were adjusted for the most recent values available at the 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,<sup>17</sup> 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 WHI<sup>9</sup> 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 WHI<sup>8</sup> 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<sup>8,18</sup> 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<sup>8,18</sup> 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,<sup>17,19</sup> and to the rationale for an ongoing randomized clinical trial to determine the effect of estrogen plus progestin on coronary calcification in younger women.<sup>20</sup> 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 estimates<sup>17</sup> 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 <sup>a</sup>	Selected <sup>b</sup> 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)

<sup>a</sup>From Table 3. Follow-up starts at time of therapy initiation. In all other columns follow starts at time of questionnaire return.

<sup>b</sup>Women 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,<sup>16</sup> 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,<sup>17</sup> which may explain the smaller change from 0.96 to 0.84. This exclusion bias may be quantified through simulations,<sup>21</sup> reduced by stratification of the analysis on duration of therapy at baseline,<sup>21</sup> and eliminated by making the start of follow-up coincident with the time of treatment initiation, as discussed by Robins<sup>22,23</sup> and Ray.<sup>24</sup> The approach we present here and elsewhere<sup>10,25</sup> 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)
P 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)
P 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 US<sup>26</sup> and UK<sup>10</sup> 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 weighting<sup>16</sup> 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.<sup>27</sup> Appendix Figure 1 summarizes the results.

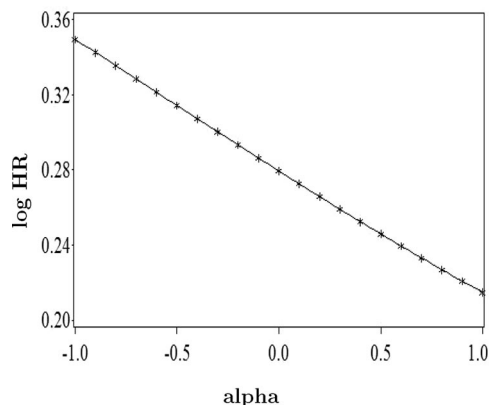
The ITT HR of CHD varies from 1.42 for  $\alpha = -1$  to 1.24 for  $\alpha = 1$ , where  $\alpha$  is the log odds ratio for the hypothesized association between treatment arm and death before returning the questionnaire, conditional on the other covariates. Our analysis in Appendix Table 4 corresponds 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,<sup>10,28</sup> 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.<sup>29</sup> 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.<sup>29,30</sup> 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.

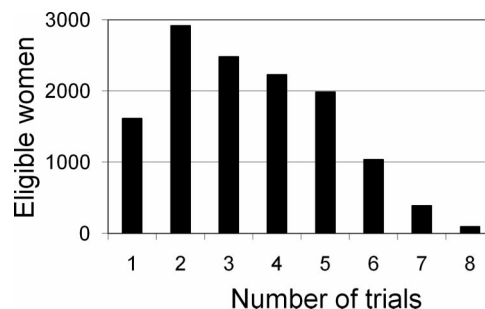
We also repeated the analysis in column vi of Table 6 without adjustment for measured lifestyle factors other than smoking (alcohol intake, physical activity, aspirin use, multivitamin use, vitamin E intake). The HR was 0.67 (CI = 0.53–0.85). When we also omitted adjustment for our measures of socioeconomic 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