



Published in final edited form as:

Epidemiology. 2010 July ; 21(4): 528–539. doi:10.1097/EDE.0b013e3181df1b69.

Estimating absolute risks in the presence of nonadherence: An application to a follow-up study with baseline randomization

Sengwee Toh¹, Sonia Hernández-Díaz¹, Roger Logan¹, James M. Robins^{1,2}, and Miguel A. Hernán^{1,3}

¹Department of Epidemiology, Harvard School of Public Health, Boston, MA

²Department of Biostatistics, Harvard School of Public Health, Boston, MA

³Harvard-MIT Division of Health Sciences and Technology, Boston, MA

Abstract

The intention-to-treat (ITT) analysis provides a valid test of the null hypothesis and naturally results in both absolute and relative measures of risk. However, this analytic approach may miss the occurrence of serious adverse effects that would have been detected under full adherence to the assigned treatment. Inverse probability weighting of marginal structural models has been used to adjust for nonadherence, but most studies have provided only relative measures of risk. In this study, we used inverse probability weighting to estimate both absolute and relative measures of risk of invasive breast cancer under full adherence to the assigned treatment in the Women's Health Initiative estrogen-plus-progestin trial. In contrast to an ITT hazard ratio (HR) of 1.25 (95% confidence interval [CI] = 1.01 to 1.54), the HR for 8-year continuous estrogen-plus-progestin use versus no use was 1.68 (1.24 to 2.28). The estimated risk difference (cases/100 women) at year 8 was 0.83 (−0.03 to 1.69) in the ITT analysis, compared with 1.44 (0.52 to 2.37) in the adherence-adjusted analysis. Results were robust across various dose-response models. We also compared the dynamic treatment regime “take hormone therapy until certain adverse events become apparent, then stop taking hormone therapy” with no use (HR= 1.64; 95% CI = 1.24 to 2.18). The methods described here are also applicable to observational studies with time-varying treatments.

The primary analysis of most randomized trials follows the intention-to-treat (ITT) principle. The ITT analysis is favored because it provides a valid test of the null hypothesis in placebo-controlled trials -- bypassing the problems associated with imperfect adherence to the assigned treatment -- and because the absence of adjustment for covariates naturally yields both absolute and relative measures of risk.

However, the ITT effect is a biased estimate of any true non-null effect that would have been observed under full adherence to the assigned treatment.¹ The greater the nonadherence to the assigned treatment, the closer to the null the ITT effect is expected to be in placebo-controlled studies. Thus, in studies whose goal is evaluating a treatment's safety, one could naively conclude that a treatment is safe because the ITT effect is close to null, even if the treatment causes serious adverse effects that would have been detected in the absence of nonadherence.

Address for correspondence: Sengwee Darren Toh, Department of Population Medicine, Harvard Medical School/Harvard Pilgrim Health Care Institute, 133 Brookline Ave 6th Floor, Boston, MA 02215, USA, darrentoh@post.harvard.edu, Tel: +1 617 509 9818, Fax: +1 617 859 8112.

To deal with nonadherence, one can attempt to estimate the effect that would have been observed had all study participants adhered to their assigned treatment throughout the follow-up, sometimes referred to as the effect of continuous treatment. Inverse probability weighting can be used to consistently estimate the effect of continuous treatment,^{2–5} but only under exchangeability and modeling assumptions that are not required to estimate the ITT effect. G-estimation of structural nested models that uses assigned treatment as an instrumental variable can also be used under a different set of assumptions.^{4,6–8} The wish to conduct an analysis whose validity does not rely on those assumptions might explain the widespread use of the ITT analysis, despite its shortcomings.

Here we describe the application of inverse probability weighting of marginal structural models to estimate both absolute and relative measures of risk under full adherence. To illustrate the use of inverse probability weighting, we estimated the effect of continuous postmenopausal hormone therapy on the risk of invasive breast cancer (henceforth, breast cancer) in the Women's Health Initiative estrogen-plus-progestin trial.

METHODS

Study population

The Women's Health Initiative estrogen-plus-progestin trial was a double-blinded, placebo-controlled, and multi-centered primary-prevention trial in which 16,608 postmenopausal women aged 50–79 years with an intact uterus at baseline were randomized to either a daily hormone regime of 0.625mg conjugated equine estrogens plus 2.5mg medroxyprogesterone acetate (n = 8506) or matching placebo (n = 8102) between 1993 and 1998.⁹ A detailed description of the trial has been published elsewhere.^{9,10} The limited access dataset we used (obtained from the National Heart, Lung, and Blood Institute) includes follow-up information through 7 July 2002, for an average follow-up of 5.6 years.

Data were collected at baseline and during the follow-up period on demographic characteristics; medical, reproductive, and family history; hormone use; dietary intake; and physical examinations. Safety and adherence data were recorded first at six weeks after randomization, followed by scheduled semi-annual interviews and annual clinical visits during which health-related information was also updated. For each follow-up year, the dataset contains indicators for discontinuation of assigned study pills and initiation of non-study hormone therapy, as well as the proportion of study pills taken (estimated by weighing of returned bottles) and self-reported frequency of use. (We re-coded doses less than 1% in a given year as zero.) Physician adjudicators at local clinics first confirmed self-reported breast cancer cases by reviewing medical records and pathology reports, and all cases were subsequently centrally adjudicated using the Surveillance, Epidemiology, and End Results (SEER) coding system.^{11,12}

ITT analysis

For quality assurance purposes, we first replicated the published estimates of the average ITT hazard ratio of hormone therapy versus placebo^{12,13} through a pooled logistic regression model¹⁴ that included months since randomization (modeled by cubic splines), age at baseline, and randomization status in a parallel diet-modification trial. We estimated the average ITT hazard ratio over the first 2-, 6-, and 8-year periods. We repeated the analysis stratified by prior hormone use (yes, no), age (<60 and ≥60 years old), years since menopause (<10 and ≥10 years).

To estimate breast cancer-free survival curves for initiators and non-initiators, we followed two separate approaches. First, we constructed unadjusted Kaplan-Meier survival curves, as previously reported by the Women's Health Initiative investigators.^{12,13} Second, we

estimated standardized survival curves from a pooled logistic model that allowed for time-varying hazard ratios by including product (“interaction”) terms between treatment arm and months since randomization, in addition to the covariates listed above. To estimate the standardized (by the baseline covariates) survival at month m for therapy initiation, we multiplied the predicted conditional probabilities of surviving to month $t \leq m$ given survival through $t-1$ for each study participant including those in the placebo arm (by setting the treatment arm indicator equal to 1 in the fitted pooled logistic model), and then averaged the estimate across all participants. To estimate the standardized survival under no initiation, we repeated the process with the treatment arm indicator set equal to 0. In the absence of model misspecification, the unadjusted and the standardized curves should be similar.

We then computed, from each standardized curve, the estimated absolute risk of breast cancer (cumulative incidence) as 1 minus the disease-free survival at 2, 6, and 8 years, and the corresponding risk differences for initiators versus non-initiators. The 95% confidence interval (CI) for the risk difference was estimated by 200 bootstrap samples (with replacement).

Adherence-adjusted analysis

We estimated the hazard ratio for continuous hormone use versus no use (an adherence-adjusted effect) using inverse probability weighting.²⁻⁵ Informally, the method weights each woman at each time period by the inverse of the density of having received her actual treatment history through that time. The density was computed as the product of the probability of receiving any hormone therapy during each follow-up year and, for those with non-zero use during that year, the density of receiving the proportion of pills she actually took. We estimated these quantities by fitting, separately for each arm, (1) a logistic regression model to estimate the probability of receiving any hormone therapy, and (2) a linear regression model that assumed independent normal errors with constant variance to estimate the density of receiving the log-transformed proportion of pills taken among those with non-zero use during that year.^{15,16} A participant contributed as many observations to the models as years she was in the study, i.e., from baseline to the occurrence of breast cancer, death, loss to follow-up, or end of study, whichever occurred first.

Both models included years since randomization (linear and quadratic) and covariates listed in Table 1 measured at baseline and, for time-varying covariates, at the most recent visit. Fitting a richer model with body pain (none, very mild, mild, moderate/severe) or use of statins, aspirin, selective estrogen receptor modulators, bisphosphonates, or multivitamin (yes, no) did not materially change the results (not shown). As noted above, we assumed a normal density for the error term in the linear regression model, with a constant variance across all combinations of the covariates. The results (not shown) were similar when we used alternative distributions (gamma, log-normal), when we used an arcsin-root transformation instead of log transformation, or when we allowed the error variance to depend on a subset of the covariates.

To improve statistical efficiency, the weights were stabilized^{2-5,17} by setting the numerator weight equal to the estimated probability (density) of received treatment history conditional on the baseline covariates included in the ITT model plus the following subset of the baseline covariates used in the model for the denominator of the weights: race/ethnicity, marital status, body mass index, physical activity, cigarette smoking, alcohol intake, parity, age at menarche, family history of breast cancer, mammography use, presence of vasomotor symptoms, use of oral contraceptive, prior hormone use, and years since menopause. Adding to the numerator weight model the covariates family history of fracture, region, education level, number of children breastfed, age at first birth, personal history of benign breast diseases, bilateral oophorectomy, general health, fruit and vegetable intake, and number of

first-degree relatives with breast cancer yielded similar results (not shown). The mean of the estimated stabilized inverse probability weights for adherence adjustment was 1.00 (standard deviation = 0.29). We did not attempt to adjust for selection bias arising from differential drop-out²⁻⁴ because of the low proportion of participants lost to follow-up (3.3%).

We then fitted a weighted model identical to the one used to estimate average ITT hazard ratios, except that (1) the treatment arm indicator was replaced by a time-varying covariate for cumulative use of hormone therapy, calculated as the sum of the annual proportion of pills taken since baseline, and (2) included the additional baseline covariates used to estimate the numerator of the weights. We used a nonparametric bootstrap estimator¹⁸ to calculate conservative 95% CIs for the average hazard ratio. We added product terms between cumulative use and indicators for prior hormone use (yes, no), age (<60, ≥60 years), and time since menopause (<10, ≥10 years) to obtain the corresponding stratum-specific estimates of the average hazard ratio. Results were similar under a number of sensitivity analyses described in Appendix 1.

Note that we specified a structural model with a dose-response function (cumulative use of hormone therapy), rather than using a “model-free” approach in which participants are censored when they become non-adherent, as in previous studies.¹⁹⁻²¹ The model-free approach could not be used because the Women’s Health Initiative data did not include sufficient information to establish the temporal sequence between nonadherence to assigned treatment and breast cancer diagnosis within each 1-year observation period. On the other hand, the uncertainty surrounding the time sequence has a small effect on our estimates because (1) the uncertainty is restricted to a short interval during which only a relatively small amount of treatment can be taken, and (2) our dose-response function (cumulative treatment use) assumes that a small amount of treatment has little effect compared with a large amount of treatment. “Average treatment use” is an example of an alternative dose-response function that meets condition 2. “Current treatment use”, is an example that does not.

This application extends previous applications of inverse probability weighting for survival analysis^{5,22} to the case of time-varying treatments and dose-response functions for the effect of treatment on disease-free survival. Annotated SAS programs used for this analysis can be found in the eAppendix (<http://links.lww.com>).

Secondary adherence-adjusted analysis

In the Women’s Health Initiative study protocol, participants were required to permanently stop their assigned hormone therapy if they developed the following events: deep vein thrombosis, pulmonary embolus, endometrial hyperplasia with atypia, malignant melanoma, endometrial cancer, breast cancer, triglycerides above 1,000 mg/dL, or starting anticoagulant medications, estrogen, progesterone, testosterone, tamoxifen, or other selective estrogen-receptor modulators. We performed a secondary analysis to estimate the survival curve that would have been observed if all women in the hormone therapy arm had fully complied with this protocol. More specifically, we estimated survival under the dynamic regime “take hormone therapy until one of the above events occurs, then stop taking hormone therapy.” To do so, we artificially censored participants in the hormone arm at the time they deviated from the protocol (i.e., did not stop taking their assigned study hormone after they had one of these events). This artificial censoring may result in selection bias because the distribution of risk factors of breast cancer may differ between the censored and the uncensored.

To adjust for such potential selection bias, one would estimate time-varying, subject-specific inverse probability weights whose denominator is the subject’s estimated probability of

remaining uncensored at each time, conditional on past joint predictors of censoring and the outcome. Note, however, that the predictors of censoring at time t are in fact the predictors of hormone therapy continuation at t because those who continue to take their study pills are precisely those who are censored. Therefore there is no need to estimate separate inverse probability weights to adjust for selection bias due to artificial censoring because the treatment weights estimated in the primary analysis already adjust for the potential time-varying selection bias due to artificial censoring. Also note that most protocol-mandated reasons for stopping treatment were not risk factors for breast cancer and thus need not be used to estimate the weights.

The specification of a marginal structural model for dynamic regimes is straightforward when only two regimes are compared²³ or when the goal of the structural model is to smooth over a set of regimes that can be placed in 1-to-1 correspondence with an indexing continuous variable.^{8,24} The situation is not so straightforward when, as is required in this example, the dose-response function depends on a summary dose measure (such as cumulative use) that can take the same value for many different regimes. See Appendix 2 for further discussion and caveats. For simplicity, we used the same specification as in the primary analysis; specifically we assumed the log discrete hazard is a linear function of cumulative dose.

RESULTS

ITT estimates

We reproduced the ITT hazard ratio estimates published previously by the Women's Health Initiative investigators (Table 2). Compared with women assigned to placebo, women assigned to estrogen-plus-progestin were 25% more likely to be diagnosed with breast cancer during the first 8 years of follow-up, but 29% less likely to have the diagnosis during the first 2 years. Only 2 cases occurred after 8 years, and thus the hazard ratios over the entire study period (not shown) were virtually identical to those in Table 2.

Figure 1 shows breast cancer-free survival curves for initiators versus non-initiators among all women, as well as women without and with prior hormone use. Kaplan-Meier (unadjusted) and standardized (adjusted) curves were similar, which suggests that our model for the standardized curves was adequately specified. The p-value from a log-rank test for the equality of the survival curves between initiators and non-initiators was 0.04 for all women, 0.45 for women without prior hormone use, and 0.01 for women with prior hormone use. The ITT risk differences at 2, 6, and 8 years are shown in Table 2. By the end of 8 years of follow-up, the estimated risk difference (cases per 100 women) of breast cancer was 0.83 (95% CI = -0.03 to 1.69).

Adherence-adjusted estimates

A selected list of potential predictors of hormone use is shown in Table 3 (with all covariates coded as categorical for simplicity). Women aged 70 years or more and non-whites were less likely to use hormone therapy in both treatment arms. Examples of time-varying predictors of hormone use in the treatment arm include normal mammogram/breast exam (odds ratio of receiving hormone was 0.3 for participants with versus without abnormal results); no new lumps, nipple discharge, or skin changes (odds ratio = 0.6); oophorectomy (odds ratio = 1.6); weight loss (odds ratio = 0.7 for women whose body mass index decreased by more than 0.5 kg/m² from baseline); and changes in physical activity (odds ratio = 0.7 for women who reduced their physical activity level by more than 0.5 metabolic equivalent units/week).

In general, the adherence-adjusted hazard ratios (Table 4) were further away from the null than the ITT estimates, but the corresponding 95% CIs were also wider. The adherence-adjusted survival curves (Figure 2) crossed at about 4 years in women without prior hormone use before randomization, and diverged at approximately 1 year after initiation in women with prior hormone use. The p-value from a log-rank test for the equality of the survival curves was <0.0001 for all women, <0.01 for women without prior hormone use, and <0.001 for women with prior hormone use. By the end of 8 years of follow-up, the estimated risk difference (cases per 100 women) of breast cancer was 1.44 (95% CI = 0.52 to 2.37) (Table 4). Risk differences obtained from different dose-response models (Table 5) were qualitatively similar.

In our secondary adherence-adjusted analysis, the hazard ratio for 8-year continuous hormone use was 1.64 (95% CI = 1.24 to 2.18) for all women, 1.59 (1.13 to 2.24) for women without prior hormone use, and 2.04 (1.34 to 3.10) for those with prior hormone use. The corresponding 8-year risk differences (cases per 100 women) were 1.56 (95% CI = 0.68 to 2.44), 1.45 (0.30 to 2.60), and 1.61 (0.60 to 2.62).

DISCUSSION

We have presented an application of inverse probability weighting to adjust for incomplete adherence to the assigned treatment in randomized trials. Our analysis estimated that, relative to no hormone therapy, the incidence of breast cancer under 8-year continuous estrogen-plus-progestin therapy was 1.7 times greater (the ITT estimate was 1.3). In absolute terms, we estimated that continuous use of estrogen-plus-progestin therapy for 8 years caused an excess of 1.4 breast cancer cases per 100 women (the ITT estimate was 0.8). Interestingly, women on estrogen-plus-progestin therapy had a lower incidence of breast cancer during the first two years of use, which is consistent with the hypothesis that estrogen-plus-progestin use delays breast cancer diagnosis, possibly by compromising the diagnostic performance of mammogram and breast biopsies.^{12,25} Thus, as in previous Women's Health Initiative analyses,^{12,13} it may be more appropriate to say that our analysis estimated the effect of continuous therapy on the diagnosis, rather than the true incidence, of breast cancer.

Inverse probability weighting adjusts for the joint time-varying predictors of adherence and breast cancer that were measured in the study, but cannot adjust for unmeasured predictors of adherence.⁸ G-estimation (a general form of instrumental variable analysis) is another approach that can adjust for measured and unmeasured predictors⁶⁻⁸ and that has been applied in several randomized trials.^{4,26-31} We have previously compared the assumptions for the validity of adherence-adjusted estimates when using adjusted inverse probability weighting and g-estimation in randomized trials in which the outcome of interest is a continuous variable.⁴

In contrast with inverse probability weighting and g-estimation, other adjustment methods may not appropriately adjust for measured predictors of adherence and survival.³² For example, a previously-used approach applies to the Women's Health Initiative data censored participants 6 months after they took <80% of their study pills or started receiving non-study hormone. The hazard ratio of breast cancer was 1.49 (95% CI = 1.13 to 1.96).¹² Such approach requires the choice of an arbitrary censoring period, as well as the assumption that reasons for stopping are not related to time-varying risk factors for breast cancer and the measured time-varying risk factors that also predict future adherence are not themselves affected by prior adherence. This assumption may be violated if, for example, women in the hormone-treatment arm were less likely to adhere to their assigned treatment. In this study, 42% of the women in the hormone arm stopped taking their study pills some time during the

follow-up, compared with 38% in the placebo arm; 6% and 11%, respectively initiated non-study hormone therapy.⁹ Further, women with joint determinants of hormone use and breast cancer that are also affected by previous hormone therapy, such as abnormalities in mammogram and breast exam, were less likely to adhere to their assigned treatment.

Besides appropriately adjusting for the measured time-varying determinants of adherence, our analytic approach allows estimation of the absolute risk under continuous therapy and the exploration of the sensitivity of the estimates to the functional form of dose-response function. Had the available Women's Health Initiative data been sufficient to establish the temporal sequence between nonadherence and breast cancer diagnosis, we could have done away with the requirement to specify a dose-response function, at the expense of wider confidence intervals for our estimates.⁴ In previous studies^{19,20} we did not have to specify a dose-response function because the temporal sequence could be inferred from the data.

Both inverse probability weighting and g-estimation can be extended to the estimation of the effect of dynamic treatment regimes.^{8,23} This extension is crucial because in some cases estimating the effect of continuous treatment (a non-dynamic or static regime) may be of little interest. For example, if many participants stop taking the treatment because it causes serious adverse effects, one would not want to estimate the effect under the static regime "always adhere to the baseline treatment" (the effect of continuous treatment) but rather under the dynamic regime "adhere to the baseline treatment unless adverse effects become apparent." Further, when certain types of participants will always discontinue treatment given certain adverse events, then estimating the effect under static regimes such as "always adhere to the baseline treatment" is problematic because the positivity assumption is violated.^{17,33}

The Women's Health Initiative protocol required participants who experienced certain adverse events to stop permanently their assigned hormone treatment. A treatment regime that changes with patients' prognosis and response to previous therapy may increase the effectiveness of the treatment or may reduce the adverse effects associated with the treatment. In our breast cancer analysis, the effect on breast cancer risk of the dynamic regime "adhere to the baseline hormone treatment unless adverse effects occurred" was similar to the effect of continuous hormone use.

Loss to follow-up is another common problem in randomized trials. In the presence of loss to follow-up, an analysis under the ITT principle is not feasible because some participants' outcomes are unknown. As a result, some studies use a "pseudo-ITT" analysis⁴ that does not preserve the desirable properties of the ITT analysis. Inverse probability weighting can also be used to adjust for selection bias due to differential loss to follow-up,²⁻⁴ but that was not necessary in this study because few participants dropped out.

In conclusion, we described the application of inverse probability weighting of a dose-response marginal structural model to estimate both absolute and relative measures of effect on a failure-time outcome. Although we focused on randomized trials for simplicity, the methods described here are also applicable to observational studies with time-varying treatments.

Acknowledgments

Sources of financial support: This study was partially funded by National Institutes of Health (NIH) grant R01 HL080644-01

The Women's Health Initiative: Estrogen plus Progestin Trial is conducted and supported by the NHLBI in collaboration with the WHI Study Investigators. This manuscript was prepared using data obtained from the NHLBI and does not necessarily reflect the opinions or views of the WHI or the NHLBI.

References

1. Robins JM. Correction for non-compliance in equivalence trials. *Stat Med.* 1998; 17(3):269–302. discussion 387-9. [PubMed: 9493255]
2. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000; 11(5):550–60. [PubMed: 10955408]
3. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000; 11(5):561–70. [PubMed: 10955409]
4. Toh, S.; Hernán, MA. Causal inference from longitudinal studies with baseline randomization; *Int J Biostat.* 2008. p. Article 22 Available at: <http://www.bepress.com/ijb/vol4/iss1/22>
5. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics.* 2000; 56(3):779–88. [PubMed: 10985216]
6. 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., editors. *Health Service Research Methodology: A Focus on AIDS.* Washington, D.C: NCHSR, U.S. Public Health Service; 1989. p. 113-159.
7. Robins, JM. Analytic methods for estimating HIV treatment and cofactor effects. In: Ostrow, DG.; Kessler, R., editors. *Methodological Issues of AIDS Mental Health Research.* New York: Plenum Publishing; 1993. p. 213-290.
8. Robins, JM.; Hernán, MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G.; Molenberghs, G., editors. *Longitudinal Data Analysis.* New York: Chapman and Hall/CRC Press; 2008. p. 553-599.
9. The Writing Group for the Women’s Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women’s Health Initiative randomized controlled trial. *JAMA.* 2002; 288(3):321–33. [PubMed: 12117397]
10. The Women’s Health Initiative Study Group. Design of the Women’s Health Initiative clinical trial and observational study. *Control Clin Trials.* 1998; 19(1):61–109. [PubMed: 9492970]
11. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S. Outcomes ascertainment and adjudication methods in the Women’s Health Initiative. *Ann Epidemiol.* 2003; 13(9 Suppl):S122–8. [PubMed: 14575944]
12. Chlebowski RT, Hendrix SL, Langer RD, Stefanick ML, Gass M, Lane D, Rodabough RJ, Gilligan MA, Cyr MG, Thomson CA, Khandekar J, Petrovitch H, McTiernan A. Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women’s Health Initiative Randomized Trial. *JAMA.* 2003; 289(24):3243–53. [PubMed: 12824205]
13. Anderson GL, Chlebowski RT, Rossouw JE, Rodabough RJ, McTiernan A, Margolis KL, Aggerwal A, David Curb J, Hendrix SL, Allan Hubbell F, Khandekar J, Lane DS, Lasser N, Lopez AM, Potter J, Ritenbaugh C. Prior hormone therapy and breast cancer risk in the Women’s Health Initiative randomized trial of estrogen plus progestin. *Maturitas.* 2006; 55(2):103–15. [PubMed: 16815651]
14. Thompson WA Jr. On the treatment of grouped observations in life studies. *Biometrics.* 1977; 33(3):463–70. [PubMed: 911970]
15. Cotter D, Zhang Y, Thamer M, Kaufman J, Hernan MA. The effect of epoetin dose on hematocrit. *Kidney Int.* 2008; 73(3):347–53. [PubMed: 18004296]
16. Toh S, Hernández-Díaz S, Logan R, Rossouw JE, Hernán MA. Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: does the increased risk ever disappear? A randomized trial. *Ann Intern Med.* 2010; 152(4):211–7. [PubMed: 20157135]
17. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health.* 2006; 60(7):578–86. [PubMed: 16790829]
18. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 1986; 73(1):13–22.
19. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM. Observational studies analyzed like randomized experiments: an application to postmenopausal

- hormone therapy and coronary heart disease. *Epidemiology*. 2008; 19(6):766–79. [PubMed: 18854702]
20. Hernán MA, Robins JM, García Rodríguez LA. Discussion of Statistical issues arising in the Women's Health Initiative by Prentice RL, Pettinger M, Andreson GL. *Biometrics*. 2005; 61(4): 922–930.
 21. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Stat Med*. 2009; 28(12):1725–38. [PubMed: 19347843]
 22. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004; 75(1):45–9. [PubMed: 15158046]
 23. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006; 98(3):237–42. [PubMed: 16611197]
 24. Robins J, Orellana L, Rotnitzky A. Estimation and extrapolation of optimal treatment and testing strategies. *Stat Med*. 2008; 27(23):4678–721. [PubMed: 18646286]
 25. Chlebowski RT, Anderson G, Pettinger M, Lane D, Langer RD, Gilligan MA, Walsh BW, Chen C, McTiernan A. Estrogen plus progestin and breast cancer detection by means of mammography and breast biopsy. *Arch Intern Med*. 2008; 168(4):370–7. [PubMed: 18299491]
 26. 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(2):79–97. [PubMed: 8500308]
 27. Mark SD, Robins JM. Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Stat Med*. 1993; 12(17): 1605–28. [PubMed: 8235180]
 28. Korhonen PA, Laird NM, Palmgren J. Correcting for non-compliance in randomized trials: an application to the ATBC Study. *Stat Med*. 1999; 18(21):2879–97. [PubMed: 10523748]
 29. Joffe MM. Administrative and artificial censoring in censored regression models. *Stat Med*. 2001; 20(15):2287–304. [PubMed: 11468764]
 30. Cole SR, Chu H. Effect of acyclovir on herpetic ocular recurrence using a structural nested model. *Contemp Clin Trials*. 2005; 26(3):300–10. [PubMed: 15911464]
 31. Greenland S, Lanes S, Jara M. Estimating effects from randomized trials with discontinuations: the need for intent-to-treat design and G-estimation. *Clin Trials*. 2008; 5(1):5–13. [PubMed: 18283074]
 32. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004; 15(5):615–25. [PubMed: 15308962]
 33. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008; 168(6):656–64. [PubMed: 18682488]

Appendix 1. Primary analysis: Sensitivity analyses

Missing data on adherence

For women with missing proportion of study pills taken as estimated by weighing of returned bottles (28% of the total person-time), we estimated this proportion based on their self-reported frequency of use (none, <1, 1–2, 3–4, 5–6, 7 days/week). When the self-report use was also missing (71% of the person-time in women with missing estimated proportion of study pills), we randomly assigned a dose for that year using a uniform distribution. In addition, because the proportion of pills taken is unknown for women who initiated non-study hormone therapy, we randomly assigned a dose for such use. We repeated our analysis under various assumptions for the imputation of missing study and non-study hormone use (using all, half, or none of the usual proportion of pills, or using the same proportion as the previous year). Results (not shown) were similar under all these assumptions. We were also not able to accurately identify the dose a participant received during the year she had the outcome. A number of sensitivity analyses were performed to test the robustness of our

results. These included assuming (1) all participants stopped taking the study pills at the time of outcome; (2) all participants kept taking the study pills until the end of the year; and (3) participants stopped taking the study pills at a random month. Our results were similar under these different scenarios. The third approach was used as the main analysis.

Density estimation

We evaluated the sensitivity of our estimates to the method of density estimation by varying several components of the analysis separately. First, we used a gamma, rather than a normal, distribution of the log-transformed proportion of dose received. Second, we considered a log-normal distribution for the original, non-transformed, proportion of dose received. Third, we used an arcsin-root, rather than a log, transformation. For women with both study and non-study hormone use whose combined proportion of hormone use was greater than 1, we recoded their proportion as 1. Fourth, we relaxed the assumption of constant variance across treatment and covariate histories. We estimated the conditional variances by regressing the squared residuals from the linear model for the (log-transformed) dose on the covariates. Results were similar under all these models although, when estimating conditional variances, we had to restrict the analysis to a subset of the covariates to avoid extreme stabilized weights.

Dose-response function

To assess the robustness of our estimates to the dose-response function, we also considered models that included quadratic term of cumulative use, average linear cumulative use, or average linear and quadratic cumulative use. Results were qualitatively similar under all these models (Table 5).

Appendix 2. Secondary analysis: Model specification and dynamic treatment regimes

Given a static treatment regime $\bar{a} = \{a_t; t \geq 0\}$, let the dynamic treatment regime $d(\bar{a})$ be the regime “follow the static regime \bar{a} until the occurrence of an adverse event; then take no more hormone therapy.” The marginal structural model used in our secondary adherence-adjusted analysis is formally a model for the logit of the discrete hazard of the counterfactual time to breast cancer diagnosis $T_{d(\bar{a})}$ under regime $d(\bar{a})$ given the baseline covariates V . Specifically, the simplest version of the model assumes that

$$\text{logitPr}(T_{d(\bar{a})} \leq m+1 | T_{d(\bar{a})} > m, V) = \beta_0(m) + \beta_1 \text{cum}(\bar{a}_m) + \beta_2 V$$

where $\text{cum}(\bar{a}_m) = \sum_{j=0}^m a_j$ is the cumulative dose under regime $d(\bar{a})$ up to month m if no adverse event occurs. The corresponding marginal structural model used in our primary analysis was the static regime model $\text{logitPr}(T_{\bar{a}} \leq m+1 | T_{\bar{a}} > m, V) = \alpha_0(m) + \alpha_1 \text{cum}(\bar{a}_m) + \alpha_2 V$.

We now argue by example that, even if (1) the primary analysis marginal structural model was correct with $\alpha_1 > 0$ (so increasing cumulative dose increases the risk of breast cancer for static regimes), and (2) the counterfactual times $C_{\bar{a}}$ to an adverse event and $T_{\bar{a}}$ to breast cancer are independently distributed given V , we have no guarantee that our secondary analysis model is correctly specified (even qualitatively). To show this, let $\bar{a} = c$ denote the treatment history equal to the constant c at all times so $\text{cum}(\bar{a}) = cm$ when $\bar{a} = c$. Also note

that, by the definition of the regime $d(\bar{a})$, $T_{\bar{a}=0} = T_{d(\bar{a}=0)}$. Suppose that, for all subjects, an adverse event will never occur if the daily dose rate is maintained below $1/2$, but will occur immediately if a dose of $1/2$ or greater is ever taken. Then one can see that, to a close approximation, $T_{\bar{a}=0} = T_{d(\bar{a}=c)}$, for $c \geq 1/2$ (since the cumulative dose actually taken will be essentially zero as the adverse events will dictate no further treatment). In contrast $T_{d(\bar{a}=c)} = T_{\bar{a}=c}$ for $c < 1/2$ because no adverse events will occur. Thus we conclude that regime $d(\bar{a} = c)$ increases the risk of breast cancer as c increases to $1/2$, and then abruptly, beginning at $c=1/2$, has no adverse effect on breast cancer. This dose response is a highly non-monotone function of c and thus of the cumulative dose cm up to any time m ; showing that the model used in the secondary analysis is badly misspecified in this example.

Furthermore two different dynamic regimes $d(\bar{a}^{(1)})$ and $d(\bar{a}^{(2)})$ that have essentially the same cumulative exposure in the absence of an adverse event can have very different effects on breast cancer. Specifically, in the context of the previous example, consider the regimes $\bar{a}^{(1)}$ and $\bar{a}^{(2)}$ that specify a dose of $3/4$ every other day and a dose of $3/8$ every day respectively. Then the regimes $d(\bar{a}^{(1)})$ and $d(\bar{a}^{(2)})$ will have essentially identical cumulative exposures in the absence of an adverse event. However $T_{d(\bar{a}^{(1)})}$ is essentially equal to $T_{\bar{a}=0}$; while $T_{d(\bar{a}^{(2)})}$ is equal to $T_{\bar{a}=c}$ for $c=3/8$, so the second regime causes much more breast cancer than the first. We conclude there is no function of cumulative exposure, even a non-monotonic one, that can be used to correctly model the discrete hazard of $T_{d(\bar{a})}$. Although, for pedagogic purposes, this example was purposely chosen to be extreme, it shows that we must be very concerned about the specification of the secondary analysis model for the Women's health Initiative.

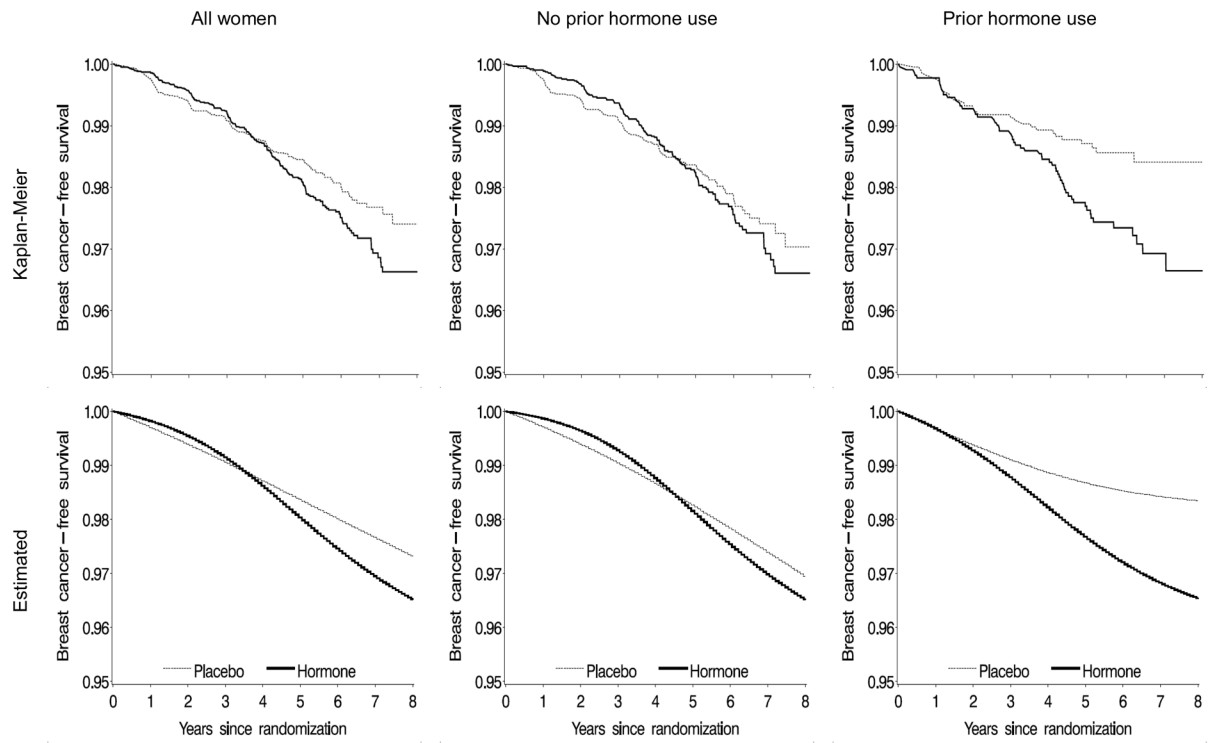


Figure 1. Proportion of women free of invasive breast cancer--intention-to-treat analysis, Women's Health Initiative estrogen-plus-progestin randomized trial. Kaplan-Meier (unadjusted) curves on top panel; standardized (adjusted) curves on bottom panel.

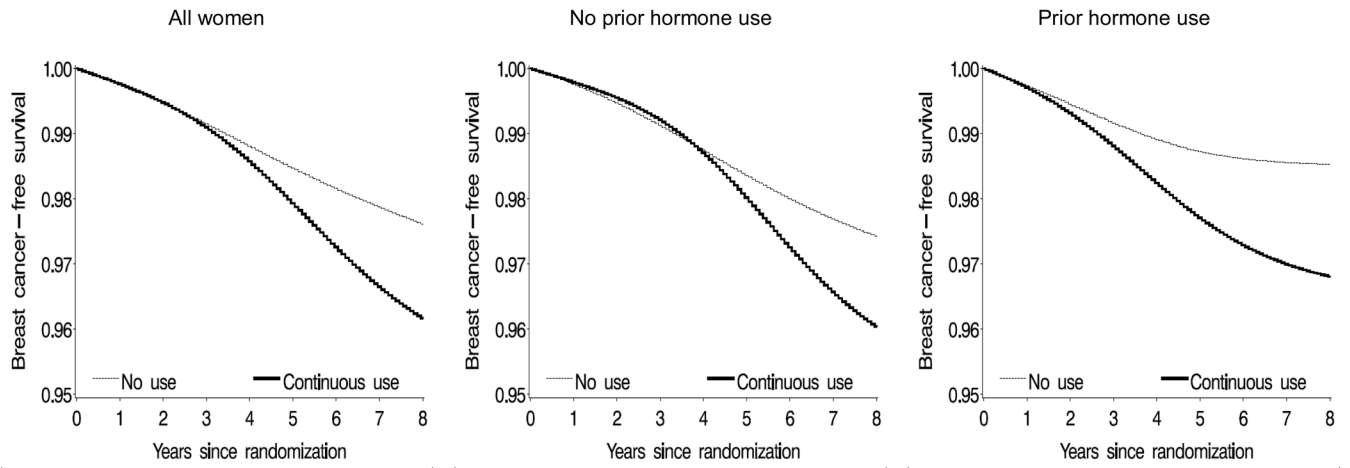


Figure 2. Estimated proportion of women free of invasive breast cancer under continuous adherence to the assigned treatment, Women’s Health Initiative estrogen-plus-progestin randomized trial

Table 1

Variables included in weight models, Women's Health Initiative estrogen-plus-progestin randomized trial

No.	Variables	Categories	Baseline	Time-varying
1	Randomization status in the diet-modification trial	Yes/No	Yes	No
2	Age	Five-year category	Yes	No
3	Region	Northeast, South, Midwest, West	Yes	No
4	Race/ethnicity	White, African-American, Hispanic, other	Yes	No
5	Education	High school or less, some college or associate degree, college graduate and above	Yes	No
6	Marital status	Never married, divorced or separated, widowed, presently married or married-like relationship	Yes	No
7	Cigarette smoking	Never, past, current	Yes	Yes
8	Physical activity	Quartiles of metabolic equivalent units/week	Yes	Yes
9	Alcohol intake	None, past, <1 drink/month, <1 drink/week, 1–7 drinks/week, >7 drinks/week	Yes	Yes
10	Body mass index	<25, 25–29, 30–34, 35–39, ≥40 kg/m ²	Yes	Yes
11	Age at menarche	<11, 12–13, ≥14 years	Yes	No
12	Number of term pregnancies	Never pregnant, never had term pregnancy, 1–2, 3–4, ≥5	Yes	No
13	Number of children breastfed	0, 1–2, ≥3	Yes	No
14	Age at first birth	Never pregnant/no term pregnancy, <20, 20–29, ≥30 years	Yes	No
15	Family history of breast cancer	Yes/No	Yes	No
16	Family history of fracture	Yes/No	Yes	No
17	Personal history of benign breast diseases	Yes/No	Yes	Yes
18	Personal history of cancer	Yes/No	Yes	Yes
19	Personal history of fracture	Yes/No	Yes	Yes
20	Personal history of osteoporosis	Yes/No	Yes	Yes
21	Personal history of cardiovascular diseases	Yes/No	Yes	Yes
22	Personal history of diabetes	Yes/No	Yes	Yes
23	Oophorectomy (unilateral or bilateral)	Yes/No	Yes	Yes
24	General health	Excellent, very good, good, poor/fair	Yes	Yes
25	Physical functioning construct	Quartiles of quality of life subscale on physical functioning in the RAND36-Item Health Survey	Yes	Yes
26	Prior hormone use	None, <5 years, 5–9 years, ≥10 years	Yes	No
27	Use of oral contraceptives ever	Yes/No	Yes	No
28	Fruit intake	Quartiles of daily intake	Yes	No
29	Vegetable intake	Quartiles of daily intake	Yes	No

No.	Variables	Categories	Baseline	
			Time-varying	Yes ^a
30	Mammogram	Yes/No	Yes ^a	Yes
31	Breast exam	Yes/No	No	Yes
32	Breast biopsy	Yes/No	No	Yes
33	Results from mammogram/breast exam	Normal/benign change, possibly malignant/cancer	No	Yes
34	Number of screening/diagnostic procedures	0, 1, 2, 3, 4, 5, ≥6	Yes	Yes
35	Years since menopause	<10, 10–19, ≥20 years	Yes	No
36	Vasomotor symptoms	None, mild, moderate/severe	Yes	Yes
37	Other menopausal symptoms	Quartiles of summary score of 15 menopausal symptoms	Yes	Yes
38	Changes in breasts ^b	Yes/No	No	Yes
39	Breast tenderness	None, mild, moderate/severe	No	Yes
40	Vaginal bleeding	None, spotting, light, moderate, severe	No	Yes

^aMammogram in last two years

^bIncludes new lumps, nipple discharge, and skin changes

Table 2

Intention-to-treat average hazard ratios (HR) and risk differences (RD) and risk differences (RD) (cases per 100 women) and their 95% confidence intervals (CIs) of invasive breast cancer, Women's Health Initiative estrogen-plus-progestin (E+P) randomized trial^a

	Follow-up period		
	0–2 years	0–6 years	0–8 years
Overall			
No. of cases; E+P/placebo	38/51	181/140	198/149
HR (95% CI)	0.71 (0.46 to 1.08)	1.23 (0.98 to 1.53)	1.25 (1.01 to 1.54)
RD (95% CI)	-0.16 (-0.37 to 0.05)	0.53 (0.03 to 1.03)	0.83 (-0.03 to 1.69)
Prior hormone use			
No			
No. of cases; E+P/placebo	22/36	127/112	140/120
HR (95% CI)	0.59 (0.34 to 1.00)	1.08 (0.84 to 1.39)	1.10 (0.86 to 1.40)
RD (95% CI)	-0.25 (-0.47 to -0.02)	0.27 (-0.33 to 0.87)	0.46 (-0.63 to 1.55)
Yes			
No. of cases; E+P/placebo	16/15	54/28	58/29
HR (95% CI)	1.02 (0.51 to 2.07)	1.82 (1.15 to 2.87)	1.87 (1.20 to 2.92)
RD (95% CI)	0.11 (-0.36 to 0.57)	1.31 (0.47 to 2.15)	1.84 (0.62 to 3.06)
Years since menopause^b			
< 10			
No. of cases; E+P/placebo	15/18	52/40	55/43
HR (95% CI)	0.80 (0.41 to 1.60)	1.26 (0.83 to 1.90)	1.23 (0.83 to 1.84)
RD (95% CI)	-0.11 (-0.47 to 0.24)	0.42 (-0.37 to 1.20)	0.48 (-0.61 to 1.57)
≥ 10			
No. of cases; E+P/placebo	20/31	109/91	118/95
HR (95% CI)	0.63 (0.36 to 1.10)	1.17 (0.89 to 1.55)	1.21 (0.93 to 1.59)
RD (95% CI)	-0.22 (-0.51 to 0.07)	0.57 (-0.09 to 1.24)	1.09 (-0.38 to 2.57)
Age group			
< 60			
No. of cases; E+P/placebo	10/15	47/35	52/39
HR (95% CI)	0.62 (0.28 to 1.38)	1.25 (0.81 to 1.94)	1.23 (0.81 to 1.86)

	Follow-up period		
	0–2 years	0–6 years	0–8 years
RD (95% CI)	–0.16 (–0.47 to 0.15)	0.41 (–0.28 to 1.10)	–0.19 (–1.51 to 1.14)
≥ 60			
No. of cases; E+P/placebo	28/36	134/105	146/110
HR (95% CI)	0.74 (0.45 to 1.22)	1.22 (0.94 to 1.57)	1.25 (0.98 to 1.61)
RD (95% CI)	–0.15 (–0.42 to 0.12)	0.64 (0.01 to 1.27)	1.46 (0.39 to 2.53)

^a Average hazard ratios in a given period (e.g., 0–2 years) were estimated over the entire period, while risk differences were estimated at the end of the period

^b 1,420 women (827 in estrogen-plus-progestin group and 593 in placebo group) had missing years since menopause

Characteristics associated with hormone use by randomized arm, Women's Health Initiative estrogen-plus-progestin randomized trial^a

Table 3

Covariates measured at baseline	Estrogen-plus-progestin arm ^b		Placebo arm ^c
	OR	(95% CI)	OR (95% CI)
Age (years)			
50–59 ^d	1.00		1.00
60–69	0.91 (0.76 to 1.09)		0.84 (0.71 to 1.00)
70–79	0.85 (0.67 to 1.10)		0.73 (0.55 to 0.96)
White race	1.18 (0.99 to 1.40)		1.25 (1.02 to 1.52)
Family history of breast cancer	1.06 (0.90 to 1.26)		0.88 (0.74 to 1.06)
Personal history of cancer	0.57 (0.41 to 0.80)		1.19 (0.76 to 1.87)
Personal history of benign breast diseases	1.08 (0.90 to 1.30)		1.05 (0.88 to 1.25)
Personal history of bilateral oophorectomy	0.80 (0.34 to 1.85)		0.96 (0.26 to 3.53)
Use of oral contraceptives	1.08 (0.95 to 1.23)		0.97 (0.85 to 1.11)
Mammogram in the last two years	0.94 (0.82 to 1.08)		0.73 (0.62 to 0.85)
Breastfeeding	0.95 (0.83 to 1.08)		0.95 (0.83 to 1.09)
Age at menarche (<12 vs. ≥12 years old)	0.98 (0.84 to 1.14)		0.96 (0.82 to 1.11)
At least one term pregnancies	0.80 (0.59 to 1.08)		0.93 (0.69 to 1.27)
Age at first birth (years)			
Never pregnant ^d	1.00		1.00
<30	1.22 (1.00 to 1.49)		0.97 (0.77 to 1.21)
≥30	1.21 (0.92 to 1.59)		1.05 (0.77 to 1.42)
Cigarette smoking			
Never ^d	1.00		1.00
Past	0.89 (0.77 to 1.01)		1.04 (0.91 to 1.19)
Current	0.85 (0.65 to 1.11)		0.95 (0.72 to 1.27)
Alcohol intake			
None ^d	1.00		1.00
Past	1.17 (0.93 to 1.48)		1.17 (0.90 to 1.51)

	Estrogen-plus-progestin arm ^b	Placebo arm ^c
	OR (95% CI)	OR (95% CI)
Current	1.12 (0.90 to 1.39)	1.09 (0.85 to 1.39)
Body mass index (kg/m ²)		
<25 ^d	1.00	1.00
25–29	0.95 (0.82 to 1.11)	0.86 (0.75 to 1.00)
30–34	1.18 (0.98 to 1.42)	0.72 (0.59 to 0.87)
≥35	0.91 (0.73 to 1.13)	0.61 (0.48 to 0.79)
Physical activity (metabolic equivalent units/week)		
None ^d	1.00	1.00
<50 th percentile	0.95 (0.78 to 1.16)	1.09 (0.89 to 1.35)
≥50 th percentile	0.92 (0.75 to 1.14)	1.15 (0.92 to 1.43)
Covariates measured at the most recent visit		
Other cancer	0.67 (0.47 to 0.96)	0.62 (0.32 to 1.17)
Abnormal mammogram or breast exam ^e	0.32 (0.19 to 0.55)	0.85 (0.20 to 3.65)
Changes in breasts ^f	0.56 (0.43 to 0.72)	1.76 (1.20 to 2.59)
Oophorectomy (unilateral or bilateral)	1.62 (0.79 to 3.30)	1.01 (0.31 to 3.29)
Cigarette smoking (yes/no)	0.78 (0.56 to 1.08)	1.02 (0.71 to 1.47)
Alcohol intake (yes/no)	0.99 (0.83 to 1.19)	0.93 (0.76 to 1.13)
Change in body mass index from baseline		
within 0.5 kg/m ² ^d	1.00	1.00
increased by more than 0.5 kg/m ²	0.81 (0.67 to 0.98)	0.87 (0.72 to 1.06)
decreased by more than 0.5 kg/m ²	0.74 (0.62 to 0.88)	0.98 (0.82 to 1.16)
Change in physical activity from baseline		
within 0.5 metabolic equivalent units (MEU)/week ^d	1.00	1.00
increased by more than 0.5 MEU/week	0.87 (0.68 to 1.11)	1.25 (0.96 to 1.62)
decreased by more than 0.5 MEU/week	0.74 (0.59 to 0.94)	1.19 (0.93 to 1.53)

^a Adjusted for all covariates in Table 1

^b Use of both study pills and non-study hormone

Toh et al.

Page 20

^c Use of non-study hormone

^d Reference category

^e Possible malignant or cancer

^f New lumps, nipple discharge, or skin changes

Table 4

Adherence-adjusted average hazard ratios (HR) and risk differences (RD) (cases per 100 women) and 95% confidence intervals (CIs) of invasive breast cancer, Women's Health Initiative estrogen-plus-progestin randomized trial^a

	Follow-up period		
	0–2 years	0–6 years	0–8 years
Overall			
HR (95% CI)	0.98 (0.72 to 1.33)	1.55 (1.21 to 1.97)	1.68 (1.24 to 2.28)
RD (95% CI)	-0.01 (-0.15 to 0.13)	0.89 (0.38 to 1.41)	1.44 (0.52 to 2.37)
Prior hormone use			
No			
HR (95% CI)	0.92 (0.63 to 1.34)	1.34 (1.01 to 1.78)	1.51 (1.06 to 2.14)
RD (95% CI)	-0.06 (-0.22 to 0.11)	0.73 (0.06 to 1.39)	1.31 (0.14 to 2.48)
Yes			
HR (95% CI)	1.15 (0.69 to 1.91)	1.92 (1.26 to 2.93)	2.04 (1.27 to 3.28)
RD (95% CI)	0.12 (-0.18 to 0.43)	1.33 (0.52 to 2.14)	1.69 (0.60 to 2.77)
Years since menopause^b			
< 10			
HR (95% CI)	1.30 (0.77 to 2.19)	1.43 (0.90 to 2.27)	1.34 (0.74 to 2.42)
RD (95% CI)	0.12 (-0.13 to 0.36)	0.66 (-0.16 to 1.47)	0.64 (-0.75 to 2.03)
≥ 10			
HR (95% CI)	0.88 (0.60 to 1.29)	1.52 (1.10 to 2.09)	1.84 (1.14 to 2.99)
RD (95% CI)	-0.09 (-0.28 to 0.10)	1.00 (0.23 to 1.77)	2.13 (0.14 to 4.12)
Age group			
< 60			
HR (95% CI)	0.96 (0.54 to 1.71)	1.58 (0.96 to 2.61)	1.44 (0.84 to 2.48)
RD (95% CI)	0.00 (-0.22 to 0.21)	0.53 (-0.24 to 1.29)	0.41 (-0.62 to 1.44)
≥ 60			
HR (95% CI)	0.99 (0.71 to 1.38)	1.49 (1.13 to 1.97)	1.81 (1.29 to 2.52)
RD (95% CI)	-0.02 (-0.20 to 0.16)	1.11 (0.44 to 1.78)	2.18 (0.90 to 3.45)

^a Average hazard ratios in a given period (e.g., 0–2 years) were estimated over the entire period, while risk differences were estimated at the end of the period

^b 1,420 women (827 in estrogen-plus-progestin group and 593 in placebo group) had missing years since menopause

Table 5

Adherence-adjusted risk differences (cases per 100 women) of invasive breast cancer for various dose-response models, Women's Health Initiative estrogen-plus-progestin randomized trial^a

	Base model	Model 1	Model 2	Model 3
		0-2 years		
Overall	-0.01 (-0.15 to 0.13)	-0.03 (-0.20 to 0.15)	-0.17 (-0.39 to 0.05)	-0.18 (-0.40 to 0.04)
Prior hormone use				
No	-0.06 (-0.22 to 0.11)	-0.08 (-0.28 to 0.13)	-0.31 (-0.54 to -0.08)	-0.33 (-0.55 to -0.10)
Yes	0.12 (-0.18 to 0.43)	0.10 (-0.29 to 0.50)	0.19 (-0.28 to 0.66)	0.22 (-0.28 to 0.71)
Years since menopause^b				
< 10	0.12 (-0.13 to 0.36)	0.12 (-0.16 to 0.39)	-0.12 (-0.38 to 0.15)	-0.09 (-0.37 to 0.18)
≥ 10	-0.09 (-0.28 to 0.10)	-0.16 (-0.41 to 0.08)	-0.23 (-0.53 to 0.07)	-0.25 (-0.55 to 0.06)
Age group (years)				
< 60	0.00 (-0.22 to 0.21)	0.04 (-0.18 to 0.26)	-0.18 (-0.44 to 0.09)	-0.17 (-0.41 to 0.07)
≥ 60	-0.02 (-0.20 to 0.16)	-0.05 (-0.29 to 0.19)	-0.18 (-0.45 to 0.10)	-0.19 (-0.47 to 0.10)
		0-6 years		
Overall	0.89 (0.38 to 1.41)	0.86 (0.32 to 1.41)	0.77 (0.25 to 1.30)	0.74 (0.20 to 1.27)
Prior hormone use				
No	0.73 (0.06 to 1.39)	0.63 (-0.04 to 1.30)	0.56 (-0.11 to 1.23)	0.45 (-0.21 to 1.12)
Yes	1.33 (0.52 to 2.14)	1.51 (0.57 to 2.44)	1.36 (0.52 to 2.20)	1.56 (0.60 to 2.53)
Years since menopause^b				
< 10	0.66 (-0.16 to 1.47)	0.71 (-0.16 to 1.59)	0.48 (-0.35 to 1.30)	0.54 (-0.33 to 1.42)
≥ 10	1.00 (0.23 to 1.77)	0.84 (0.06 to 1.63)	0.89 (0.11 to 1.67)	0.81 (0.03 to 1.59)
Age group				
< 60	0.53 (-0.24 to 1.29)	0.60 (-0.20 to 1.41)	0.38 (-0.40 to 1.16)	0.39 (-0.41 to 1.19)
≥ 60	1.11 (0.44 to 1.78)	1.03 (0.32 to 1.74)	1.00 (0.32 to 1.68)	0.94 (0.24 to 1.64)
		0-8 years		
Overall	1.44 (0.52 to 2.37)	1.35 (0.39 to 2.31)	1.31 (0.37 to 2.25)	1.28 (0.32 to 2.24)
Prior hormone use				
No	1.31 (0.14 to 2.48)	1.11 (-0.08 to 2.31)	1.05 (-0.17 to 2.28)	0.94 (-0.27 to 2.16)
Yes	1.69 (0.60 to 2.77)	1.95 (0.65 to 3.25)	1.77 (0.72 to 2.82)	2.04 (0.73 to 3.35)

	Base model	Model 1	Model 2	Model 3
Years since menopause^b				
< 10	0.64 (-0.75 to 2.03)	0.76 (-0.79 to 2.31)	0.27 (-0.92 to 1.46)	0.58 (-0.89 to 2.05)
≥ 10	2.13 (0.14 to 4.12)	2.16 (-0.23 to 4.55)	2.25 (0.27 to 4.23)	2.13 (0.09 to 4.16)
Age group				
< 60	0.41 (-0.62 to 1.44)	0.42 (-0.62 to 1.45)	0.26 (-0.81 to 1.33)	0.27 (-0.81 to 1.35)
≥ 60	2.18 (0.90 to 3.45)	2.02 (0.72 to 3.32)	2.08 (0.79 to 3.37)	1.99 (0.70 to 3.28)

Base model: Linear cumulative hormone use

Model 1: Linear and quadratic cumulative hormone use

Model 2: Average linear cumulative hormone use

Model 3: Average linear and quadratic cumulative hormone use

^aRisk differences in a given period (e.g., 0–2 years) were estimated at the end of the period

^b1,420 women (827 in estrogen-plus-progestin group and 593 in placebo group) had missing years since menopause