Women's Health Initiative estrogen plus progestin clinical trial: a study that does not allow establishing relevant clinical risks

Sócrates Aedo, MD, MSc,[1] Gabriel Cavada, MSc,[2] Juan E. Blümel, MD, PhD,[3] Peter Chedraui, MD, MSc, PhD,[4] Juan Fica, MD,[5] Patricio Barriga, MD,[6] Sergio Brantes, MD,[1] Cristina Irribarra, MD,[1] María Vallejo, MD,[7] and Ítalo Campodónico, MD[1]

Abstract

Objective: This study aims to determine time differences (differences in restricted mean survival times [RMSTs]) in the onset of invasive breast cancer, coronary heart disease, stroke, pulmonary embolism, colorectal cancer, and hip fracture between the placebo group and the conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg group of the Women's Health Initiative (WHI) trial based on survival curves of the original report and to provide adequate interpretation of the clinical effects of a given intervention.

Methods: Distribution of survival function was obtained from cumulative hazard plots of the WHI report; Monte Carlo simulation was performed to obtain censored observations for each outcome, in which assumptions of the Cox model were evaluated once corresponding hazard ratios had been estimated. Using estimation methods such as numerical integration, pseudovalues, and flexible parametric modeling, we determined differences in RMSTs for each outcome.

Results: Obtained cumulative hazard plots, hazard ratios, and outcome rates from the simulated model did not show differences in relation to the original WHI report. The differences in RMST between placebo and conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg (in flexible parametric modeling) were 1.17 days (95% CI, −2.25 to 4.59) for invasive breast cancer, 7.50 days (95% CI, 2.90 to 12.11) for coronary heart disease, 2.75 days (95% CI, −0.84 to 6.34) for stroke, 4.23 days (95% CI, 1.82 to 6.64) for pulmonary embolism, −2.73 days (95% CI, −5.32 to −0.13) for colorectal cancer, and −2.77 days (95% CI, −5.44 to −0.1) for hip fracture.

Conclusions: The differences in RMST for the outcomes of the WHI study are too small to establish clinical risks related to hormone therapy use.

Key Words: Women's Health Initiative – Menopausal hormone therapy – Menopause – Difference in restricted mean survival time.

I n 2002, the main results of the Women's Health Initiative (WHI) randomized controlled trial were published.[1] This trial was carried out to assess the major risks and benefits of using a combined hormone regimen of conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg (CEE/MPA) in healthy postmenopausal women. After a mean follow-up of 5.2 years, outcomes were evaluated using hazard ratios (HRs), which were obtained by Cox proportional hazards model analysis.[2] Compared with placebo, the use of CEE/MPA combination showed favorable effects on hip fracture and colorectal cancer and a higher risk for invasive breast cancer, coronary heart disease, stroke, and pulmonary embolism. These results caused an overall negative impact on menopausal hormone therapy prescription.[3-5]

Cox proportional hazards analysis is widely used for survival analysis in clinical studies. It uses HRs to estimate treatment effects, assuming that risk rates do not change across time among treatment groups (proportional hazards assumptions).[6] Nevertheless, several authors agreed that times using HRs may pose difficulty in interpreting clinical results.[6-9] Indeed, results that are statistically significant within a large sample size may not translate into a significant clinical difference.[10,11]

Restricted mean survival time (RMST) is a measure of mean survival from time 0 to a specified time point. In contrast to HRs obtained in Cox proportional hazards analysis, RMST does not require proportional hazards assumptions.[7-9] Royston and Parmar[7,8] proposed estimating and reporting RMST and expressing treatment effects as the difference in RMST between studied randomized arms at a suitable follow-up time point. Time difference in the onset of events should not be evaluated by HRs.[6] Rather, difference in RMST is used as a measure to determine interpretable differences in treatment effects (complementing the information provided by HRs) to ensure a correct interpretation of the intervention.[7-9]

To assess the clinical effects of CEE/MPA combination on postmenopausal women in the WHI study,[1] we proposed to determine differences in RMST[7,8] for the outcomes of the WHI study (invasive breast cancer, coronary heart disease, stroke, pulmonary embolism, colorectal cancer, and hip fracture) between placebo and CEE/MPA at 5.2-year follow-up.

METHODS

In the present study, we proceeded to obtain log-hazard and time values for the outcomes of the WHI study per treatment group (placebo and CEE/MPA) by means of Digitizelt version 2.03 (Ingo Bormann, 2013), using original published cumulative hazard plots.[1] With these results, the Stata program (Stata/IC version 13.1 for Windows; StataCorp LP, 2013) was used to perform log-hazard linear regression on log survival time. The linear regression provided rough estimates of the scale and shape parameters of the two survival distributions (placebo and CEE/MPA)[12,13]; Monte Carlo simulation was performed[12] to obtain randomized censored observations for each outcome of the WHI study.

Upon simulation, Nelson-Aalen survival estimator is used to obtain cumulative hazard plots for 7 years per outcome.[14] In addition, corresponding HRs through the Cox model[2] were determined per year of follow-up, and the respective log-rank tests were performed.[14] Proportional risk assumptions per
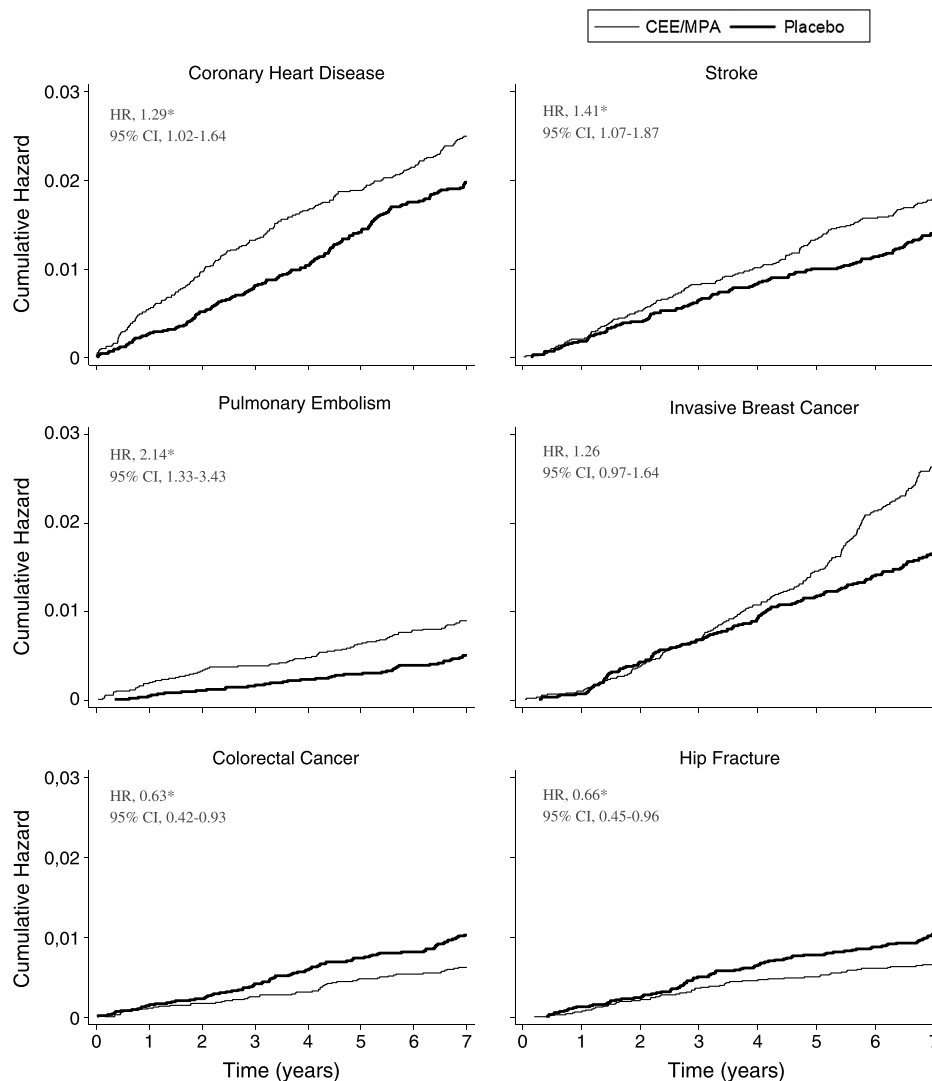
FIG. Kaplan-Meier survival curves for cumulative hazard function (Nelson-Aalen estimator), by outcome, for simulated observations of the Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women (Women's Health Initiative) study.[1] CEE/MPA, conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg; HR, hazard ratio at 5.2-year follow-up. *Log-rank test, $P < 0.05$.

TABLE 1. *Hazard ratios per outcome upon simulation and real values reported in the Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women (WHI) study*[1]

Outcome	Per year upon simulation in the WHI study							After 5.2 y (real values reported in the WHI study)
	1 y	2 y	3 y	4 y	5.2 y	6 y	7 y	
Coronary heart disease	2.04 (1.23 to 3.38)[a]	1.86 (1.29 to 2.7)[a]	1.62 (1.20 to 2.20)[a]	1.61 (1.23 to 2.10)[a]	1.29 (1.02 to 1.64)[a]	1.23 (0.99 to 1.63)	1.26 (1.03 to 1.55)[a]	1.29 (1.02 to 1.63)[a]
Stroke	1.14 (0.58 to 2.27)	1.30 (0.83 to 2.04)	1.28 (0.9 to 1.84)	1.21 (0.88 to 1.66)	1.41 (1.07 to 1.87)[a]	1.38 (1.06 to 1.80)[a]	1.28 (1.00 to 1.64)[a]	1.41 (1.07 to 1.85)[a]
Pulmonary embolism	4.05 (1.36 to 12.04)[a]	2.97 (1.40 to 6.29)[a]	2.25 (1.20 to 4.2)[a]	2.06 (1.20 to 3.55)[a]	2.14 (1.33 to 3.43)[a]	2.00 (1.31 to 3.05)[a]	1.77 (1.21 to 2.59)[a]	2.13 (1.39 to 3.25)[a]
Invasive breast cancer	1.43 (0.51 to 4.02)	0.87 (0.54 to 1.41)	1.00 (0.69 to 1.45)	1.60 (0.85 to 1.57)	1.26 (0.97 to 1.64)	1.52 (1.20 to 1.92)[a]	1.60 (1.29 to 1.99)[a]	1.26 (1 to 1.59)
Colorectal cancer	0.79 (0.34 to 1.84)	0.75 (0.38 to 148)	0.62 (0.36 to 1.05)	0.54 (0.33 to 0.86)	0.63 (0.42 to 0.93)[a]	0.66 (0.46 to 0.97)[a]	0.61 (0.43 to 0.86)[a]	0.63 (0.43 to 0.92)[a]
Hip fracture	0.52 (0.19 to 1.40)	0.86 (0.45 to 1.62)	0.72 (0.45 to 1.15)	0.70 (0.46 to 1.06)	0.66 (0.45 to 0.96)[a]	0.70 (0.49 to 1.00)[a]	0.65 (0.46 to 0.90)[a]	0.66 (0.45 to 0.98)[a]

Data are presented as hazard ratio (95% CI).
WHI, Women's Health Initiative.
[a]Log-rank test, $P < 0.05$.

outcome are verified graphically and using Grambsch-Therneau test.[15] Likewise, incidence rates and corresponding CIs (Poisson exact distribution)[16] are determined per outcome and treatment group (placebo and CEE/MPA) at the 5.2-year time limit.

RMST is the mean time for the event to occur; its clinical interpretation depends on the studied event and follow-up time.[7,8] It is determined by calculating the area under the survival curve.[7-9] This area can be estimated as a nonparametric curve (Kaplan-Meier survival curve) using numerical integration[7,14] and/or pseudovalues.[7,17] The area can also be calculated as a parametric curve using the so-called "flexible parametric" (FP) modeling,[7,8] which, unlike other parametric models, is easy to implement and is more flexible, allowing for a better fit—the reasons that it was selected for our analysis. FP modeling should be understood as a regression methodology in which the dependent variable is the survival measure for the studied outcome. This methodology uses transformation of independent variables (restricted cubic splines).[18,19] For our case, we used intervention (treatment group) and iteration between treatment times. Transformation of independent variables generates different FP models. Final selection is based on likelihood (<0.05), Bayesian information criteria (BIC), and Akaike information criteria (AIC), which determine the best adjustment. Degrees of freedom (*df*) and degrees of freedom for each time-dependent effect (DFTVC) indicate the transformation (number of knots) of independent variables.[18,19]

RMST (in days) is determined per treatment group (placebo and CEE/MPA) and outcome at 5.2 years, using estimation methods such as numerical integration,[7,14] pseudovalues,[7,17] and FP modeling.[7,8] Differences in RMST (in days) at 5.2 years between placebo and CEE/MPA and the corresponding CIs (normal distribution) were calculated for each outcome of the WHI study.[7,8]

RESULTS

Before simulation, Kaplan-Meier survival curves for different outcomes were obtained from cumulative hazard plots of the original WHI report.[1] These Kaplan-Meier survival curves showed a small separation area between placebo and treatment curves that, at first glance, seemed to overlap.

In our simulation of the WHI trial report, HRs obtained from Cox proportional hazards analysis at 5.2 years showed values with significant CIs for coronary heart disease, stroke, pulmonary embolism, colorectal cancer, and hip fracture (Fig.). In Table 1, one can observe a wide variation in HRs for various outcomes. Similarly, log-rank test also showed a trend similar to that observed for HRs (Table 1).

Grambsch-Therneau test was used to assess the assumptions of the Cox model. This showed significant values (*P* < 0.05) for coronary heart disease and invasive breast cancer outcomes at 7 years. Significant values were not found for the other outcomes. Graphic examination of cumulative hazards of simulated data for invasive breast cancer curves showed evident cross-linking. A similar cross-linking was observed

TABLE 2. *Rates (per 10,000 person-years) per outcome (reported and simulated) in the Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women (WHI) study[1]*

	Actual reported rates (95% CI) for the WHI		Simulation rates (95% CI) for the WHI	
Outcome	Placebo	CEE/MPA	Placebo	CEE/MPA
Coronary heart disease	30.00 (24.91 to 35.82)	37.00 (31.55 to 43.12)	29.16 (24.22 to 34.82)	37.73 (32.9 to 43.95)
Stroke	21.00 (16.77 to 25.97)	29.00 (24.18 to 34.5)	19.33 (15.35 to 24.03)	27.31 (22.65 to 32.66)
Pulmonary embolism	8.00 (5 to 10.44)	16.00 (12.34 to 20)	5.94 (3.85 to 8.77)	12.71 (9.6 to 16.5)
Invasive breast cancer	30.00 (24.95 to 35.77)	38.00 (32.44 to 44.24)	23.63 (19.21 to 28.77)	29.81 (24.92 to 35.37)
Colorectal cancer	16.00 (12.4 to 20.32)	10.00 (7.29 to 13.38)	14.77 (11.33 to 18.94)	9.29 (6.67 to 12.61)
Hip fracture	15.00 (11.5 to 19.23)	10.00 (7.27 to 13.43)	15.49 (11.96 to 19.74)	10.20 (7.44 to 13.65)

CIs were estimated by Poisson's exact distribution.
WHI, Women's Health Initiative; CEE/MPA, conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg.

for stroke and colorectal cancer; however, it was mild and initial in the survival curves (Fig.). In our simulation model, rates per outcome showed significant differences in pulmonary embolism between the placebo group and the CEE/MPA group (Table 2).

Adjustment measures and criteria used for selecting the FP model for each outcome were as follows: invasive breast cancer (df_2; DFTVC$_2$; AIC, 3,899.97; BIC, 3,946.28), coronary heart disease (df_1; DFTVC$_1$; AIC, 4,385.74; BIC, 4,416.61), stroke (df_1; AIC, 3,242.98; BIC, 3,266.14), pulmonary embolism (df_1; DFTVC$_1$; AIC, 1,624.07; BIC, 1,654.95), colorectal cancer (df_1; DFTVC$_1$; AIC, 1,849.18; BIC, 1,880.05), and hip fracture (df_2; AIC, 1,888.8; BIC, 1,919.67).

RMST showed differences of less than 1 day for each studied outcome when methodologies (numerical integration, pseudovalues, and FP modeling) were compared at 5.2 years (Table 3).

Using estimation methods (numerical integration, pseudovalues, and FP modeling), we found significant differences in coronary heart disease, pulmonary embolism, and colorectal cancer between RMST for placebo and RMST for CEE/MPA ($P < 0.05$). Using the FP estimation method, we found differences in RMST for hip fracture to be equally significant ($P < 0.05$; Table 3). The power for hypothesis testing of the difference of two means (two-tailed; $P < 0.05$) for colorectal cancer and hip fracture was below 54%. Contrary to this, for coronary heart disease and pulmonary embolism, power was higher than 88% (Table 3).

Differences in RMST at 5.2 years for each outcome are presented in Table 3. Using estimation methods (pseudovalues and numerical integration), we found the largest observed absolute difference for coronary heart disease (mean, 8.28 d) and the smallest observed absolute difference for invasive breast cancer (mean, 1.07 d; Table 3).

TABLE 3. *RMST and difference in RMST between placebo and CEE/MPA at 5.2-year follow-up by outcome upon simulation in the Risks and Benefits of Estrogen Plus Progestin in Healthy Postmenopausal Women (Women's Health Initiative) study[1]*

		RMST (d)			
Outcome	Method	Control placebo	Experimental CEE/MPA	Power (%)[a]	Difference in RMST (95% CI) (d)[b]
Coronary heart disease	NI	1,886.08	1,877.80[c]	94.69	8.28 (3.71 to 12.85)
	PV	1,886.08	1,877.80[c]	94.69	8.28 (3.71 to 12.85)
	FP	1,885.61	1,878.10[c]	89.48	7.50 (2.90 to 12.11)
Stroke	NI	1,888.99	1,886.46		2.53 (−1.12 to 6.19)
	PV	1,888.99	1,886.46		2.53 (−1.12 to 6.19)
	FP	1,889.53	1,886.78		2.75 (−0.84 to 6.34)
Pulmonary embolism	NI	1,896.43	1,892.50[c]	88.95	3.93 (1.48 to 6.37)
	PV	1,896.43	1,892.50[c]	88.95	3.93 (1.48 to 6.37)
	FP	1,896.47	1,892.24[c]	93.56	4.23 (1.82 to 6.64)
Invasive breast cancer	NI	1,888.36	1,887.29		1.07 (−2.37 to 4.51)
	PV	1,888.36	1,887.29		1.07 (−2.37 to 4.51)
	FP	1,888.44	1,887.27		1.17 (−2.25 to 4.59)
Colorectal cancer	NI	1,892.19	1,894.82[c]	51.44	−2.63 (−5.20 to −0.06)
	PV	1,892.19	1,894.82[c]	51.44	−2.63 (−5.20 to −0.06)
	FP	1,892.09	1,894.82[c]	53.56	−2.73 (−5.32 to −0.13)
Hip fracture	NI	1,891.71	1,893.84		−2.13 (−4.81 to 0.54)
	PV	1,891.71	1,893.84		−2.13 (−4.81 to 0.54)
	FP	1,891.43	1,894.20[c]	52.51	−2.77 (−5.44 to −0.10)

RMST, restricted mean survival time; CEE/MPA, conjugated equine estrogens 0.625 mg plus medroxyprogesterone acetate 2.5 mg; NI, numerical integration[7,14]; PV, pseudovalues[7,17]; FP, flexible parametric modeling.[7,8]
[a]Power for hypothesis testing of the difference of two means[16] when the null hypothesis is rejected (two-tailed significance set to 0.05). RMSTs and differences in RMST for NI and PV differ in ten thousands, which cannot be shown in the decimals presented in this table.
[b]Difference = control − experimental.
[c]$P < 0.05$, CEE/MPA versus placebo.

DISCUSSION

The original observations of the WHI study[1] were presented as cumulative hazard plots for each clinical outcome. The simulation technique performed in our study used an inverse mechanism that reproduced observations from original plots.[12] Upon simulation, we obtained cumulative hazard plots (Fig.), HR years, and event rates at 5.2 years (Tables 1 and 2), which were similar to those of the original WHI study,[1] therefore indicating the plausibility of inferring our results from those found in the WHI study.[1]

The WHI report[1] and our simulation estimated HRs using Cox proportional hazards analysis.[2] Both studies showed curve cross-linking for invasive breast cancer, indicating that HR was not constant across time. This observation was corroborated by other statistical tests (linear tendency test and Grambsch-Therneau test). When hazards are not maintained across time, it would be misleading to report treatment effects through HRs obtained by Cox proportional hazards analysis.[6,7,20,21] In these cases, using other methodologies is advised.[7,21]

As shown in Table 1, the observed changes in HRs for each outcome on follow-up may reflect the natural variation of HRs and/or the fact that, for each outcome, there are subpopulations that possibly do not share the same risk profile. This will hinder the interpretation of HRs obtained from Cox proportional hazards analysis (under proportional hazards assumptions).

RMSTs at 5.2 years and their corresponding 95% CIs for the various outcomes of the WHI simulation display similar values (differences in thousands) with two different estimation methods used (numerical integration and pseudovalues). This is attributable to the fact that both methods quantify the area under the nonparametric curve.[7,14,17]

The Royston-Parmar methodology (restricted cubic splines)[18,19] generates FP regression models for independent variables (for this case, intervention and iteration intervention with treatment). Likelihood, AIC, and BIC are measures assessing the prediction capability of FP regression for real data, allowing the selection of the best-fitting model. For example, the models selected for hip fracture and stroke used intervention (treatment) as independent variable. Invasive breast cancer, coronary heart disease, pulmonary embolism, and colorectal cancer used intervention and the interaction between intervention and time as independent variables. Therefore, FP models allow for better adjustment because they consider that time may vary with intervention and is a continuous variable. RMST values in the FP model are slightly different (differences of <1 d) from those estimated in the nonparametric curve. This agreement between the different methods used for estimating RMST indicates the robustness of the methodology employed.

In our analysis (FP model) at 5.2-year follow-up, the RMST or mean time of onset of coronary heart disease was 1,878.1 and 1,885.6 days for CEE/MPA and placebo, respectively, with a difference of 7.5 days between groups ($P < 0.05$; Table 3). This means that, at 5.2 years of follow-up, coronary heart event occurs 7.5 days earlier in women using hormone therapy compared with placebo. Our simulation—same as the WHI trial[1]—shows that, for a mean follow-up of 5.2 years, the HR for coronary heart disease is 1.29 (95% CI, 1.02 to 1.63; Table 1). There was statistical significance for RMST and HR. However, one must consider that this fact does not mean causality[22]; in studies with great sample size, statistical significance can be achieved with minimal differences.[10,11] In this sense, RMST offers a measure of time when an event occurs, whereas HRs can be interpreted as the instantaneous relative risk of an event per unit time for an individual with risk factors present compared with an individual with risk factors absent, given that both individuals have survived across time and are similar in all other covariates.[16] For this reason, interpreting HRs to establish clinical differences becomes difficult. Thus, when RMST for placebo is compared with RMST for CEE/MPA at 5.2 years for outcomes such as coronary heart disease, pulmonary embolism, colorectal cancer, and hip fracture, P values lower than 0.05 do not provide a significant clinical difference (Table 3). Likewise, between-group (placebo and CEE/MPA) differences in RMST at 5.2 years for the time of onset of outcomes in the WHI study showed a small difference that does not allow establishment of a reliable clinical prediction of risk.

Difference in RMST is the difference in Kaplan-Meier survival curve areas between intervention groups (placebo and CEE/MPA). In the original report (published cumulative hazard plots were transformed into Kaplan-Meier survival curves), the observation for each outcome of small areas of separation between intervention groups confirms the aforementioned findings. In addition to the difficulty of interpreting HRs, one must add measure and selection biases[23-25] reported for the WHI trial, which may possibly explain the observed results.

CONCLUSIONS

Upon simulation of WHI results, differences in RMST at 5.2 years for the onset of invasive breast cancer, coronary heart disease, stroke, and pulmonary embolism are too small to establish risks related to hormone therapy use.

REFERENCES

1. Rossouw JE, Anderson GL, Prentice RL, et al. 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;288:321-333.
2. Cox D. Regression models and life-tables. *J R Stat Soc* 1972;34:187-220.
3. Rossouw JE, Manson JE, Kaunitz AM, Anderson GL. Lessons learned from the Women's Health Initiative trials of menopausal hormone therapy. *Obstet Gynecol* 2013;121:172-176.
4. Aedo S, Schiattino I, Cavada G, Porcile A. Quality of life in climacteric Chilean women treated with low-dose estrogen. *Maturitas* 2008;61:248-251.
5. Sturmberg JP, Pond DC. Impacts on clinical decision making—changing hormone therapy management after the WHI. *Aust Fam Physician* 2009;38:249-251, 253-255.
6. Spruance SL, Reid JE, Grace M, Samore M. Hazard ratio in clinical trials. *Antimicrob Agents Chemother* 2004;48:2787-2792.
7. Royston P, Parmar MK. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med* 2011;30:2409-2421.

8. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013;13:152.

9. Tian L, Zhao L, Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics* 2014; 15:222-233.

10. Ferrill MJ, Brown DA, Kyle JA. Clinical versus statistical significance: interpreting *P* values and confidence intervals related to measures of association to guide decision making. *J Pharm Pract* 2010;23:344-351.

11. Sarmukaddam SB. Interpreting "statistical hypothesis testing" results in clinical research. *J Ayurveda Integr Med* 2012;3:65-69.

12. Royston P. Tools to simulate realistic censored survival-time distributions. *Stata J* 2012;12:639-654.

13. Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Stat Med* 2013;32:4118-4134.

14. Cleves M, Gould W, Gutierrez RG, Mrachenko YV. *An Introduction to Survival Analysis Using Stata*. 3rd ed. College Station, TX: Stata Press; 2010:412.

15. Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994;81:515-526.

16. Rosner B. *Fundamentals of Biostatistics*. 7th ed. Boston, MA: Cengage Learning; 2010:859.

17. Andersen PK, Perme MP. Pseudo-observations in survival analysis. *Stat Methods Med Res* 2010;19:71-99.

18. Lambert PC, Royston P. Further development of flexible parametric models for survival analysis. *Stata J* 2009;9:265-290.

19. Royston P, Lambert PC. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. College Station, TX: Stata Press; 2011:347.

20. Bellera CA, MacGrogan G, Debled M, de Lara CT, Brouste V, Mathoulin-Pélissier S. Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol* 2010;10:20.

21. Uno H, Claggett B, Tian L, et al. Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. *J Clin Oncol* 2014;32:2380-2385.

22. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health* 2005;95(suppl 1):S144-S150.

23. Shapiro S, Farmer RD, Mueck AO, Seaman H, Stevenson JC. Does hormone replacement therapy cause breast cancer? An application of causal principles to three studies: Part 2. The Women's Health Initiative: estrogen plus progestogen. *J Fam Plann Reprod Health Care* 2011;37: 165-172.

24. Tan O, Harman SM, Naftolin F. What can we learn from design faults in the Women's Health Initiative randomized clinical trial? *Bull NYU Hosp Jt Dis* 2009;67:226-229.

25. Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 2007;297:1465-1477.