

An Analytic Framework for Aligning Observational and Randomized Trial Data: Application to Postmenopausal Hormone Therapy and Coronary Heart Disease

Sengwee Toh · Joann E. Manson

Received: 4 May 2012 / Accepted: 24 September 2012 / Published online: 13 October 2012
© International Chinese Statistical Association 2012

Abstract We describe a conceptual analytic framework for aligning observational and randomized controlled trial (RCT) data. The framework allows one to (1) use observational data to estimate treatment effects comparable to their RCT counterparts, (2) properly include early events that occur soon after treatment initiation in the analysis of observational data, (3) estimate various treatment effects that are of clinical and scientific relevance while appropriately adjusting for time-varying confounders in both the RCT and observational analyses, (4) assess the generalizability of RCT findings in the more diverse populations generally found in the observational data, and (5) combine both types of data to study associations that cannot be addressed by one study or a single data set. We describe the theoretical application of this framework to the Women’s Health Initiative data to examine the relation between postmenopausal hormone therapy and coronary heart disease. The analytic framework can be tailored to specific exposure-outcome associations and data sources, and may be refined as more is learned about its strengths and limitations.

Keywords Postmenopausal hormone therapy · Coronary heart disease · Marginal structural model · Comparative effectiveness research · Pharmacoepidemiology · Pragmatic trials

S. Toh (✉)

Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, 133 Brookline Avenue 6th Floor, Boston, MA 02215, USA
e-mail: darrentoh@post.harvard.edu

J.E. Manson

Division of Preventive Medicine, Brigham and Women’s Hospital, Harvard Medical School, 900 Commonwealth Avenue East 3rd Floor, Boston, MA 02215, USA

1 Introduction

Although randomized controlled trials (RCTs) are considered the gold standard in biomedical research, observational data remain an important, sometimes the only, source to generate valid information on the comparative safety and effectiveness of therapeutics [5, 9, 15, 17, 27, 44, 63, 76]. When observational studies produce results that are not consistent with RCT findings, they are often criticized for their inability to adjust sufficiently for confounding and other biases. Although this is true in some cases, the randomized-observational study discrepancy can arise from reasons that do not necessarily invalidate observational findings, such as differences in study populations and analytic approaches [63]. Specifically, analytic differences are an underappreciated source of disagreement between observational and RCT results, which often lead to “apples and oranges” comparisons as different treatment effects are being estimated in these studies.

In this paper, we use the relation between postmenopausal estrogen-plus-progestin therapy and coronary heart disease (CHD) as a “case study” to describe a conceptual analytic framework for aligning observational and RCT data. We describe how the framework may theoretically be applied to the Women’s Health Initiative data, but note that it can be used for other exposure-outcome associations and in other data sources. The framework allows one to (1) use observational data to estimate treatment effects comparable to their RCT counterparts, (2) appropriately include early events that occur soon after treatment initiation in the analysis of observational data, (3) estimate an array of treatment effects of interest, (4) assess the generalizability of RCT findings, and (5) integrate observational and RCT data to answer research questions that are otherwise not addressable by individual studies due to limited statistical power.

2 Postmenopausal Hormone Therapy and Coronary Heart Disease

2.1 A Tale of Two Study Designs

Initiated in 1993, the Women’s Health Initiative (WHI) comprises the Estrogen-plus-Progestin Trial, the Estrogen-Only Trial, the Diet Modification Trial, the Calcium-plus-Vitamin D Trial (nested within the two hormone trials and the Diet Modification Trial), and a large Observational Study, which includes women who were not willing to participate in or not eligible for the trials [69]. The two parallel hormone trials were in part motivated by previous observational findings that suggested a 40–50 % lower risk for CHD associated with hormone use [4, 18, 65]. Unlike observational studies, however, the Estrogen-plus-Progestin Trial found that women randomized to hormone therapy had a 24 % increased risk for CHD after an average follow-up of 5.6 years [34], whereas the Estrogen-Only Trial found no increase or decrease in risk for CHD with hormone use after a mean follow-up of 6.8 years [25]. In the post-WHI era, the relation between postmenopausal hormone therapy and CHD remains one of the most controversial public health issues and a widely cited example casting doubt on the validity of observational studies.

2.2 Possible Explanations for the Randomized-Observational Study Discrepancy

Potential explanations for the conflicting WHI and observational findings have been discussed extensively [1, 7, 11, 16, 19, 20, 26, 31, 35, 37, 40, 57, 64], and many of them can be applied to the randomized-observational study discrepancies in general.

Confounding bias in observational studies? While women in the hormone and placebo arm of the WHI hormone trials were comparable in their baseline characteristics, hormone users in previous observational studies were likely to be different from non-users in their underlying risk for CHD. Specifically, a “healthy user effect,” which argues that hormone users are generally healthier or more health-conscious than non-users, may at least partly explain a lower CHD risk with hormone therapy in some observational studies [20].

Different treatment regimens? The Estrogen-plus-Progestin Trial, which found an increased CHD risk with hormone therapy, studied one combined hormone regimen, whereas most observational studies in the pre-WHI period examined estrogen-only therapy. The two regimens may have different effects on CHD [33].

The timing hypothesis? The timing hypothesis argues that the hormone therapy-CHD relation may vary by the stage of coronary atherosclerosis: estrogen may reduce the risk for CHD (through, for example, its effects on lipid profile or endothelial function) in younger women who do not yet have advanced atherosclerotic plaque in their coronary arteries, but trigger CHD (through, for example, its effects on thrombosis, inflammation, and plaque rupture) in the presence of advanced lesions in older women [33, 36, 43]. This hypothesis is biologically plausible [36] and is supported by nonhuman primate studies [38] and some data from the WHI hormone trials (especially the Estrogen-only Trial) [25, 32, 34, 58]. Whether existing evidence conclusively proves the timing hypothesis has been debated [3]. Recent studies have indicated that both the timing of treatment initiation and the duration of treatment might be important in determining the benefits and risks of postmenopausal hormone therapy [21, 42, 60, 66, 74].

Were they estimating the same treatment effects? The primary analysis of the WHI hormone trials followed an intention-to-treat (ITT) principle, ignoring changes in treatment status during follow-up and essentially estimating the effect of hormone initiation. In contrast, most pre-WHI observational studies have employed an “as-treated” approach by comparing current hormone users with non-users, effectively estimating the effect of current hormone use. The two effects can be very different, especially when an elevated risk emerges soon after treatment initiation and nonadherence is common. Specifically, in some observational studies, women who experienced CHD following hormone initiation and subsequently stopped the treatment might not have been identified as cases, or might have been systematically misclassified as unexposed cases, leading to an underestimation of the true adverse effect of hormone therapy [20, 49].

3 Prentice’s Work on Resolving the Randomized-Observational Study Discrepancy

Prentice and colleagues have pioneered a series of analyses that combined the trial and observational data in the WHI to address some of the potential sources of the

randomized-observational study discrepancies discussed above [45–47]. The WHI offers a unique opportunity to address these issues as participants were followed contemporaneously, using a similar protocol, regardless of whether they were in the trials or the observational study. Prentice and colleagues were able to achieve better alignment of trial and observational results when they examined the effects of hormone therapy by time from initiation of the current hormone therapy, and adjusted for an extensive list of potential confounders in the observational data. They suggested that the tendency for RCTs to have predominantly short-term follow-up (characterized by increased CHD risk from hormone therapy) and observational studies to have predominantly long-term follow-up (characterized by neutral or reduced CHD risk) might explain a large proportion of the discrepancy between the two study designs. However, their observational analyses started follow-up at the time participants entered the study, not at the time they initiated hormone therapy. As a result, CHD events that occurred from the time of treatment initiation to the time of study entry were not appropriately included in the analyses.

4 An Analytic Framework for Aligning Observational and Randomized Data

Building in part upon work pioneered by Prentice and colleagues [45–47], Hernán and colleagues [21, 24, 74] and Tannen and colleagues [67], we develop a conceptual analytic framework that can be used to align observational and randomized data. The analytic framework requires that one be able to conceptualize a hypothetical trial using observational data (Fig. 1). It tailors the design of observational studies to emulate their RCT counterparts at the design phase, and maps their analysis to the ITT approach used in RCTs at the analysis phase. The goal is to design a cohort study identical to an actual or hypothetical RCT, except that the assignment of treatment status is neither random nor blinded. We refer to such a cohort study as a “simulated trial.”

An inception cohort design and a restriction approach form the backbone of the design phase of the framework. The inception cohort design identifies treatment initiators following a “wash-out” period (with length defined by investigators), allowing one to follow individuals at the time of treatment initiation to identify early events and assure that confounders can be measured prior to treatment initiation [49]. The restriction approach identifies the subset of the inception cohort who meets the eligibility criteria of the actual or hypothetical RCT. Restriction reduces confounding by removing individuals who are ineligible for certain treatment due to contraindications or unmeasured risk factors [48, 61, 76]. The analysis phase is guided by the ITT approach, the primary analysis used in RCTs that preserves baseline comparability of participants in different treatment groups.

The analytic framework allows one to assess the generalizability of RCT findings in the more diverse populations generally found in the observational data by gradually relaxing the eligibility criteria or by using the approach developed by Cole and Stuart (Sect. 5.4) [8]. It also enables one to use both randomized and observational data to estimate an array of treatment effects that are of clinical and scientific relevance.

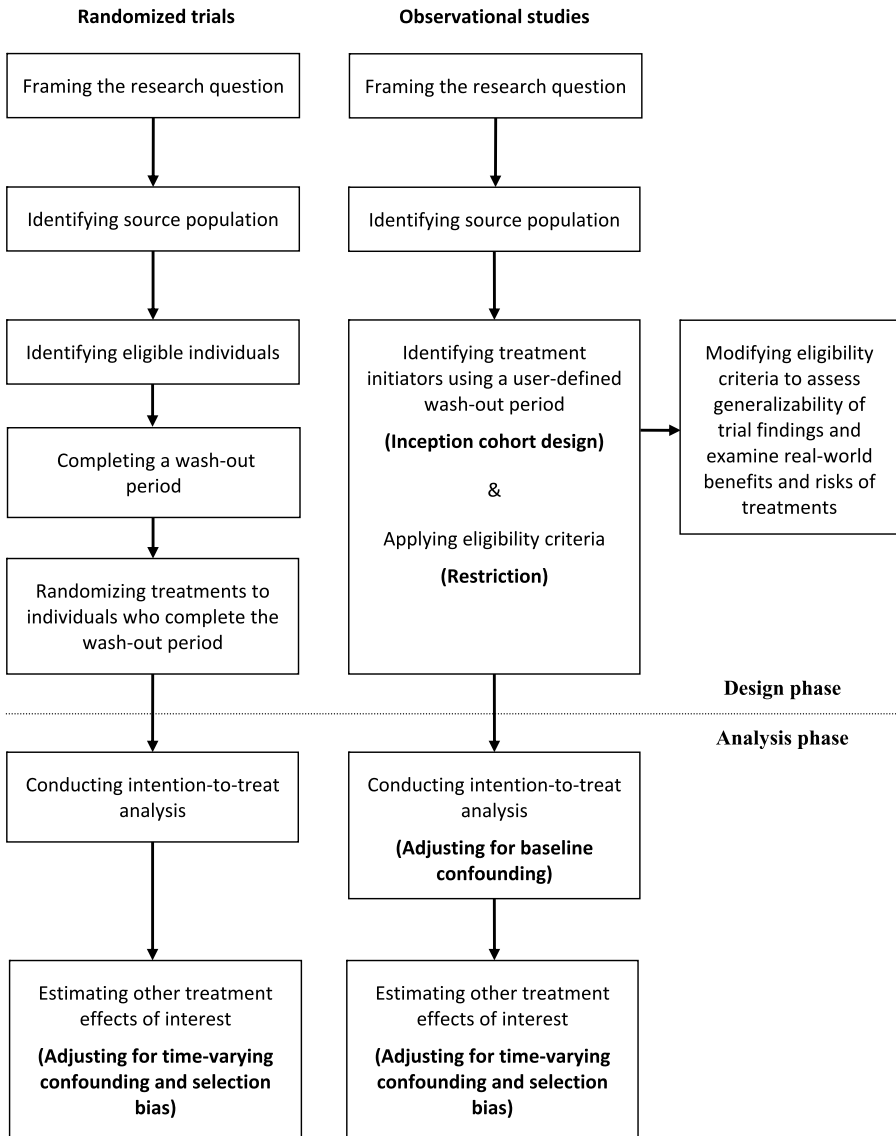


Fig. 1 A conceptual analytic framework for aligning randomized and observational data

We have previously applied part of the analytic framework to the Nurses' Health Study—a large prospective observational study—to obtain ITT results that are consistent with the results from the WHI Estrogen-plus-Progestin Trial [21] (Table 1). Since the ITT estimates might underestimate the true effects of hormone therapy in the presence of treatment nonadherence and since these effect might not be directly comparable across studies in the presence of differential nonadherence, we further used the framework to estimate the adherence-adjusted effects in both the Nurses'

Table 1 Intention-to-treat effect estimates of postmenopausal hormone therapy on coronary heart disease from the Women’s Health Initiative (WHI) Estrogen-plus-Progestin Trial and the observational Nurses’ Health Study (NHS)

		Hazard ratio by follow-up period					
Overall		≤2 years		>2 years			
	WHI ^a	NHS ^a	WHI ^a	NHS ^a	WHI ^a	NHS ^a	
All women	1.23 (0.99–1.52)	0.99 (0.80–1.22)	1.55 (1.09–2.20)	1.43 (0.92–2.23)	1.07 (0.81–1.40)	0.91 (0.72–1.16)	
<i>p</i> -value for homogeneity ^b	0.15		0.78		0.38		
Years since menopause							
<10	0.89 (0.55–1.46)	0.88 (0.63–1.21)	1.17 (0.54–2.52)	1.28 (0.62–2.64)	0.74 (0.39–1.40)	0.81 (0.56–1.17)	
<i>p</i> -value for homogeneity ^b	0.97		0.87		0.81		
≥10	1.44 (1.10–1.87)	1.13 (0.85–1.49)	1.74 (1.15–2.63)	1.50 (0.84–2.68)	1.27 (0.90–1.79)	1.06 (0.77–1.44)	
<i>p</i> -value for homogeneity ^b	0.21		0.68		0.44		
Age at baseline (years)							
50–59	1.30 (0.79–2.15)	0.89 (0.67–1.19)	1.60 (0.73–3.55)	1.36 (0.71–2.57)	1.14 (0.60–2.16)	0.82 (0.59–1.14)	
<i>p</i> -value for homogeneity ^b	0.20		0.76		0.37		
≥60	1.21 (0.95–1.54)	1.15 (0.85–1.57)	1.54 (1.04–2.29)	1.49 (0.79–2.80)	1.05 (0.77–1.42)	1.08 (0.77–1.53)	
<i>p</i> -value for homogeneity ^b	0.80		0.93		0.91		

^aThe WHI results are obtained from Toh et al. [74] paper, the NHS results are obtained from Hernán et al. [21] paper

^bWald test for homogeneity of the hazard ratios from two studies

Table 2 Adherence-adjusted effect estimates of postmenopausal hormone therapy on coronary heart disease from the Women’s Health Initiative (WHI) Estrogen-plus-Progestin Trial and the observational Nurses’ Health Study (NHS) (from [74])

		Hazard ratio by follow-up period					
Overall		≤2 years		>2 years			
WHI ^a	NHS	WHI ^b	NHS	WHI ^c	NHS	WHI ^c	NHS
All women	1.69 (0.98–2.89)	1.30 (0.76–2.21)	1.71 (1.03–2.83)	2.36 (1.55–3.62)	1.71 (1.03–2.83)	1.33 (0.88–2.01)	1.07 (0.44–2.63)
<i>p</i> -value for homogeneity ^d	0.50	0.34		0.34		0.67	
Years since menopause							
<10	0.64 (0.21–1.99)	0.68 (0.24–1.91)	1.28 (0.43–3.86)	1.29 (0.52–3.18)	1.28 (0.43–3.86)	0.63 (0.27–1.52)	0.20 (0.03–1.54)
<i>p</i> -value for homogeneity ^d	0.94			0.99		0.31	
≥10	2.22 (1.20–4.11)	1.57 (0.86–2.85)	1.97 (1.11–3.47)	2.82 (1.73–4.60)	1.97 (1.11–3.47)	1.64 (1.01–2.65)	1.37 (0.54–3.45)
<i>p</i> -value for homogeneity ^d	0.43			0.35		0.74	
Age at baseline (years)							
50–59	1.47 (0.57–3.77)	0.91 (0.49–1.69)	1.80 (0.83–3.87)	2.69 (1.14–6.36)	1.80 (0.83–3.87)	1.22 (0.59–2.56)	0.54 (0.20–1.49)
<i>p</i> -value for homogeneity ^d	0.40			0.49		0.20	
≥60	1.76 (0.97–3.19)	1.92 (0.90–4.10)	1.69 (0.87–3.32)	2.30 (1.43–3.72)	1.69 (0.87–3.32)	1.36 (0.86–2.15)	2.10 (0.68–6.50)
<i>p</i> -value for homogeneity ^d	0.86			0.47		0.49	

^aEight-year cumulative use; the WHI Estrogen-plus-Progestin Trial began in 1993 and ended in 2001, which was approximately 8 years

^bTwo-year cumulative use

^cSix-year cumulative use

^dWald test for homogeneity of the hazard ratios from two studies

Health Study [21] and the WHI Estrogen-plus-Progestin Trial [74] (Table 2). In addition to highlighting the importance and usefulness of the analytic framework, these analyses provide insight into the relation of hormone therapy and CHD by age, year since menopause, treatment duration.

5 Theoretical Application of the Analytic Framework Using the WHI Data

In this section, we describe how the analytic framework may theoretically be applied to the WHI data. We simplify the discussion by assuming that non-methodological issues, including data availability and reliability, are minor or can be addressed adequately. Where applicable, however, we highlight certain data issues based on our knowledge of the data, and their implications on the feasibility of applying the framework in practice.

5.1 Emulating the WHI Estrogen-Plus-Progestin Trial Using the Observational Study Data

The design and analysis of the WHI Estrogen-plus-Progestin Trials (“the Trial”) have been described in detail elsewhere [34, 69, 70]. Briefly, in this double-blinded trial, postmenopausal women aged 50–79 years with an intact uterus and without certain exclusion conditions (described below) at baseline were randomly assigned to receive oral conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d or placebo. They were followed for occurrence of a number of outcomes, such as CHD, cancer, fracture, and mortality. As with other RCTs, the primary analysis was guided by an ITT principle.

Using this analytic framework, we first identify postmenopausal women aged 50–79 years with an intact uterus at the baseline visit from the WHI Observational Study cohort. Among these women, we identify those who reported either use of the same hormone regimen as in the Trial or no use of hormone therapy at baseline. To mimic the 3-month wash-out period used in the Trial, these women must also report having no use of any hormone therapy in the last 3 months. We further restrict the cohort according to the eligibility criteria of the Trial, including no myocardial infarction, stroke, transient ischemic attack in the previous 6 months, and no breast cancer or other cancers (except non-melanoma skin cancer) within the past 10 years.

The remaining women form the study cohort of the simulated trial. They are followed from the baseline visit to the earliest occurrence of CHD, death, loss to follow-up, or end of follow-up (July 7, 2002, the day the Trial was terminated). We compare the risk for CHD between hormone initiators and non-initiators, regardless of whether these women subsequently stopped or initiated hormone therapy. Specifically, we estimate the average hazard ratio of CHD in hormone initiators versus non-initiators, and its 95 % confidence interval (CI), by fitting a Cox proportional hazards model that includes a non-time-varying indicator for hormone initiation.

To obtain valid effect estimates in the simulated trial, however, we need to adjust for baseline confounders, which include sociodemographic, lifestyle, dietary, and medical factors [21, 24, 45–47, 74]. There are several ways to incorporate baseline

confounders in the analysis, inducing matching, stratification, modeling, or weighting [28, 72]. A common approach is to adjust for them in the outcome regression model, either as individual covariates or as confounder summary scores (e.g., propensity scores [29, 56] or disease risk scores [2]).

5.2 The Sequential Simulated Trial Design

The approach described above produces imprecise effect estimates if there are few eligible hormone initiators at the baseline visit. However, we can produce an additional simulated trial if we apply the framework again to subsequent follow-up contacts in the WHI Observational Study (Fig. 2). This sequential simulated trial design has been shown to improve statistical efficiency [12, 21]. We construct additional simulated trials at each subsequent follow-up contact. In each trial, we use the updated information to apply the eligibility criteria and identify hormone initiators and non-initiators. We then pool all trials in a single analysis and use the robust variance estimator [30] to account for within-person correlation because some women may participate in multiple trials. We assess the potential heterogeneity of the ITT estimates across trials by estimating a separate parameter in each trial and testing for heterogeneity of the parameters, or by creating a product term between trial and hormone therapy indicator and testing for the product term being different from zero [21].

5.3 Differences Between the Simulated Trial and the Actual Randomized Trial

There are a number of differences between the simulated trial and the actual Trial. First, unlike in the actual Trial, the distributions of baseline risk factors for CHD are likely to be different between hormone initiators and non-initiators in the simulated trial. Therefore, additional adjustment for potential confounders is necessary. The validity of the simulated trial results depend heavily on the availability of and appropriate adjustment for all the joint determinants of hormone therapy and CHD.

Second, treatment assignment in the simulated trial is not blinded, i.e., patients and clinicians know what patients receive. Bias may arise if outcome diagnosis varies

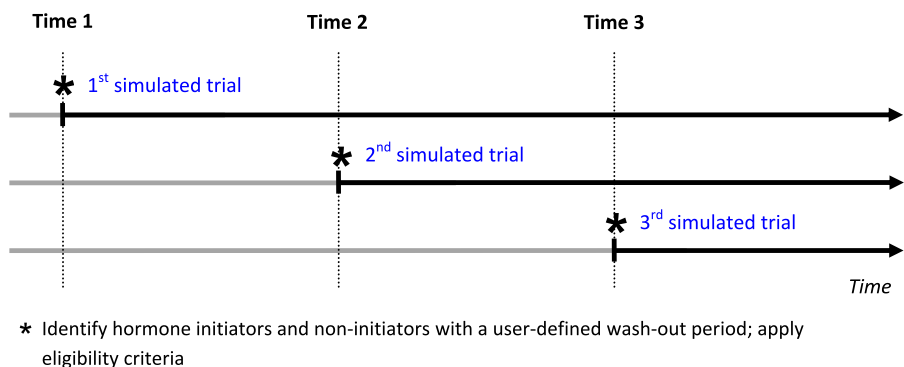


Fig. 2 A sequential simulated trial design

by treatment status, but this is not likely to be a major issue in the WHI Observational Study because the follow-up protocol was developed carefully to ensure that the identification, reporting, and validation of the outcome are independent of hormone use status [69]. However, the awareness of treatment status may lead to behavioral changes that may also impact the outcome risk. As a result, the ITT effect observed in the simulated trial is likely not solely from the treatment itself, but also that from the associated behavioral changes [12]. This is less of a concern if the goal is to emulate an open-label trial. (We note that even though the actual Trial was designed to be double-blinded, blinding might not be complete. For example, women with menopausal symptoms in the placebo arm may assume that they were receiving placebo.)

Third, it is not possible to identify placebo initiators in the simulated trial. To further mimic the Trial, we may use initiators of another drug not thought to be associated with CHD as the comparison group. Initiators of glaucoma drugs, which have been previously used to adjust for biases arising from healthy-user effects [62], may be a potential comparison group, but others may also be considered. The comparison group should be similar to the hormone group in their baseline characteristics. This can be achieved in part by applying the same eligibility for both the hormone and the comparison group. Identifying an appropriate comparison group is usually not an issue if the goal is to mimic RCTs with active comparators. Fourth, in the Trial a pre-randomization, placebo-only wash-out period was used to identify individuals who were likely to adhere to their assigned treatment during the Trial. Therefore, we might further require participants to also report using a non-study drug in the previous 3 months. However, the granularity of drug use information in the WHI Observational Study might be too coarse to be used to deal with the last two issues.

5.4 Assessing the Generalizability of RCT Findings

We can use the analytic framework to assess the generalizability of the RCT results. For example, if the interest is in a specific subgroup of individuals excluded from the Trial (e.g., individuals with breast cancer in the past 10 years), the simulated trial can be constructed as described above, except that these individuals are no longer excluded from the analysis. The eligibility criteria can be further modified to include other individuals who are excluded from the RCTs. We can study the average treatment effects in the overall population, or within strata of baseline characteristics to examine treatment heterogeneity.

Alternatively, we can use the approach proposed by Cole and Stuart to deal with multiple exclusion criteria simultaneously [8]. The approach models the conditional probability of being selected from the target population into the RCT, then uses inverse probability weighting (described in greater detail in the next section) to standardize RCT results to the target population under the assumptions that determinants of selection that reflect treatment heterogeneity be measured and modeled correctly. If the framework is applied to a comparative effectiveness question, it may also provide insight into the “efficacy–effectiveness gap” often observed in studies of intended effects of therapeutics [14, 41].

5.5 Estimating Other Treatment Effects of Interest

In addition to ITT effect, other treatment effects of interest can also be estimated using both the observational and the RCT data. These treatment effects can be estimated by a number of methods that appropriately adjust for time-dependent confounders that are also affected by prior treatments, including inverse probability weighting of marginal structural models [22, 54, 72], g-estimation of structural nested models [51–53], or the parametric g-formula [50, 55, 68, 78]. This section describes how to use inverse probability weighting to estimate two treatment effects that are both scientifically and clinically relevant.

5.5.1 Treatment Effect Under Full Treatment Adherence

We use inverse probability weighting to estimate the effect if all women had adhered to their initial assigned treatment throughout the follow-up: this effect is sometimes referred to as the effect of continuous treatment. As RCTs are also vulnerable to time-dependent confounding and selection biases that arise from differential loss to follow-up, inverse probability weighting is used in both the simulated trial and the actual Trial. (Note: Since loss to follow-up was minimal in the Estrogen-plus-Progestin Trial and the Observational Study during the study period, we do not use inverse probability weighting to adjust for selection bias due to study dropout. Readers are referred to [22, 54, 72] for more information.)

Informally, the inverse probability weighting approach weighs each woman at each follow-up time by the inverse of the conditional probability (or more generally, density) of having her observed treatment history through that time. A valid weight is required to provide an unbiased estimate of the treatment effect. To obtain a valid weight, we need to include in the weight estimation models all the joint determinants of treatment and outcome. This method produces valid estimates provided that treatment status at each follow-up time is unrelated to unmeasured risk factors for the outcome conditional on the measured covariates. The weight will be invalid, if, for example, LDL (which is not available for all WHI participants at all follow-up timepoints) is still predictive of hormone use after adjusting for all measured factors such as body mass index and use of lipid-lowering medications.

We describe hormone use as an annualized proportion with a person-year data structure (i.e., each observation is a person-year contributed by eligible participants). In the Trial, this is computed from the proportion of study pills taken obtained from weighing of returned bottles and the self-reported treatment duration of non-study hormone use. In the simulated trial conducted within the WHI Observational Study, information about the proportion of pills taken is not available, so we estimate the annualized proportion based on the self-reported treatment duration in a given year. Many women, especially those in the placebo arm of the Trial, reported no hormone use, resulting in a skewed distribution. Thus, we use a “two-part model” [13] to estimate the inverse probability (density) weights by fitting, separately for each arm, (1) a logistic regression model to estimate each woman’s probability of receiving any hormone therapy during each follow-up year, and (2) a linear regression model to estimate each woman’s density of receiving her actual proportion of pills taken (some

transformations, e.g., arcsin-root transformation, may be required) among those with non-zero use during that year [10, 54, 73, 74]. Both models include years since initiation, proportion of study pills taken in the previous year, as well as the potential confounders measured at baseline and, for time-varying covariates, at the most recent visit. A list of baseline and time-varying confounders can be found in a previous study [74].

The weight for each woman at each year is calculated as the inverse of the probability (or density) of having received her actual treatment history through that time. To improve statistical efficiency, we stabilize the weights [22, 54, 72] by adding to their numerator the estimated density of received treatment history conditional on the proportion of study pills taken in the previous year and selected baseline covariates included in the model for the denominator of the weights. A woman contributes as many observations to the models as person-years she was in the study, i.e., from baseline to the earliest occurrence of CHD, death, loss to follow-up, or end of study period.

To estimate the effect under continuous hormone therapy, we need to assume a “dose-response” outcome model. This is required in this version of the inverse probability weighting approach that does not censor participants when they become non-adherent [12]. An alternative approach censors participants when they deviate from their initial treatment and uses inverse probability weighting to adjust for potential selection bias that arises from such censoring. That approach does not require a dose-response outcome model but generally has a smaller sample size in the final outcome analysis [12].

The dose-response outcome model should be specified based on subject-matter knowledge whenever possible. If we assume a cumulative effect of hormone therapy on the risk for CHD, we can fit a weighted pooled logistic regression model [71] (to approximate the Cox model) that includes a time-varying variable for cumulative hormone therapy, calculated as the sum of the proportion of pills taken since baseline, and the baseline variables used to estimate the numerator of the weights. Other dose-response models can be specified. Fitting different dose-response models allow us to examine the robustness of study findings. To estimate the average hazard ratios, we use the parameter estimates from the model to simulate a Monte Carlo sample of, say, 100,000 women, and use a non-parametric bootstrap estimator [77] to calculate the 95 % CIs for the average hazard ratios.

In applying the analytic framework in this theoretical exercise, the use of a person-year data structure and a dose-response outcome model is not by choice but rather by necessity. The information available in the WHI does not allow us to establish with confidence the temporal sequence of treatment nonadherence and CHD in a given year. Therefore, we are not able to censor participants at the exact time they became nonadherent.

5.5.2 *Effects of Dynamic Treatment Regimens*

Sometimes the effect under continuous treatment may not be clinically meaningful because patients may have to stop the treatment due to, for example, severe side effects. Therefore, the effects of treatment regimens that evolve with patients' changing prognosis and indications for treatment may be of greater interest. In the WHI

Estrogen-plus-Progestin Trial, participants in the hormone arm who developed an adverse outcome (e.g., endometrial hyperplasia with atypia) were required to permanently stop their study pills. We can estimate the effect of hormone therapy that would have been observed had all women fully adhered to this protocol using inverse probability weighting [6, 23, 73, 75]. More specifically, we can estimate the effects under the dynamic regimen “take hormone therapy until an adverse event occurs, then stop taking hormone therapy.” To do so, we artificially censor participants in the hormone arm at the time they deviated from the protocol (i.e., did not stop taking their study pills after they had an adverse event).

Such artificial censorings may result in selection bias because the distribution of risk factors of CHD may differ between the censored and the uncensored women. To adjust for this potential selection bias, we would estimate time-varying, subject-specific inverse probability weights whose denominator is the women’s estimated conditional probability of remaining uncensored at each time. However, the predictors of censoring at time t are in fact the predictors of hormone therapy continuation at t because those who continue taking 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 because the treatment weights estimated above already adjust for selection bias due to such artificial censoring. Thus, to estimate effect of this dynamic hormone treatment regimen, we fit a weighted pooled logistic regression model only for women who remained uncensored.

Inverse probability weighting, g-estimation of structural nested models, and other methods can be used to estimate the effects of other dynamic treatment regimens. Readers are referred to [23, 39, 51–53, 75] for additional information.

5.6 Performing Stratified or Subgroup Analyses

It is common to perform stratified or subgroup analyses in comparative safety and effectiveness research. For example, we may be interested in estimating the effects of postmenopausal hormone therapy by the timing of treatment initiation and the duration of treatment as recent studies have suggested that both factors might be important in determining its risk and benefit profile [21, 42, 60, 66, 74]. It is straightforward to conduct stratified or subgroup analyses by baseline characteristics under the analytic framework. For example, we can stratify the analysis according to the recency of menopause at the time of treatment initiation (e.g., <5 vs. ≥ 5 years since menopause) or age (e.g., 50–59 vs. ≥ 60 years) to estimate the treatment effects by timing of use. The effects of hormone therapy by duration of (continuous) use are in fact time-varying treatment effects, even though the treatment status remains unchanged over that duration. Therefore, we can use the method described in Sect. 5.5.1 to estimate the effects of hormone therapy by treatment duration (e.g., first 2 years of continuous use) [21, 74].

5.7 Combining Observational and RCT Data

We examine if there is heterogeneity in effect estimates from the simulated trial and the actual Trial. This can be done by Wald test for homogeneity [59]. If there is little

evidence of heterogeneity of a specific treatment effect (e.g., the ITT effect among women aged 50–59 years in the first 2 years following treatment initiation), the log hazard ratios from the studies can be weighted by the inverse of their variances to obtain pooled estimates [59]. Other pooling approaches can also be considered.

6 Conclusion

We have described a conceptual analytic framework for aligning randomized and observational data. Under this framework, we can use observational and RCT data to estimate comparable treatment effects, assess the generalizability of RCT findings, and combine both types of data to study associations that require larger sample size. The proposed framework may be used to answer some of the unresolved questions about the hormone-CHD relation, e.g., whether the timing hypothesis is supported by existing data. It can be tailored to specific exposure-outcome associations, and may be refined as more is learned about its strengths and limitations.

Acknowledgements Dr. Toh is partially supported by the Agency for Healthcare and Research Quality (R03HS019024). Dr. Manson is partially supported by the National Heart, Lung, and Blood Institute (R01HL034594, R01HL088521, and NOI-WH32109).

References

1. Allison MA, Manson JE (2006) Observational studies and clinical trials of menopausal hormone therapy: can they both be right? *Menopause* 13(1):1–3
2. Arbogast PG, Ray WA (2009) Use of disease risk scores in pharmacoepidemiologic studies. *Stat Methods Med Res* 18(1):67–80
3. Barrett-Connor E (2007) Hormones and heart disease in women: the timing hypothesis. *Am J Epidemiol* 166(5):506–510
4. Barrett-Connor E, Grady D (1998) Hormone replacement therapy, heart disease, and other considerations. *Annu Rev Public Health* 19:55–72
5. Benson K, Hartz AJ (2000) A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 342(25):1878–1886
6. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA (2010) When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat* 6(2):18
7. Col NF, Pauker SG (2003) The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? *Ann Intern Med* 139(11):923–929
8. Cole SR, Stuart EA (2010) Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol* 172(1):107–115
9. Concato J, Shah N, Horwitz RI (2000) Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 342(25):1887–1892
10. Cotter D, Zhang Y, Thamer M, Kaufman J, Hernán MA (2008) The effect of epoetin dose on hematocrit. *Kidney Int* 73(3):347–353
11. Creasman WT, Hoel D, Disaia PJ (2003) WHI: Now that the dust has settled: a commentary. *Am J Obstet Gynecol* 189(3):621–626
12. Danaei G, Garcia Rodriguez LA, Canero OF, Logan R, Hernan MA (2011) Observational data for comparative effectiveness research: an emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res*. doi:10.1177/0962280211403603
13. Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY (1999) Methods for analyzing health care utilization and costs. *Annu Rev Public Health* 20:125–144

14. Eichler HG, Abadie E, Breckenridge A, Flamion B, Gustafsson LL, Leufkens H, Rowland M, Schneider CK, Bloechl-Daum B (2011) Bridging the efficacy-effectiveness gap: a regulator's perspective on addressing variability of drug response. *Nat Rev Drug Discov* 10(7):495–506
15. Federal Coordinating Council for Comparative Effectiveness Research (2009) Report to the President and Congress on comparative effectiveness research. Department of Health and Human Services
16. Garbe E, Suissa S (2004) Hormone replacement therapy and acute coronary outcomes: methodological issues between randomized and observational studies. *Hum Reprod* 19(1):8–13
17. Golder S, Loke YK, Bland M (2011) Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. *PLoS Med* 8(5):e1001026
18. Grady D, Rubin SM, Petitti DB, Fox CS, Black D, Ettinger B, Ernster VL, Cummings SR (1992) Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 117(12):1016–1037
19. Grimes DA, Lobo RA (2002) Perspectives on the Women's Health Initiative trial of hormone replacement therapy. *Obstet Gynecol* 100(6):1344–1353
20. Grodstein F, Clarkson TB, Manson JE (2003) Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med* 348(7):645–650
21. Hernán MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, Manson JE, Robins JM (2008) Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 19(6):766–779
22. Hernán MA, Brumback B, Robins JM (2000) Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 11(5):561–570
23. Hernán MA, Lanoy E, Costagliola D, Robins JM (2006) Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol* 98(3):237–242
24. Hernán MA, Robins JM, García Rodríguez LA (2005) Discussion of Statistical issues arising in the Women's Health Initiative by Prentice RL, Pettinger M, Andreson GL. *Biometrics* 61(4):922–930
25. Hsia J, Langer RD, Manson JE, Kuller L, Johnson KC, Hendrix SL, Pettinger M, Heckbert SR, Greep N, Crawford S, Eaton CB, Kostis JB, Caralis P, Prentice R (2006) Conjugated equine estrogens and coronary heart disease: the Women's Health Initiative. *Arch Intern Med* 166(3):357–365
26. Humphrey LL, Chan BK, Sox HC (2002) Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med* 137(4):273–284
27. Institute of Medicine (2009) Initial national priorities for comparative effectiveness research. The National Academies Press, Washington
28. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, Robins JM (2006) Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 163(3):262–270
29. Kuss O, Legler T, Borgermann J (2011) Treatments effects from randomized trials and propensity score analyses were similar in similar populations in an example from cardiac surgery. *J Clin Epidemiol* 64(10):1076–1084
30. Liang KY, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73(1):13–22
31. Machens K, Schmidt-Gollwitzer K (2003) Issues to debate on the Women's Health Initiative (WHI) study. Hormone replacement therapy: an epidemiological dilemma? *Hum Reprod* 18(10):1992–1999
32. Manson JE, Allison MA, Rossouw JE, Carr JJ, Langer RD, Hsia J, Kuller LH, Cochrane BB, Hunt JR, Ludlam SE, Pettinger MB, Gass M, Margolis KL, Nathan L, Ockene JK, Prentice RL, Robbins J, Stefanick ML (2007) Estrogen therapy and coronary-artery calcification. *N Engl J Med* 356(25):2591–2602
33. Manson JE, Bassuk SS (2007) Invited commentary: hormone therapy and risk of coronary heart disease why renew the focus on the early years of menopause? *Am J Epidemiol* 166(5):511–517
34. Manson JE, Hsia J, Johnson KC, Rossouw JE, Assaf AR, Lasser NL, Trevisan M, Black HR, Heckbert SR, Detrano R, Strickland OL, Wong ND, Crouse JR, Stein E, Cushman M (2003) Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med* 349(6):523–534
35. McDonough PG (2002) The randomized world is not without its imperfections: reflections on the Women's Health Initiative Study. *Fertil Steril* 78(5):951–956
36. Mendelsohn ME, Karas RH (2007) HRT and the young at heart. *N Engl J Med* 356(25):2639–2641
37. Michels KB, Manson JE (2003) Postmenopausal hormone therapy: a reversal of fortune. *Circulation* 107(14):1830–1833
38. Mikkola TS, Clarkson TB (2002) Estrogen replacement therapy, atherosclerosis, and vascular function. *Cardiovasc Res* 53(3):605–619

39. Murphy SA (2003) Optimal dynamic treatment regimes. *J R Stat Soc B* 65(2):331–355
40. Naftolin F, Taylor HS, Karas R, Brinton E, Newman I, Clarkson TB, Mendelsohn M, Lobo RA, Judelson DR, Nachtigall LE, Heward CB, Hecht H, Jaff MR, Harman SM (2004) The Women's Health Initiative could not have detected cardioprotective effects of starting hormone therapy during the menopausal transition. *Fertil Steril* 81(6):1498–1501
41. Nallamothu BK, Hayward RA, Bates ER (2008) Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 118(12):1294–1303
42. North American Menopause Society (2012) The 2012 hormone therapy position statement of the North American Menopause Society. *Menopause* 19(3):257–271
43. Phillips LS, Langer RD (2005) Postmenopausal hormone therapy: critical reappraisal and a unified hypothesis. *Fertil Steril* 83(3):558–566
44. Platt R, Carnahan RM, Brown JS, Chrischilles E, Curtis LH, Hennessy S, Nelson JC, Racoosin JA, Robb M, Schneeweiss S, Toh S, Weiner MG (2012) The U.S. food and drug administration's mini-sentinel program: status and direction. *Pharmacoepidemiol Drug Saf* 21(Suppl 1):1–8
45. Prentice RL, Langer R, Stefanick ML, Howard BV, Pettinger M, Anderson G, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J (2005) 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* 162(5):404–414
46. Prentice RL, Langer RD, Stefanick ML, Howard BV, Pettinger M, Anderson GL, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J (2006) Combined analysis of Women's Health Initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *Am J Epidemiol* 163(7):589–599
47. Prentice RL, Manson JE, Langer RD, Anderson GL, Pettinger M, Jackson RD, Johnson KC, Kuller LH, Lane DS, Wactawski-Wende J, Brzyski R, Allison M, Ockene J, Sarto G, Rossouw JE (2009) Benefits and risks of postmenopausal hormone therapy when it is initiated soon after menopause. *Am J Epidemiol* 170(1):12–23
48. Psaty BM, Siscovick DS (2010) Minimizing bias due to confounding by indication in comparative effectiveness research: the importance of restriction. *JAMA* 304(8):897–898
49. Ray WA (2003) Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 158(9):915–920
50. Robins JM (1986) A new approach to causal inference in mortality studies with sustained exposure periods—application to control of the health worker survivor effect. *Math Model* 7:1393–1512
51. Robins JM (1989) 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 service research methodology: a focus on AIDS*. NCHSR, US Public Health Service, Washington, pp 113–159
52. Robins JM (1993) Analytic methods for estimating HIV treatment and cofactor effects. In: Ostrow DG, Kessler R (eds) *Methodological issues of AIDS mental health research*. Plenum, New York, pp 213–290
53. Robins JM, Hernán MA (2008) Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds) *Longitudinal data analysis*. Chapman and Hall/CRC Press, New York, pp 553–599
54. Robins JM, Hernán MA, Brumback B (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology* 11(5):550–560
55. Robins JM, Hernán MA, Siebert U (2004) Effects of multiple interventions. In: Ezzati M, Lopez AD, Murray CJL (eds) *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*. World Health Organization, Geneva
56. Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70:41–55
57. Rossouw JE (2005) Coronary heart disease in menopausal women: implications of primary and secondary prevention trials of hormones. *Maturitas* 51(1):51–63
58. Rossouw JE, Prentice RL, Manson JE, Wu L, Barad D, Barnabei VM, Ko M, LaCroix AZ, Margolis KL, Stefanick ML (2007) Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA* 297(13):1465–1477
59. Rothman KJ, Greenland S, Lash TL (eds) (2008) *Modern epidemiology*, 3rd edn. Lippincott Williams & Wilkins, Philadelphia
60. Santen RJ, Allred DC, Ardoin SP, Archer DF, Boyd N, Braunstein GD, Burger HG, Colditz GA, Davis SR, Gambacciani M, Gower BA, Henderson VW, Jarjour WN, Karas RH, Kleerekoper M, Lobo RA,

- Manson JE, Marsden J, Martin KA, Martin L, Pinkerton JV, Rubinow DR, Teede H, Thiboutot DM, Utian WH (2010) Postmenopausal hormone therapy: an endocrine society scientific statement. *J Clin Endocrinol Metab* 95:s1–s66
61. Schneeweiss S, Patrick AR, Sturmer T, Brookhart MA, Avorn J, Maclure M, Rothman KJ, Glynn RJ (2007) Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Med Care* 45(10 Supl 2):S131–142
 62. Solomon DH, Avorn J, Sturmer T, Glynn RJ, Mogun H, Schneeweiss S (2006) Cardiovascular outcomes in new users of coxibs and nonsteroidal antiinflammatory drugs: high-risk subgroups and time course of risk. *Arthritis Rheumatol* 54(5):1378–1389
 63. Sorensen HT, Lash TL, Rothman KJ (2006) Beyond randomized controlled trials: a critical comparison of trials with nonrandomized studies. *Hepatology* 44(5):1075–1082
 64. Stampfer M (2004) Commentary: Hormones and heart disease: do trials and observational studies address different questions? *Int J Epidemiol* 33(3):454–455
 65. Stampfer MJ, Colditz GA (1991) Estrogen replacement therapy and coronary heart disease: a quantitative assessment of the epidemiologic evidence. *Prev Med* 20(1):47–63
 66. Sturdee DW, Pines A, Archer DF, Baber RJ, Barlow D, Birkhauser MH, Brincat M, Cardozo L, de Villiers TJ, Gambacciani M, Gompel AA, Henderson VW, Kluff C, Lobo RA, MacLennan AH, Marsden J, Nappi RE, Panay N, Pickar JH, Robinson D, Simon J, Sitruk-Ware RL, Stevenson JC (2011) Updated IMS recommendations on postmenopausal hormone therapy and preventive strategies for midlife health. *Climacteric* 14(3):302–320
 67. Tannen RL, Weiner MG, Xie D, Barnhart K (2007) A simulation using data from a primary care practice database closely replicated the women’s health initiative trial. *J Clin Epidemiol* 60(7):686–695
 68. Taubman SL, Robins JM, Mittleman MA, Hernan MA (2009) Intervening on risk factors for coronary heart disease: an application of the parametric g-formula. *Int J Epidemiol* 38(6):1599–1611
 69. The Women’s Health Initiative Study Group (1998) Design of the Women’s Health Initiative clinical trial and observational study. *Control Clin Trials* 19(1):61–109
 70. The Writing Group for the Women’s Health Initiative Investigators (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women’s Health Initiative randomized controlled trial. *JAMA* 288(3):321–333
 71. Thompson WA Jr (1977) On the treatment of grouped observations in life studies. *Biometrics* 33(3):463–470
 72. Toh S, Hernán MA (2008) Causal inference from longitudinal studies with baseline randomization. *Int J Biostat* 4(1):22
 73. Toh S, Hernández-Díaz S, Logan R, Robins JM, Hernán MA (2010) Estimating absolute risks in the presence of nonadherence: an application to a follow-up study with baseline randomization. *Epidemiology* 21(4):528–539
 74. Toh S, Hernández-Díaz S, Logan R, Rossouw JE, Hernán MA (2010) Coronary heart disease in postmenopausal recipients of estrogen plus progestin therapy: does the increased risk ever disappear? A randomized trial. *Ann Intern Med* 152(4):211–217
 75. van der Laan M, Petersen ML, Joffe MM (2005) History-adjusted marginal structural models and statically-optimal dynamic treatment regimens. *Int J Biostat* 1:4
 76. Vandenbroucke JP (2004) When are observational studies as credible as randomised trials? *Lancet* 363(9422):1728–1731
 77. Wasserman L (2006) All of nonparametric statistics. Springer, New York
 78. Young JG, Cain LE, Robins JM, O’Reilly EJ, Hernán MA (2011) Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Stat Biosci* 2011(3):119–143