

# On the Use of Biomarkers to Elucidate Clinical Trial Results: Examples from the Women's Health Initiative

Ross L. Prentice and Shanshan Zhao

**Abstract** Biomarkers provide opportunities to maximize the knowledge gained from randomized controlled trials. Applications may include the identification of subpopulations that experience differential treatment effects; the assessment of adherence to treatment or intervention goals; and the elucidation of key biological pathways through which the treatments affect clinical outcomes. This last biomarker role also has implications for the development and initial testing of potential treatments. These types of applications are illustrated using biomarker studies in the Women's Health Initiative postmenopausal hormone therapy and low-fat dietary pattern trials. Related topics are also described where further methodology developments would be helpful.

## 1 Introduction

The Women's Health Initiative (WHI) is a large-scale epidemiologic research program focused on the prevention of chronic disease among postmenopausal women. A total of 161,808 postmenopausal women, in the age range 50–79, were enrolled at 40 U.S. Clinical Centers during 1993–1998. The centerpiece of the WHI is a multifaceted clinical trial of four preventive interventions, in a partial factorial design [1]. A total of 10,739 post-hysterectomy women were randomized to the E-alone trial of 0.625 mg/day of conjugated equine estrogens (Premarin) or placebo; 16,608 women with uterus were randomized to the E + P trial of this same estrogen

---

R.L. Prentice (✉)  
Public Health Sciences, Fred Hutchinson Cancer Center, 1100 Fairview Ave. N.,  
Seattle, WA, 98109 USA  
e-mail: [rprentic@WHL.org](mailto:rprentic@WHL.org)

S. Zhao  
University of Washington, Seattle, WA, USA  
e-mail: [zhaoss41@gmail.com](mailto:zhaoss41@gmail.com)

**Table 1** Clinical outcomes in the WHI postmenopausal hormone therapy trials

Outcomes	E + P trial		E-alone trial	
	Hazard ratio	95% CI <sup>a</sup>	Hazard ratio	95% CI <sup>a</sup>
<b>Coronary heart disease</b>	<b>1.29</b>	<b>1.02–1.63</b>	<b>0.91</b>	<b>0.75–1.12</b>
Stroke	1.41	1.07–1.85	1.39	1.10–1.77
Venous thromboembolism	2.11	1.58–2.82	1.33	0.99–1.79
<b>Invasive breast cancer</b>	<b>1.26</b>	<b>1.00–1.59</b>	<b>0.77</b>	<b>0.59–1.01</b>
Colorectal cancer	0.63	0.43–0.92	1.08	0.75–1.55
Endometrial cancer	0.83	0.47–1.47		
Hip fracture	0.66	0.45–0.98	0.61	0.41–0.91
Death due to other causes	0.92	0.74–1.14	1.08	0.88–1.32
Global index	1.15	1.03–1.28	1.01	0.91–1.12
Number of women	8,506	8,102	5,310	5,429
Follow-up time, mean (SD), mo	62.2 (16.1)	61.2 (15.0)	81.6 (19.3)	81.9 (19.7)

<sup>a</sup> CI confidence interval, from a proportional hazards model stratified by age (5-year categories), and randomization status in the DM trial

preparation plus 2.5 mg/day medroxyprogesterone acetate (Prempro) or placebo; and 48,835 women were randomized to a low-fat dietary pattern (40%) or usual diet (60%). At their one-year anniversary following randomization into either or both of the hormone therapy (HT) or dietary modification (DM) components, participating women were given the opportunity for further randomization to a dietary supplementation trial of 1,000 mg/day calcium carbonate plus 400 international units of vitamin D3 or placebo, and 36,282 women did so. The WHI program is strengthened by the inclusion of a companion cohort study among 93,676 postmenopausal women in the same age range, recruited from essentially the same catchment populations, with much commonality with the clinical trial in methodology, and in data and biospecimen collection.

Table 1 shows key findings from the hormone therapy trials [2, 3] with findings for the designated primary CHD outcome, and the designated primary adverse outcome highlighted. The E + P trial was stopped early in 2002 when health risks were judged to exceed benefits over a 5.6-year average intervention period. The risks included an early elevation in coronary heart disease incidence, the primary trial outcome for which an important risk reduction had been hypothesized, and elevations in stroke and venous thromboembolism incidence. An elevation in breast cancer incidence and a reduction in fracture incidence were also observed, as was hypothesized in trial design [1]. A global index, defined as the time to the earliest of the outcomes listed above it in Table 1 was in the unfavorable direction, and contributed to early stopping considerations. The E-alone trial also stopped early, in 2004, after an intervention period that averaged 7.1 years, substantially because of a stroke elevation of similar magnitude as that observed for E + P, though health risks and benefits and the global index were rather balanced in this trial.

Analyses beyond the summary hazard ratios (HRs) shown in Table 1 took place for each clinical outcome, as well as for some additional important outcomes (e.g., cognition and dementia). These included analyses of HR form as a function of time,

**Table 2** Comparison of cancer incidence rates between intervention and comparison groups in the Women’s Health Initiative (WHI) dietary modification trial

Incidence per 1,000 person-years (number of cases)				
Cancer site	Intervention	Comparison	p <sup>a</sup>	HR (95% CI) <sup>b</sup>
Breast	4.15(655)	4.52(1, 072)	0.09	0.91 (0.83–1.01)
Colorectal	1.27(201)	1.18(279)	0.29	1.08 (0.90–1.29)
Ovary	0.36(57)	0.43(103)	0.03	0.83 (0.60–1.14)
Endometrium	0.79(125)	0.71(170)	0.18	1.11 (0.88–1.40)
All other sites	4.56(720)	4.81(1, 140)	0.30	0.95 (0.86–1.04)
Total cancer	10.69(1, 687)	11.22(2, 661)	0.10	0.95 (0.89–1.01)

Trial includes 19,541 women in the intervention group and 29,294 women in the comparison group

<sup>a</sup> Weighted log-rank test (two-sided) stratified by age (5-year categories) and randomization status in the WHI hormone therapy trial. Weights increase linearly from zero at random assignment to a maximum of 1.0 at 10 years

<sup>b</sup> HR hazard ratio, CI confidence interval, from a proportional hazards model stratified by age (5-year categories), and randomization status in the WHI hormone therapy trial

analyses of HRs among women adherent to their assigned intervention, and various subgroup analyses (with appropriate caveats). Participating women were actively followed beyond the cessation of intervention, giving rise to a range of additional analyses of public health importance [4, 5].

To cite but one example, the more detailed studies of breast cancer incidence in the E + P trial showed an HR that increased unfavorably and approximately linearly to about 1.6 following 5 years of use, and dropped back to basal levels by 2–3 years following trial stoppage. When analyses focused on adherent women, a more dramatic increase to an HR of about 2.5 after 5 years of use was estimated, again with dissipation by 2–3 years following cessation of use [6]. These patterns, in conjunction with the approximately six million women using this estrogen plus progestin preparation in the USA, about 70% of whom stopped shortly after initial trial results [1] were announced, projected a national reduction in breast cancer incidence of about 15,000 women per year as a result of this change in usage patterns, as agrees with subsequent U.S. breast cancer incidence rates [7].

The WHI low-fat dietary pattern trial had dietary intervention goals of equal or less than 20% of energy from fat; five or more fruit and vegetable servings/day, and six or more grain servings/day, with breast and colorectal cancer as primary outcomes, and with ovary and endometrial cancer as additional diet-related cancers that may benefit from this intervention. Table 2 shows principal cancer incidence results from this trial, which proceeded to its planned termination with an 8.1-year average intervention period. The trial design projected a reduction in breast cancer risk with an overall HR of 0.87. The principal targeted dietary change was a reduction in percent of energy from fat, but only about 70% of the hypothesized change was achieved. In correspondence the estimated breast cancer HR of 0.91 [8] differed from unity by about 70% of that projected, but was not significantly different from one (weighted logrank p = 0.09). The corresponding contrast for ovarian cancer incidence [9] was nominally significant (p = 0.03), providing an important lead for a disease having few known modifiable risk factors. For both breast [8] and

ovarian cancer [9], there was a significant interaction between baseline percent of energy from fat and HR, with stronger evidence for an intervention effect among women having a high fat content in their customary diet. These women made a comparatively larger reduction in percent of energy from fat, if assigned to the dietary intervention group.

The calcium and vitamin trial did not provide significant evidence of a treatment effect, either for its primary hip fracture outcome [10], or secondary outcomes (colorectal cancer, other fractures).

## **2 Biomarkers and Variations in Clinical Trial Intervention Effects**

Even though it is good clinical trial practice to focus primarily on overall treatment effects as opposed to those in subsets of a study population, it needs to be recognized that hazard ratios may, and often do, vary according to specific characteristics of the study population. Notably, HRs provide but one way of summarizing a treatment effect over time, and lack of variation on an HR scale may differ from corresponding lack of variation on other assessment scales. That said, however, HR or other ratio measures (e.g., odds ratios) seem particularly useful in leading to simple models, wherein the joint association of treatment and study subject characteristics or exposures on clinical outcomes often seems to depart little from a multiplicative model.

For example, for breast cancer incidence, no interacting demographic or clinical variables were found for E + P [11], whereas for E-alone a suggested reduction in risk seemed to be largely confined to lower risk women, specifically those without benign breast disease or a family history of breast cancer [12].

Several nested case-control studies within the HT trial cohorts were conducted in an attempt to identify biomarkers that may interact with hormone therapy HRs, or that may mediate the observed intervention effects on clinical outcomes. These studies primarily focused on biochemical and genetic markers that were recognized risk indicators for the clinical outcomes under study. For example, a Cardiovascular Disease Biomarker Study focused on markers of inflammation, coagulation/thrombosis, lipids and lipoproteins, and related genetic variants, for each of coronary heart disease, stroke, and venous thromboembolism. These studies [13, 14] generally confirmed associations with disease risk, but there were few interacting factors identified and none of the observed biomarker changes following intervention activities appeared to meaningfully mediate the observed treatment effects, a topic that will be discussed further below. As an example of an interacting variable, women having a relatively high baseline low-density lipoprotein (LDL) cholesterol who were assigned to active hormone therapy evidently experienced a comparatively larger early elevation in coronary heart disease risk [13].

Some high-dimensional genotype biomarker studies were also conducted, in an attempt to understand more of the biology related to observed clinical effects in

the WHI trials. For example, a breast cancer nested case-control study involved the genotyping of 9,039 single nucleotide polymorphisms (SNPs) for 2,166 women who developed breast cancer during the trial intervention period. A randomized trial context is well suited to genotype by treatment interaction testing in that “case-only” analyses, which require genotype data only on study subjects developing disease, have efficiency about the same as if genotyping had been conducted on the full cohort.

More specifically, let  $V = 1$  and  $V = 0$  denote active and control randomization assignments, and  $z = 0, 1, \text{ or } 2$  denote the number of minor alleles of an SNP. A simple Cox model that stratifies on SNP genotype, and allows a separate HR parameter for treatment at each value of  $z$ , can be written

$$\lambda(t; V, z) = \lambda_{0z}(t) \exp\{\beta_0 I(z = 0) + \beta_1 I(z = 1) + \beta_2 I(z = 2)\},$$

where  $I(\cdot)$  denotes an indicator variable, and  $e^{\beta z}$  is the HR for women having SNP genotype  $z$ , for  $z = 0, 1, \text{ or } 2$ . From this expression,

$$\text{logit}(V|X = t, z) = \text{logit}(V|X \geq t, z) + \sum_{i=0}^2 \beta_i I(z = z_i),$$

where  $X = t$  denotes disease occurrence at time  $t$  following randomization. The important feature here is that  $V$  is orthogonal to  $z$  by virtue of randomization so that if the disease is rare one has, to a good approximation

$$\text{logit}(V|T \geq t, z) = \log\{q/(1 - q)\},$$

where  $q = \text{pr}(V = 1)$  is the randomization fraction for the active treatment group. It follows that one can estimate intervention HRs at each SNP genotype by ordinary logistic regression of the randomization indicator  $V$  on indicator variables for the number of minor SNP alleles, with  $\log\{q/(1 - q)\}$  as an “offset.” Breast cancer analyses of this type yielded nominally significant variations in the intervention HR with a SNP (rs3750817) in intron 2 of the fibroblast growth factor receptor gene on chromosome 10 for both E + P and E-alone [15] and for the dietary modification intervention in the subset of women (denoted DMQ) who were in the upper quartile of percent of energy from fat in their baseline diet [16]. These analyses also drew attention (nominal  $p < 0.05$ ) to a SNP (rs7705343) in the mitochondrial ribosomal protein S30 region of chromosome five [17] for each of E-alone, DMQ, and for the calcium and vitamin D intervention (for which there was no breast cancer “main” effect). For either SNP the more favorable intervention effects were evidently localized among women who were homozygous for the SNP minor allele (TT genotype for rs3750817; AA for rs7705343). A challenge with these types of suggestive findings is the identification of a research strategy for replication. Observational studies pertinent to these intervention topics may be limited in their assessment of related exposures (e.g., hormonal or dietary exposures), and may be

subject to important confounding or measurement biases. In general, the methods for identification of genotype by environmental factor interaction evaluation are at an early stage of development, and large-scale clinical trial settings have much to offer in this arena.

### 3 Biomarkers of Intervention Adherence and Exposure

As mentioned above, the Dietary Modification trial evidently achieved only about 70% of its projected intervention versus control group difference for the principal dietary intervention target, percent of energy from fat. Even this 70% assessment relies on self-reported dietary information from participating women. Specifically, based on food frequency questionnaire (FFQ) assessments, intervention group women reported a 10.7% average lower percent of energy from fat compared to the control group at 1-year following randomization; 9.5% at 3 years after randomization; and 8.1% at 6 years after randomization. However, the FFQ data also indicate a differential total energy consumption by about 100 kilocalories/day, which is not consistent with the weight changes experienced by trial participants (2–3 kg greater weight loss in intervention group compared to usual diet control group at 1-year, which mostly dissipated over the subsequent 5 years). If the greater underreporting of energy by intervention group women pertained disproportionately to fat calories, then percent of energy from fat would also be differentially reported and the power of the DM trial accordingly affected.

In fact, the dietary assessment measurement issue is even more acute in observational nutritional epidemiology studies, where systematic and random assessment errors could well distort the very associations under study, rather than simply reducing study power as in the intervention trial setting. In either context, however, biomarkers provide important avenues for strengthening the research agenda.

Two nutritional biomarker substudies have been conducted in WHI cohorts, and a controlled human feeding study that aims to develop biomarkers for additional nutrients and foods is currently underway. The first, the Nutrient Biomarker Study (NBS) included energy [18] and protein [19] biomarkers and a concurrent FFQ, among a representative 544 weight-stable women in the DM trial. The second, the Nutrition and Physical Activity Assessment Study included these biomarkers and self-reports of dietary frequencies, records and recalls, along with a biomarker of activity-related energy expenditure and three types of physical activity self-report, among 450 representative women from the WHI Observational (cohort) Study. By simple linear regression of log-biomarker assessments on corresponding log self-reports and on readily available study subject characteristic data (body mass index, age, and ethnicity), calibrated consumption estimates were developed for energy, protein, and percent of energy from protein. For example, for energy, even though the log self-report data explained only a few percent (e.g., 3–4% for the FFQ) of the variation in log-biomarker values, the inclusion of these other factors in the regression equation raised this percentage to the 40–45% range. Upon extracting the

temporal variation in the biomarker, this increased to about 70% of the average daily energy consumption variation over a 1-year study period [20]. The NBS equations were used to develop calibrated-energy, protein, and percent of energy from protein estimates throughout the WHI cohorts, and positive associations between energy and several major cancers [21] as well as coronary disease [22] and diabetes [23] were found that were not apparent without calibration. The role of body mass index in these analyses is complex [24], and the associations just mentioned seemed substantially, if not entirely, explained by body fat accumulation over time.

This nutritional epidemiology research area and the similarly important physical activity epidemiology area are ripe for further development, with a substantial use of biomarkers providing a logical next step in the overall research agenda.

The energy biomarker data indicates severe underreporting using the FFQ, by about 30% overall, and with much greater underreporting among overweight and obese women, along with greater underreporting by younger compared to older postmenopausal women. These analyses also suggest some energy underreporting by intervention compared to control group women also, by about 100 kcal/day, allowing the weight change data mentioned above to align with corresponding calibrated energy consumption in the DM trial.

## 4 Biomarkers as Mediators of Clinical Trial Intervention Effects

Again let  $V = 1$  or  $0$  denote active and control randomization assignments in a clinical trial, but now let  $z$  denote a biomarker change following some period of intervention activities. Analyses may aim to understand the extent to which intervention effects on the clinical outcome are mediated by  $z$ . A traditional mediation analysis would compare the coefficient of  $V$  in a regression analysis that doesn't include  $z$  with a corresponding analysis including  $z$ , with evidence for mediation if the coefficient for  $V$  moves substantially toward the null when  $z$  is added to the regression model.

A key statistical difference is evident between the type of interaction analysis discussed in Sect. 2, where  $V$  and  $z$  are independent by study design, and mediation analyses where  $V$  and  $z$  may be highly correlated. This is a critical point. In the extreme, for example, if the biomarker doesn't change in the control group ( $z = 0$ ) and changes by exactly the same amount ( $z = c$ , for some  $c \neq 0$ ) among all study subjects in the intervention group, then  $z$  and  $V$  will be perfectly correlated, and it will not be possible to carry out an analysis that simultaneously models  $V$  and  $z$ . As a plausible departure from this scenario, suppose that  $z$  is constant in both groups, but that  $z$  is assessed with some technical measurement error. As amplified below, the regression analyses will then be possible, but  $z$  may then appear not to mediate, even though the biomarker change in question may be central to explaining the intervention effects on clinical outcomes.

To elaborate just a little, consider baseline,  $x_0$ , and post-intervention,  $x_1$ , biomarker values, and suppose that the biomarker fully mediates an intervention effect in a linear model for a quantitative response  $Y$ . Hence,  $E(Y; x_0, x_1, V) = a + a_0x_0 + a_1x_1$ , with  $a_1 \neq 0$ . Under a bivariate normal model for  $(x_0, x_1)$  with mean  $(\mu_0, \mu_1 + dV)$ , common variance  $\sigma^2$  and correlation  $\rho$ , one can derive

$$E(Y; x_0, V) = a' + a'_0x_0 + (a_1d)V,$$

where  $a'$  and  $a'_0$  are simple functions of the response and biomarker parameters, so that a mediation analysis would compare an estimate of  $a_1d$  to an estimate of the coefficient of  $V$  (zero) when  $x_1$  is added to the regression model.

Now suppose that  $x_0$  and  $x_1$  incorporate classical normal measurement error, so that one measures  $w_0 = x_0 + e_0$  and  $w_1 = x_1 + e_1$  where  $e_0$  and  $e_1$  are independent mean zero normal variates having variance  $\sigma_e^2$ . It is straightforward to show that the coefficient of  $V$  in  $E(Y; w_0, V)$  is again  $a_1d$ , unaffected by measurement error owing to the independence between  $x_0$  and  $V$ , but that for  $V$  in  $E(Y | w_0, w_1, V)$  is a complicated function of model parameters that approaches  $(a_0 + a_1)d/(2 + \delta^2)$  as  $\rho \rightarrow 1$ , where  $\delta^2 = \sigma_e^2/\sigma^2$ . This limiting coefficient can be very far from zero even if  $\delta^2$  is small! This rather counter-intuitive result arises because of the diminishing ability to distinguish the biomarker effect from the overall treatment effect on  $Y$ , as  $\rho \rightarrow 1$ . It follows that careful modeling of the biomarker and its measurement process may be needed to reliably assess mediation, or more generally, to assess treatment effects after allowing for certain biomarker changes.

As noted above, none of the candidate biomarkers studied appeared to mediate HT effects on cardiovascular diseases in the WHI hormone therapy trials. We undertook additional “discovery” research to identify blood biomarkers that are risk markers for these diseases, and that are affected by hormone therapy. This work focused on protein expression, using an Intact Protein Analysis System [25] having the capability of quantitatively comparing concentrations between pairs of specimens for about 350–400 proteins across a substantial dynamic range. Specifically, concentrations based on blood collected 1-year after randomization were compared to corresponding baseline concentrations for 50 women adherent to active E-alone, and 50 women adherent to active E + P, over the first year of HT trial participation. For throughput reasons IPAS analyses were based on pools formed from equal volumes of serum from 10 women. A total of 378 proteins were quantified for change. Of these, a remarkable 44.7% (169/378) had evidence of change ( $p < 0.05$ ) following intervention with E-alone or E + P [26, 27]. The protein changes were mostly quite similar for E-alone and E + P, and included proteins in multiple biological pathways relevant to observed clinical effects, including inflammation, coagulation, immune function, cell adhesion, growth factors, and osteogenesis, among others.

Corresponding analyses were then conducted, using the same proteomic platform, to compare baseline blood protein concentrations between women who went on to develop CHD or stroke and corresponding matched controls, with cases and controls drawn from the WHI Observational Study. This time larger pools of size



100 were employed. There were eight such pool pairs for each of these diseases, as well as for breast cancer. From the resulting data [28] there were 37 proteins having nominal  $p < 0.05$  for a CHD case versus control difference compared to 17.3 expected by chance; and 47 for stroke compared to 18.3 expected by chance. Several of these had estimated false discovery rates  $< 0.05$  and most of these were among the proteins evidently affected by E-alone and/or E + P. These provide novel candidates for mechanistic effects of HT on these cardiovascular diseases.

We are still at an early stage of evaluating these candidates for mediation in the HT trials. An initial evaluation involving beta-2 microglobulin, a highly ranked protein for CHD association, and insulin-like growth factor binding protein 4 (IGFBP4), a highly ranked protein for stroke association, confirmed the association of these proteins with disease incidence in the WHI trials but, once again, change in protein concentration following hormone therapy treatment did not seem to mediate intervention effects on these diseases, at least not without explicit account of the biomarker measurement error process. Further, analyses with these analytes revealed that HT hazard ratios, in the presence of baseline and 1-year biomarker measurements were quite sensitive to the ratio ( $\delta^2$ ) of the measurement error variance to the underlying biomarker variance for both E-alone and E + P, with larger  $\delta^2$  values consistent with full mediation.

These preliminary analyses reinforce the need for enhanced statistical methods for identifying the important biological intermediaries of intervention effects in clinical trials. Adequate modeling of the underlying biomarker process, and of the departure of such models from corresponding measured biomarker values, may typically require biomarker assessments at more than two time points in conjunction with large case and control sample sizes. This topic, and methods for correcting treatment hazard ratio estimates for the biomarker measurement process, will be discussed in more detail elsewhere.

## 5 Biomarkers for Intervention Development

In recent years biomarkers have come to play a rather central role in treatment development, particularly in the therapeutics area. For example, the molecular characteristics of patient tumors may identify key therapeutic targets for disruption by potential treatments. High-dimensional data, including gene expression profiles, may help to focus developmental efforts toward therapeutic benefit or toward the avoidance of certain adverse effects.

The development and initial testing of preventive interventions is a rather under-developed aspect of the chronic disease prevention research agenda. Sometimes it is attractive to move interventions from therapeutics to primary prevention. Examples include statins for heart disease prevention; tamoxifen, SERMS, or aromatase inhibitors for breast cancer prevention; or biphosphonates for fracture prevention. However, this approach seems unlikely to lead to the behavioral changes, for example, in the diet and physical activity area, that arguably provide the ultimate

preventive approaches needed. Observational epidemiology has much to offer for identifying preventive approaches, but findings may lack the needed specificity and force to fuel needed behavioral, regulatory, or policy changes. For example, one can contrast the influence of the rather extensive body of observational research on postmenopausal hormones, with that of the comparatively few clinical trials that eventually were able to be conducted.

Intermediate outcome trials, which have outcomes on putative pathways between treatments and clinical outcomes of interest, have considerable potential to add to these other data sources for preventive intervention development. For example, a trial of moderate size, called the Postmenopausal Estrogen Progestin Intervention (PEPI) trial was initiated in advance of the WHI trials to compare various hormone regimens in respect to cardiovascular disease risk factors, uterine hyperplasia, and other intermediate outcomes [29, 30]. This trial had an influence on the choice of regimens studied in the WHI trials, but it did not warn, for example, concerning the observed early elevation in CHD, or the sustained elevation in stroke that emerged in the WHI hormone therapy trials.

Intermediate outcome trials that combine changes in major risk factors for clinical outcomes of interest with more agnostic, possibly high-dimensional, changes in blood or other biospecimens, may offer a more comprehensive approach to the selection and initial evaluation of preventive interventions. For example, the agnostic aspect could entail study of potential intervention effects on the plasma proteome and metabolome. These data, whether for a few emergent candidates or for a high-dimensional set of changes, could then be merged with observational analyses relating the entire set of intermediate variables to clinical outcomes of interest, to develop projections of intervention effects on each such outcome. This approach could be considered for behavioral as well as chemopreventive interventions. The two data sources to be combined would each involve studies having a small fraction of the cost of a full-scale prevention trial. While not sufficient in itself, such an approach could augment the value of intermediate outcome trials and, in particular, may help to filter intervention options arising from the traditional data sources mentioned above, thereby permitting a focus on the more strongly justified concepts for full-scale trial consideration with clinical outcomes.

## 6 Discussion and Summary

Biomarkers have potential to play several important roles in the development, conduct, analysis, and reporting of clinical trials. Specifically, biomarkers may permit stratification of study subjects according to the magnitude of beneficial or adverse treatment effects, possibly leading to the identification of persons for whom the treatment can be particularly recommended or should be avoided.

Though not much emphasized here, biomarkers typically play a key role in the assessment of adherence to intervention goals, and in the assessment of adherence-adjusted treatment effects. Biomarkers also provide the principal approach to

identifying the important biological pathways whereby a treatment may influence a clinical outcome of interest. The utility of biomarkers for each of these purposes, but especially for the elucidation of disease mechanisms can be expected to depend strongly on the properties of the biomarker measurement process, and on the ability to adequately model and correct for measurement error in data analyses.

The adherence-adjustment and mediation applications of biomarkers are sometimes posed using a potential outcomes, and a principal stratification formulation [31]. The principal stratification “framework,” however, seems too restrictive to be very useful in this type of biomedical research context [32, 33], and the important measurement error issue discussed here does not seem to have been addressed in the potential outcomes context.

Finally, high-dimensional biomarkers from discovery platforms evidently have an important role to play in intervention development, though this concept has yet to be much explored to date for preventive intervention development.

**Acknowledgements** This work was supported by NIH/NCI grants P01-CA53996, R01-CA119171, R21HL109527, R01CA149135 U01-CA86368-06, and U19-CA148065-01; and the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contract N01WH22110.

## References

1. 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
2. Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, Jackson RD, Beresford SA, Howard BV, Johnson KC, Kotchen JM, Ockene J, 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
3. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, Bonds D, Brunner R, Brzyski R, Caan B, Chlebowski R, Curb D, Gass M, Hays J, Heiss G, Hendrix S, Howard BV, Hsia J, Hubbell A, Jackson R, Johnson KC, Judd H, Kotchen JM, Kuller L, LaCroix AZ, Lane D, Langer RD, Lasser N, Lewis CE, Manson J, Margolis K, Ockene J, O’Sullivan MJ, Phillips L, Prentice RL, Ritenbaugh C, Robbins J, Rossouw JE, Sarto G, Stefanick ML, Van Horn L, Wactawski-Wende J, Wallace R, Wassertheil-Smoller S, Women’s Health Initiative Steering Committee (2004) Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The Women’s Health Initiative randomized controlled trial. *JAMA* 291(14):1701–1712
4. Heiss G, Wallace R, Anderson GL, Aragaki A, Beresford SA, Brzyski R, Chlebowski RT, Gass M, LaCroix A, Manson JE, Prentice RL, Rossouw J, Stefanick ML, WHI Investigators (2008) Health risks and benefits 3 years after stopping randomized treatment with estrogen and progestin. *JAMA* 299(9):1036–1045
5. LaCroix AZ, Chlebowski RT, Manson JE, Aragaki AK, Johnson KC, Martin L, Margolis KL, Stefanick ML, Brzyski R, Curb JD, Howard BV, Lewis CE, Wactawski-Wende J, WHI Investigators (2011) Health outcomes after stopping conjugated equine estrogens among postmenopausal women with prior hysterectomy: a randomized controlled trial. *JAMA* 305(13):1305–1314

6. Chlebowski RT, Kuller LH, Prentice RL, Stefanick ML, Manson JE, Gass M, Aragaki AK, Ockene JK, Lane DS, Sarto GE, Rajkovic A, Schenken R, Hendrix SL, Ravdin PM, Rohan TE, Yasmeeen S, Anderson G, WHI Investigators (2009) Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med* 360(6):573–587
7. Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, Edwards BK, Berry DA (2007) The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* 356(16):1670–1674
8. Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, Margolis KL, Limacher MC, Manson JE, Parker LM, Paskett E, Phillips L, Robbins J, Rossouw JE, Sarto GE, Shikany JM, Stefanick ML, Thomson CA, Van Horn L, Vitolins MZ, Wactawski-Wende J, Wallace RB, Wassertheil-Smoller S, Whitlock E, Yano K, Adams-Campbell L, Anderson GL, Assaf AR, Beresford SA, Black HR, Brunner RL, Brzyski RG, Ford L, Gass M, Hays J, Heber D, Heiss G, Hendrix SL, Hsia J, Hubbell FA, Jackson RD, Johnson KC, Kotchen JM, LaCroix AZ, Lane DS, Langer RD, Lasser NL, Henderson MM (2006) Low-fat dietary pattern and risk of invasive breast cancer: The Women’s Health Initiative Randomized Controlled Dietary Modification Trial. *JAMA* 295(6):629–642
9. Prentice RL, Thomson CA, Caan B, Hubbell FA, Anderson GL, Beresford SAA, Pettinger M, Lane DS, Lessin L, Yasmeeen S, Singh B, Khandekar J, Shikany JM, Satterfield S, Chlebowski RT (2007) Low-fat dietary pattern and cancer incidence in the Women’s Health Initiative Dietary Modification Randomized Controlled Trial. *J Natl Cancer Inst* 99(20):1534–1543
10. Jackson RD, LaCroix AZ, Gass M, Wallace RB, Robbins J, Lewis CE, Bassford T, Beresford SA, Black HR, Blanchette P, Bonds DE, Brunner RL, Brzyski RG, Caan B, Cauley JA, Chlebowski RT, Cummings SR, Granek I, Hays J, Heiss G, Hendrix SL, Howard BV, Hsia J, Hubbell FA, Johnson KC, Judd H, Kotchen JM, Kuller LH, Langer RD, Lasser NL, Limacher MC, Ludlam S, Manson JE, Margolis KL, McGowan J, Ockene JK, O’Sullivan MJ, Phillips L, Prentice RL, Sarto GE, Stefanick ML, Van Horn L, Wactawski-Wende J, Whitlock E, Anderson GL, Assaf AR, Barad D, Women’s Health Initiative Investigators (2006) Calcium plus vitamin D supplementation and the risk of fractures. *N Engl J Med* 354(7):669–683
11. 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, WHI Investigators (2003) Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women’s Health Initiative Randomized Trial. *JAMA* 289(24):3243–3253
12. Stefanick ML, Anderson GL, Margolis KL, Hendrix SL, Rodabough RJ, Paskett ED, Lane DS, Hubbell FA, Assaf AR, Sarto GE, Schenken RS, Yasmeeen S, Lessin L, Chlebowski RT, WHI Investigators (2006) Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. *JAMA* 295(14):1647–1657
13. Rossouw JE, Cushman M, Greenland P, Lloyd-Jones DM, Bray P, Kooperberg C, Pettinger M, Robinson J, Hendrix S, Hsia J (2008) Inflammatory, lipid, thrombotic, and genetic markers of coronary heart disease risk in the Women’s Health Initiative trials of hormone therapy. *Arch Intern Med* 168(20):2245–2253
14. Kooperberg C, Cushman M, Hsia J, Robinson J, Aragaki A, Lynch J, Baird A, Johnson K, Kuller L, Beresford S, Rodriguez B (2007) Can biomarkers identify women at increased stroke risk? The Women’s Health Initiative hormone trials. *PLoS Clin Trials* 2(6):e28
15. Prentice RL, Huang Y, Hinds DA, Peters U, Pettinger M, Cox DR, Beilharz E, Chlebowski RT, Rossouw JE, Caan B, Ballinger DG (2009) Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidemiol Biomarker Prev* 18(11):3079–3085
16. Prentice RL, Huang Y, Hinds DA, Peters U, Cox DR, Beilharz E, Chlebowski RT, Rossouw JE, Caan B, Ballinger DG (2010) Variation in the FGFR2 gene and the effect of a low-fat dietary pattern on invasive breast cancer. *Cancer Epidemiol Biomarker Prev* 19:74–79
17. Huang Y, Ballinger DG, Dai J, Peters U, Hinds DA, Cox DR, Beilharz E, Chlebowski RT, Rossouw JE, McTiernan A, Rohan T, Prentice RL (2011) Genetic variants in the MRPS30 region and postmenopausal breast cancer risk. *Genome Med* 3:42

18. Schoeller DA (1999) Recent advances from application of doubly labeled water to measurement of human energy expenditure. *J Nutr* 129(10):1765–1768
19. Bingham S (1994) The use of 24-h urine samples and energy expenditure to validate dietary assessments. *Am J Clin Nutr* 59(1):2275–2315
20. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SAA, Caan B, Tinker LF, Schoeller D, Bingham S, Eaton CB, Thomson C, Johnson KC, Ockene J, Sarto G, Heiss G, Neuhauser ML (2011) Evaluation and comparison of food records, recalls and frequencies for energy and protein assessment using recovery biomarkers. *Am J Epidemiol* 174(5):591–603
21. Prentice RL, Shaw PA, Bingham SA, Beresford SA, Caan B, Neuhauser ML, Patterson RE, Stefanick ML, Satterfield S, Thomson CA, Snetselaar L, Thomas A, Tinker LF (2009) Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *Am J Epidemiol* 169(8):977–989
22. Prentice RL, Huang Y, Kuller LH, Tinker LF, Van Horn L, Stefanick ML, Sarto G, Ockene J, Johnson KJ (2011) Biomarker-calibrated energy and protein consumption and cardiovascular disease risk among postmenopausal women. *Epidemiology* 22(2):170–179
23. Tinker LF, Sarto GE, Howard BV, Huang Y, Neuhauser ML, Mossavar-Rahmani Y, Beasley JM, Margolis KL, Eaton CB, Phillips LS, Prentice RL (2011) Biomarker-calibrated dietary energy and protein intake association with diabetes risk among postmenopausal women from the WHI. *Am J Clin Nutr* 94(6):1600–1606, Epub 2011 Nov 9
24. Prentice RL, Huang Y (2011) Measurement error modeling and nutritional epidemiology association analyses. *Can J Statistics* 39:498–509. doi:[10.1002/cjs](https://doi.org/10.1002/cjs)
25. Faca V, Coram M, Phanstiel D, Glukhova V, Zhang Q, Fitzgibbon M, McIntosh M, Hanash S (2006) Quantitative analysis of acrylamide labeled serum proteins by LC-MS/MS. *J Proteome Res* 5(8):2009–2018
26. Katayama H, Paczesny S, Prentice RL, Aragaki A, Faca VM, Pitteri SJ, Zhang Q, Wang H, Silva M, Kennedy J, Rossouw J, Jackson R, Hsia J, Chlebowski R, Manson J, Hanash S (2009) Application of serum proteomics to the Women's Health Initiative conjugated equine estrogens trial reveals a multitude of effects relevant to clinical findings. *Genome Med* 1(4):47.1–47.16
27. Pitteri SJ, Hanash SM, Aragaki A, Amon L, Chen L, Busald Buson T, Paczesny S, Katayama H, Wang H, Johnson MM, Zhang Q, McIntosh M, Wang P, Kooperberg C, Rossouw JE, Jackson R, Manson JE, Hsia J, Liu S, Martin L, Prentice RL (2009) Postmenopausal estrogen and progestin effects on the serum proteome. *Genome Med* 1(12):121.1–121.14
28. Prentice RL, Paczesny SJ, Aragaki A, Amon LM, Chen L, Pitteri SJ, McIntosh M, Wang P, Busald Buson T, Hsia J, Jackson RD, Rossouw JE, Manson JE, Johnson K, Eaton C, Hanash SM (2010) Novel proteins associated with risk for coronary heart disease or stroke among postmenopausal women identified by in-depth plasma proteome profiling. *Genome Med* 2(7):48–60
29. PEPI Trial Writing Group (1995) Effects of estrogen or estrogen/progestin regimes on heart disease risk factors in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA* 273(3):199–208
30. PEPI Trial Writing Group (1996) Effects of hormone replacement therapy on endometrial histology in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA* 275(5):370–375
31. Frangakis CE, Rubin DB (2002) Principal stratification in causal inference. *Biometrics* 58:21–29
32. Pearl J (2011) Principal stratification – a goal or a tool? *Int J Biostat* 7(1). pii: Article 20
33. Prentice RL (2011) Invited commentary on pearl and principal stratification. *Int J Biostat* 7(1):Article 30