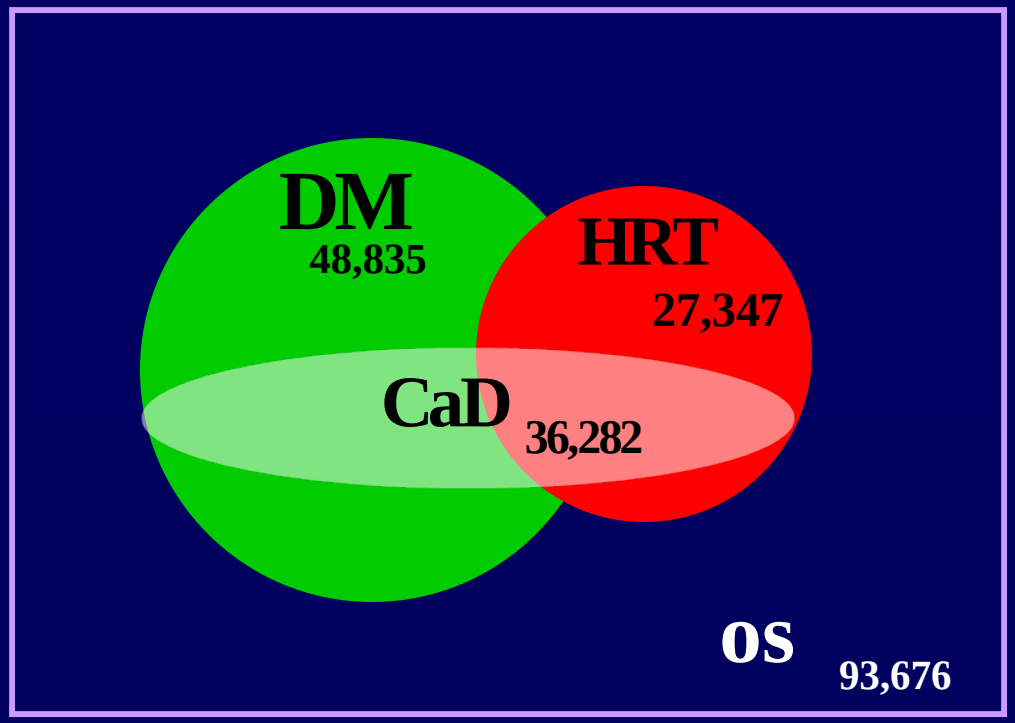
On the Use of Biomarkers to Elucidate Clinical Trial Results: Examples from the Women's Health Initiative

Ross L. Prentice
Fred Hutchinson Cancer Research Center
and University of Washington

How can we obtain answers concerning health benefits and risks of behavior changes/preventive interventions, and know that the answers are reliable?

- WHI postmenopausal hormone therapy and a low-fat dietary pattern trial findings, and cohort study interface
- Biomarkers and variation in clinical trial intervention effects
- Biomarkers for intervention exposure assessment
- Biomarkers for preventive intervention development, and for studies of intervention effects mediation

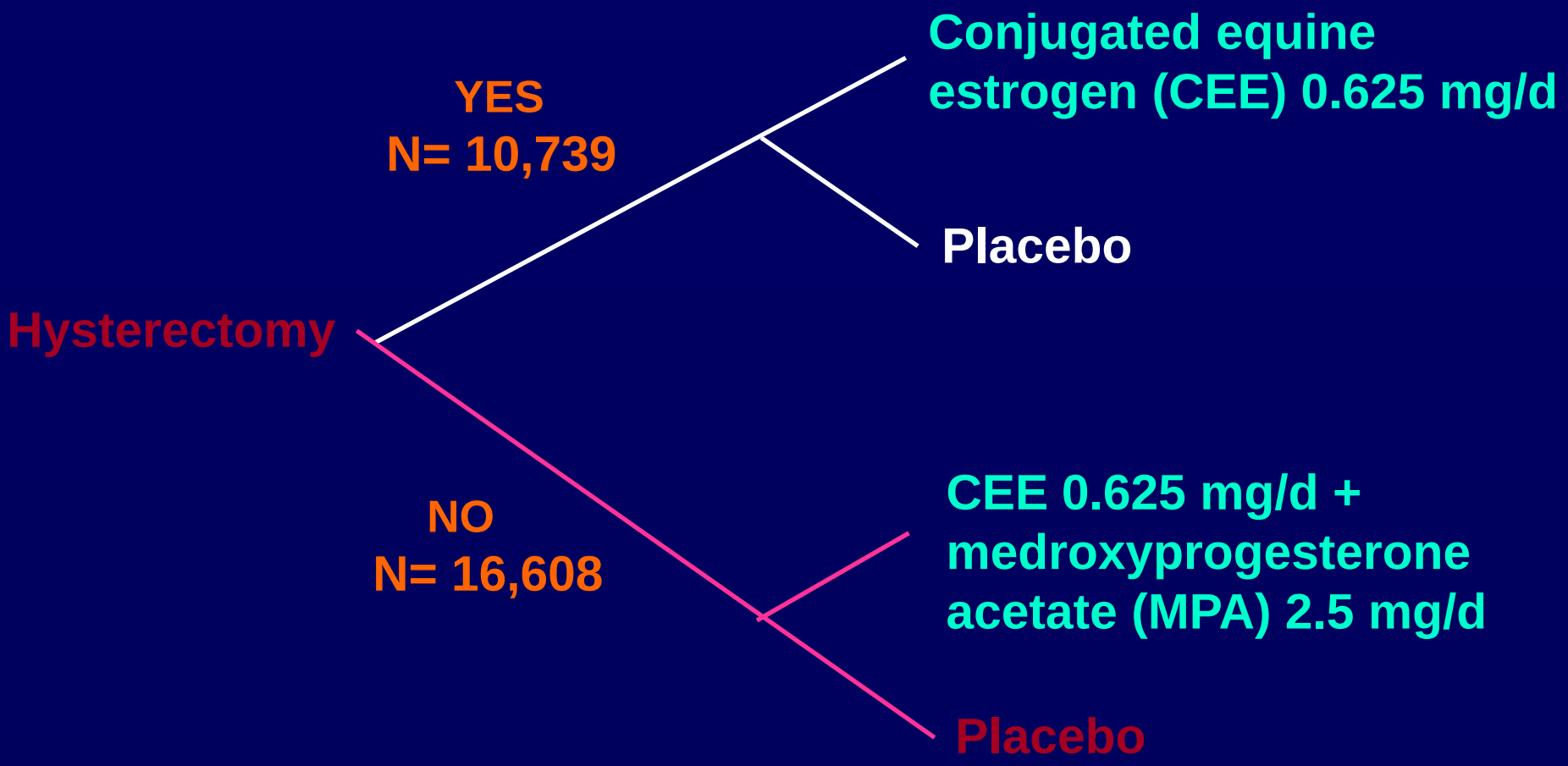
Design of WHI



CT=68,132

WHI =161,808

Design of WHI Hormone Therapy Trials

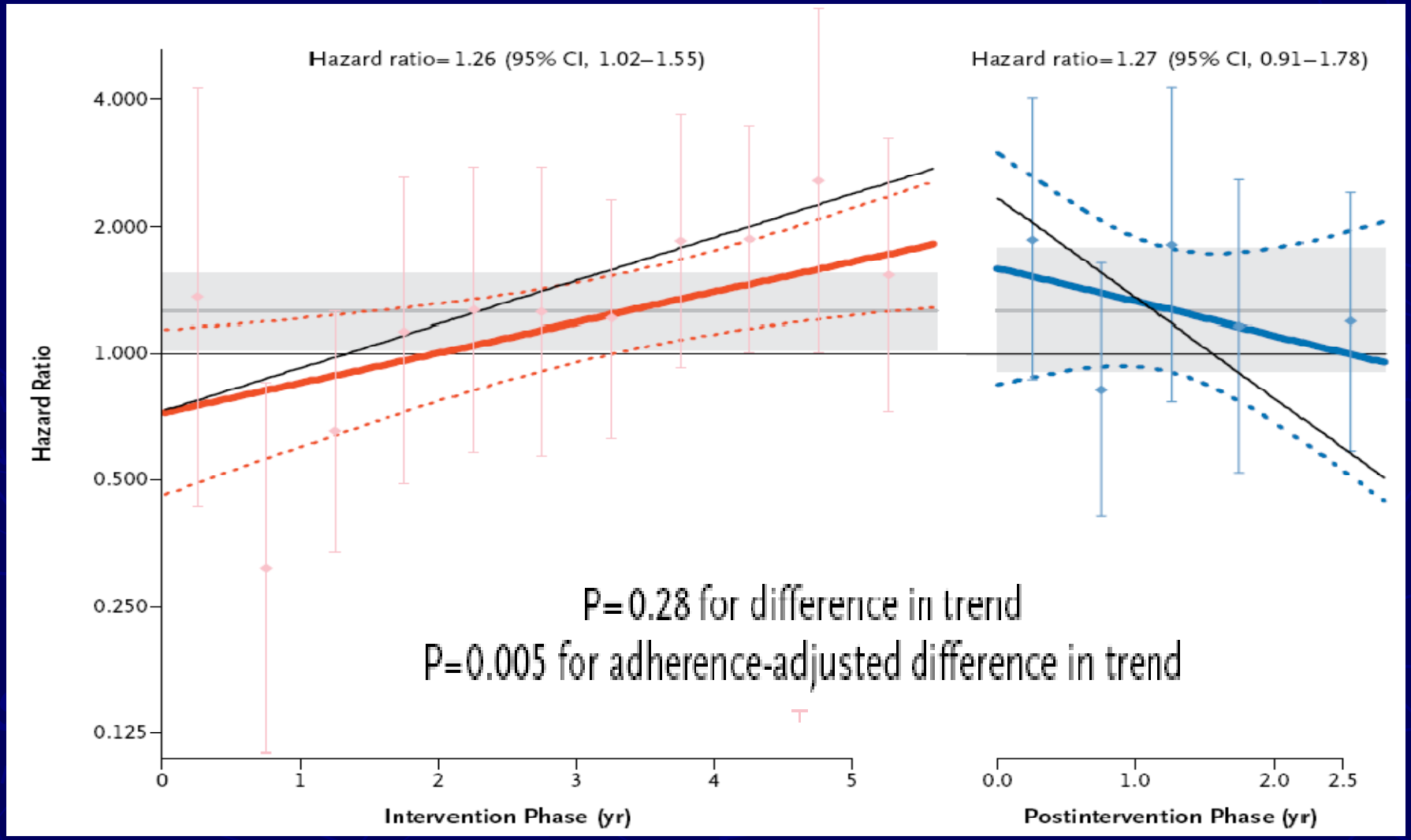


Clinical Outcomes in the WHI Postmenopausal Hormone Therapy Trials

(Rossouw et al, JAMA 2002; Anderson et al, JAMA 2004)

Outcomes	E+P Trial		E-Alone Trial	
	Hazard Ratio	95% CI	Hazard Ratio	95% CI
Coronary heart disease	**1.29**	**1.02 - 1.63**	**0.91**	**0.75 - 1.12**
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
Invasive breast cancer	**1.26**	**1.00 - 1.59**	**0.77**	**0.59 - 1.01**
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	8506	8102	5310	5429
Follow-up time, mean (SD), mo	62.2 (16.1)	61.2 (15.0)	81.6 (19.3)	81.9 (19.7)

Clinical Trial: CEE+MPA vs. Placebo Breast Cancer Risk During Intervention and Postintervention (Chlebowski et al, 2009 NEJM)



Breast Cancer Hazard Ratio Estimates according to Prior Postmenopausal Hormone Therapy Status, Years from Hormone Therapy Initiation, and Gap Time from Menopause to Hormone Therapy Initiation, among Women Adhering to their Baseline Hormone Therapy Status

	Prior HT	No Prior HT		
Gap Time Periods*		<5	5-15	>15
	HR (95% CI)	HR (95% CI)	HR (95% CI)	HR (95% CI)
CEE				
Years from HT Initiation				
< 2	1.52 (0.63, 3.65)	1.08 (0.14, 8.19)	0.56 (0.13, 2.43)	0.92 (0.34, 2.47)
2 - 5	0.76 (0.40, 1.45)	1.61 (0.67, 3.88)	0.76 (0.29, 1.99)	0.61 (0.25, 1.46)
> 5	0.85 (0.46, 1.58)	0.98 (0.54, 1.75)	0.77 (0.40, 1.50)	0.76 (0.35, 1.64)
HT in OS/HT in CT	1.07 (0.61, 1.89)			
CEE/MPA				
Years from HT Initiation				
< 2	1.61 (0.76, 3.41)	1.32 (0.56, 3.11)	0.73 (0.35, 1.52)	0.37 (0.13, 1.06)
2 - 5	3.51 (1.81, 6.81)	1.85 (1.03, 3.34)	1.61 (1.00, 2.59)	0.81 (0.44, 1.48)
> 5	2.76 (1.31, 5.79)	2.75 (1.73, 4.39)	2.00 (1.19, 3.36)	1.23 (0.53, 2.86)
HT in OS/HT in CT	1.15 (0.74, 1.80)			

*Gap time in years from menopause to first use of HT

Low-Fat Dietary Pattern Trial:
Findings and Methodology

Intervention Group Goals:

- 20% energy from fat
- 5 or more fruit and vegetable servings daily
- 6 or more grain servings daily

Photos courtesy of USDA Agricultural Research Service

Comparison of Cancer Incidence Rates between Intervention and Comparison Groups in the Women's Health Initiative (WHI) Dietary Modification Trial*

Prentice et al (JAMA, 2006; JNCI, 2007); Beresford et al (JAMA, 2006)

Incidence per 1000 person-years				
	(Number of cases)			
Cancer Site	Intervention	Comparison	p†	HR(95% CI)‡
Breast	4.15 (655)	4.52 (1072)	.09	0.91 (0.83 to 1.01)
Colorectal	1.27 (201)	1.18 (279)	.29	1.08 (0.90 to 1.29)
Ovary	0.36 (57)	0.43 (103)	.03	0.83 (0.60 to 1.14)
Endometrium	0.79 (125)	0.71 (170)	.18	1.11 (0.88 to 1.40
All other sites	4.56 (720)	4.81 (1140)	.30	0.95 (0.86 to 1.04)
Total cancer	10.69 (1687)	11.22 (2661)	.10	0.95 (0.89 to 1.01)

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

†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.

‡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.

Low-Fat Dietary Pattern Intervention Effects on Breast and Ovarian Cancer, in Relation to Baseline 4-Day Food Record % of Energy from Fat

% of Energy from Fat (4DFR)	Mean (SD) Difference Between Groups	Hazard Ratio (95% CI)	Interaction P-Value
Breast Cancer (1727 cases)			
< 27.9	9.7 (6.2)	0.97 (0.79, 1.20)	
27.9 - 32.3	10.4 (6.5)	1.08 (0.89, 1.30)	0.04
32.3 - 36.8	11.7 (6.6)	0.85 (0.70, 1.03)	
≥ 36.8	12.2 (7.0)	0.78 (0.64, 0.95)	
Ovary Cancer (160 cases)			
< 28.7		1.33 (0.76, 2.33)	
28.7 - 35.1		0.60 (0.32, 1.12)	0.05
≥ 35.1		0.58 (0.31, 1.08)	

Variations Among Participants in WHI Clinical Trial Intervention Effects

- Standard full cohort HR analyses in subsets of participants defined by demographic characteristics, personal habits, medical history
  - Timing hypothesis for hormone therapy
  - More breast cancer HR variations for E-alone than for E+P

Analyses of Variations in Hormone Therapy Effects

- Candidate risk factor biomarker studies for CHD, stroke, and VT (inflammatory markers, thrombosis and coagulation markers, lipids and lipoproteins, related genetic markers) and for breast cancer (circulating hormones and related binding proteins) in relation to hormone therapy HRs
  - Some interesting findings (e.g., women with high baseline LDL-C experienced a larger hormone therapy HR for CHD (Rossouw et al,2008).
  - Even though most biomarkers were strongly associated with disease risk, and many changed markedly following hormone therapy initiation, none appeared to mediate HT effects on these diseases.

Variations in Clinical Trial Intervention Effects *(continued)*

- Case-only analysis for categorical characteristics

  V=1 active, V=0 placebo

  z-baseline categorized variable taking values $z_1 \ldots z_k$

$$\lambda(t, V, z) = \lambda_{0z}(t) exp\left\{\sum_{i}^{k} \beta_i V \ I(z = z_i)\right\}$$

So that e^{β_i} is intervention HR at $z=z_i$.

$$logit(V = 1|X = t, z) = logit(V = 1|X \geq t, z) + \sum_1^k \beta_i I(z = z_i)$$

where X=t denotes disease occurrence at time t following randomization.

V⊥z by trial design, so if disease is rare, to a good approximation, logit(V=1|T≥t,z)=log{q/(1-q)}, where q=P(V=1)

Can estimate $\{e^{\beta_i}\}$ using ordinary logistic regression with `offset' log (q/1-q)}, with efficiency essentially equal to full cohort analysis.

Case-Only Analyses of CT Intervention Effects by Genotype

- High-dimensional SNP studies of WHI clinical trial effects on key clinical outcomes
  - 2166 breast cancer cases and matched controls through end of CT intervention period
  - 9039 SNPs from WHI pooled DNA GWAS and C-GEMS GWAS, including 17 disease associated SNPs based on other studies
  - Confirmed associations with FGFR2 region of 10q ($p<10^{-8}$), MRPS30 region of 5p ($p<0.001$), MAP3K1 region on 5q ($p<0.001$), and TOX3 region on 16q ($p<0.01$) (All FDR<0.05)
  - Case-only analyses of SNPs in these regions in relation to HRs for E+P, E-alone, and Dietary Modification among women in upper quartile of baseline % energy from fat

SNP Interactions with CT Intervention Effects

Odds ratio interactions with CT interventions

FGFR2			Odds Ratio (95% CI)			Test of Equality
	Major/		Number of Minor SNP Alleles			
SNP	Minor	MAF	0	1	2	*P*-value
Estrogen plus progestin (471 cases)						
rs2981582	G/A	0.38	1.08 (0.79, 1.48)	1.43 (1.09, 1.87)	1.43 (0.94, 2.18)	0.37
rs3750817	C/T	0.39	1.52 (1.14, 2.02)	1.33 (1.01, 1.75)	0.69 (0.41, 1.17)	0.033
Estrogen-alone (247 cases)						
rs2981582	G/A	0.38	0.51 (0.31, 0.84)	1.04 (0.73, 1.47)	0.61 (0.34, 1.07)	0.045
rs3750817	C/T	0.39	0.74 (0.51, 1.09)	0.99 (0.68, 1.44)	0.34 (0.15, 0.76)	0.046
Dietary Modification (baseline % energy from fat in upper quartile, 428 cases)						
rs2981582	G/A	0.38	0.51 (0.34, 0.77)	0.80 (0.61, 1.06)	1.04 (0.68, 1.59)	0.05
rs3750817	C/T	0.39	1.06 (0.80, 1.41)	0.53 (0.38, 0.74)	0.62 (0.33, 1.15)	0.005

Prentice et al, CEBP 2009, and CEBP 2010

Biomarkers as Objective Measures of Dietary Consumption and Physical Activity Patterns

- Interpretation of DM trial results
  - What are actual dietary differences between randomization groups?
  - Was a sufficiently powerful test implemented?
- What are the desirable next steps in nutritional and physical activity epidemiology research?

Mean (SD) of Nutrient Consumption by Randomization Group

	Year 1		Year 1	Year 3	Year 6
	Intervention	Control	Difference	Difference	Difference
Fat (% of calories)	24.3 (7.5)	35.1 (6.9)	-10.7* (7.0)	-9.5* (7.4)	-8.1* (7.8)
Total Fat (g)	40.8 (21.4)	63.0 (31.0)	-22.4* (31.1)	-20.1* (32.0)	-18.4* (33.5)
Energy (kcal)	1500 (544)	1593 (644)	-95.8* (616.2)	-92.5* (632.1)	-119.9* (662.9)

*Difference significant at $p<0.001$ from a two sample t-test

Nutrient and Physical Activity Biomarkers in the WHI

- 544 DM trial women completed two-week DLW protocol with urine and blood collection and with FFQ and other questionnaire data collection (50% intervention, 50% control). A 20% reliability subsample repeated protocol separated, by about 6 months from original data collection.
- Biomarker study among 450 women in the WHI Observational Study for calibrating baseline FFQ, 4DFR, and PA questions, and for evaluating measurement properties of prominent dietary and physical activity assessment approaches (frequencies, records, and recalls) and their combination.

Measurement Models for Nutritional Epidemiology

(Carroll, Freedman, Kaaks, Kipnis, Spiegelman, Rosner, Prentice…)

Recovery Biomarkers:

$$W_{biomarker} = Z + e$$
$$Q_{self\text{-}report} = a_0 + a_1 Z + a_2 V + a_3 ZV + (r + \varepsilon)$$

Can estimate odds ratios *(Sugar et al, 2007, Biometrics)*, or hazard ratios *(Shaw et al, 2007)*, corresponding to Z from cohort data on W and subcohort data on X.

Hazard Ratio Estimation

Under a joint normality assumption for (Z, r + e) given V, conditional expectation of Z of the form

$$E(Z \mid Q,V) = b_0 + b_1 Q + b_2 V + b_3 QV$$

and

$$E(Z \mid Q,V) = E(W \mid Q,V)$$

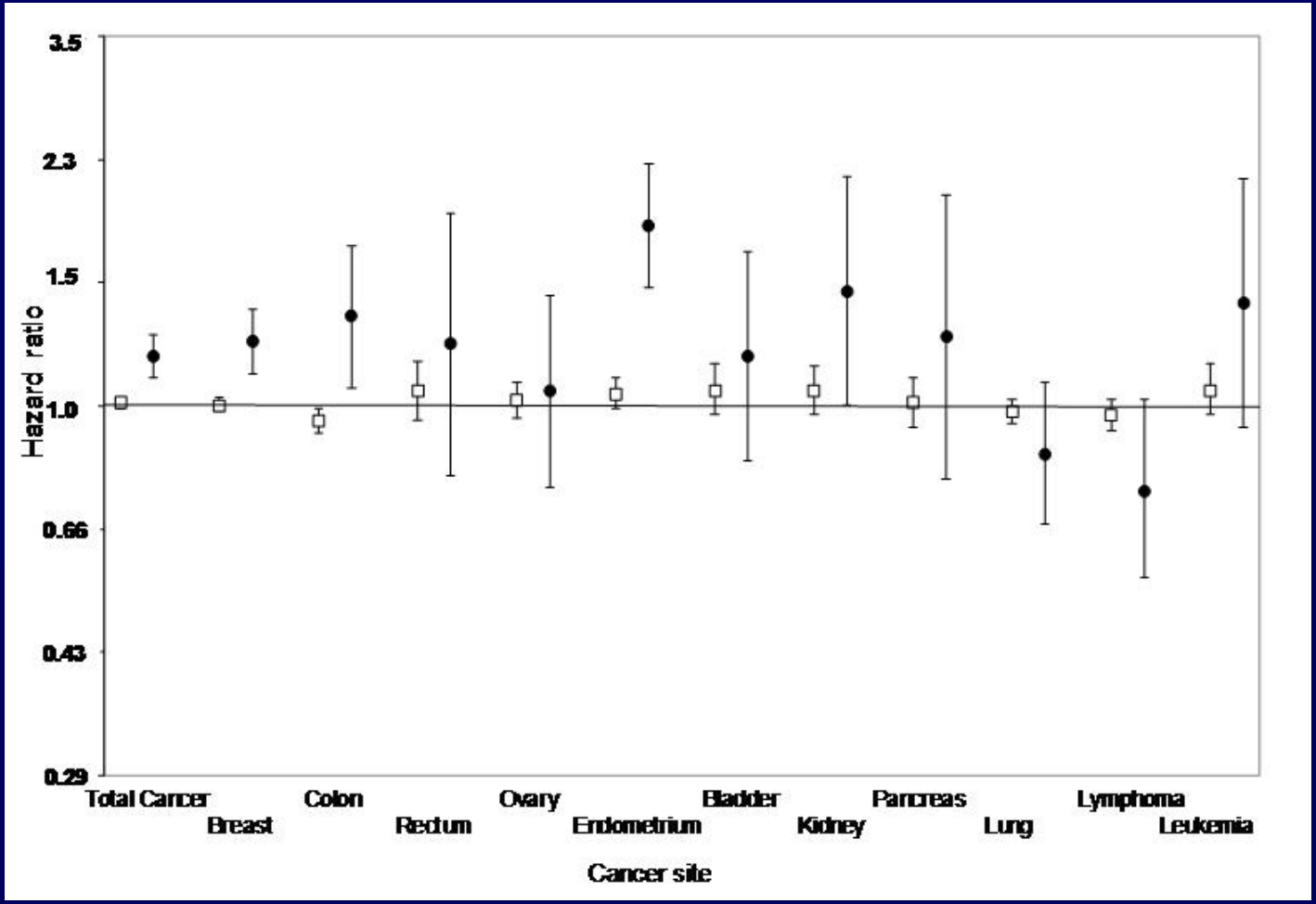
- Calibrated estimates of Z from linear regression of W on (Q,V) in the biomarker subsample
- Regression calibration estimation of hazard ratios by inserting calibrated consumption estimates in Cox regression, and using a bootstrap procedure for standard error estimation

Regression Calibration Coefficients for Log-Transformed Total Energy, Total Protein and Percent Energy from Protein

(Neuhouser et al, AJE, 2008)

Characteristic	Coefficient (SD) Log Total Energy	Coefficient (SD) Log Protein	Coefficient (SD) Log % Energy from Protein
Intercept	7.61 (0.13)	4.28 (0.024)	2.66 (0.01)
Log FFQ	**0.062 (0.018)**	**0.212 (0.032)**	**0.439 (0.058)**
BMI	0.013 (0.001)	0.012 (0.002)	-0.004 (0.002)
Age	-0.005 (0.001)	-0.008 (0.002)	-0.005 (0.002)
Black	-0.016 (0.017)	-0.130 (0.047)	__
Hispanic	-0.004 (00.30)	-0.021 (0.056)	__
Other race	-0.093 (0.027)	-0.100 (0.058)	__

Estimated Hazard Ratios and 95% Confidence Intervals for a 20% Increase in **Energy Consumption** from Combined Analysis of Data from the Women's Health Initiative Dietary Modification Trial Comparison Group and Observational Study , Without and With Biomarker Calibration of Consumption: *(Open box – uncalibrated, Black circle – calibrated) (Prentice et al, 2009, AJE)*



Biomarker Development for Other Nutrients/Dietary Components

Human Feeding Study for Biomarker Development

- Provide all food and drink over a two-week feeding period
- Use blood and urine nutritional measures and study subject characteristics to explain variation in provided nutrient consumption
- Use a highly individualized diet that aims to approximate usual diet so that blood and urine nutritional measures will stabilize quickly, and to preserve nutrient consumption variation in the study cohort

Statistical model: $W = c_0 + c_1X + c_2V + e_W$

W is log-provided nutrient

X is comprised of log-pertinent urine or blood nutritional measures

V is a vector of study subject characteristics

- A potentially useful biomarker should be able to explain substantial variation in W (e.g., ≥ 50%), and without obvious omissions to (X,V)
- Feeding study among 150 WHI women in Seattle initiated 7/1/10

Development and Testing of New Preventive Interventions

- Newer forms of high-dimensional biologic data have potential to add to screening and initial testing of preventive interventions

Intermediate Outcome Trials Having High-Dimensional Responses

- Evaluate impact of candidate preventive interventions on high-dimensional response (e.g., plasma proteome)
- Develop knowledge base to relate high-dimensional response to risk of a broad range of clinical outcomes
- Predict intervention effects on clinical outcomes of interest, from high-dimensional response (including established disease risk factors), to help determine whether a full-scale intervention trial is merited

Hormone Therapy Proteomics Project (collaboration with Sam Hanash, FHCRC)

- 50 E-alone women; 50 E+P women

  Compare baseline to 1-year serum proteome

  in pools of size 10.

- 800 cases of each of breast cancer, CHD and stroke, and 1-1 matched controls

  Compare case versus control plasma proteome

  in pools of size 100

Hormone Therapy Proteomics Project

- Following the application of strict criteria for protein identification, 378 proteins were evaluated for change with E+P or E-alone
- Of these, a remarkable 44.7% (169/378) had evidence of change ($p<0.05$) with E+P and/or E-alone.
- Altered proteins were in multiple biological pathways relevant to observed clinical effects: coagulation/inflammation, immune function, cell adhesion, osteogenesis, growth factors,… (Katayama et al, 2009; Pitteri et al, 2009, Genome Medicine).

Case-control Plasma Proteome Comparisons

800 cases and 1-1 matched controls for each of CHD, stroke and breast cancer (using pools of size 100).

- CHD 37 with $p<0.05$ vs. 17.3 by chance
- Stroke 47 with $p<0.05$ vs. 18.3 by chance
- Most proteins having small FDRs for disease association were affected by hormone therapy, and provide novel candidates to explain HT effects on these diseases (Prentice et al, 2010 Genome Medicine)

ELISA Replication in HT trials

- Beta-2 microglobulin related to CHD with FDR <0.05, confirmed in HT cohort
- OR (95% CI) for 16% increase in B2M of 1.30 (1.11, 1.54)
- IGFBP4 related to stroke with FDR<0.05, and confirmed in HT cohort
- OR (95% CI) for 20% increase in IGFBP4 of 1.40 (1.06, 1.85)

Biomarkers and Clinical Trial Mediation Analyses
(joint work with doctoral student Shanshan Zhao)

- Examine the impact of including post-randomization biomarker change on treatment effect parameters.

  V = 1 active, V = 0 control

  z – change from baseline to post-intervention initiation as potential mediator

- Here V and z may be strongly positively correlated.
- If z is measured with (technical) error, or if most temporal variation in z in the control group is not relevant to disease risk, then regression analysis will tend to put ‘weight’ on V rather than z, and the biomarker change will appear not to mediate the intervention effect.

Measurement Error and Mediation Analysis: Quantitative Response Special Case

x_0 – (underlying risk-relevant) baseline biomarker value

x_1 – corresponding post-intervention biomarker value

V = 1 intervention, V = 0 control

Y – quantitative response

Suppose (x_0, x_1) fully mediates effect of V on Y

E (Y | x_0, x_1, V) = a + a_0x_0 + a_1x_1 + (0)V, and $a_1 \neq 0$

Also suppose

$$\begin{pmatrix} x_0 \\ x_1 \end{pmatrix} | V \sim BVN \left\{ \begin{pmatrix} \mu_0 \\ \mu_1 + dV \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{pmatrix} \right\}$$

from which $E(Y|x_0, V) = a' + a_0'x_0 + (a_1 d)V$

Mediation analysis compares estimate of a_1d to estimate of 0.

Measurement Error and Mediation Analysis

Now suppose x_0 and x_1 incorporate classical measurement error giving measurements

$$z_0 = x_0 + e_0 \text{ and } z_1 = x_1 + e_1, e_0 \perp e_1, \text{var } e_0 = var\ e_1 = \sigma_e^2$$

Then

$$E(Y|z_0, z_1, V) = b + b_0 z_0 + b_1 z_1 + b_2 V \text{ and } E(Y|z_0, V) = b' + b_0' z_0 + b_2' V$$

where $b_2' = a_1 d$ (not affected by measurement error)
and

$$b_2 = a_0 d k^{-1} \rho \delta^2 + a_1 d\{1 - k^{-1}(1 + \delta^2 - \rho^2)\}$$

For specified δ^2 = var e / var x, $b_2 \rightarrow (a_0 + a_1) d / (2 + \delta^2)$ as $\rho \rightarrow 1$, which may be very far from zero, even if δ^2 is small.

Can valid mediation assessment be recovered through measurement error correction?

IGFBP4 and Mediation of Hormone Therapy Effects on Stroke Risk

- Baseline and 1-year post-randomization IGFBP4 measures from 122 stroke cases in E-alone trial and 136 stroke cases in E+P trial and 1-1 matched controls
- IGFBP4 values found to increase by about 20% in treatment versus placebo group, for both E-alone and E+P
- Baseline IGFBP4 associated with stroke risk

	Odds Ratios (95% CI) for Hormone Therapy	
	E-alone	E+P
Adjusted for baseline IGFBP4	1.59 (0.95-2.67)	1.42 (0.87-2.32)
Adjusted for baseline and 1-yr IGFBP4	1.58 (0.94-2.66)	1.39 (0.85-2.27)

- No evidence of mediation !

Hormone Therapy Odds Ratios Following Regression Calibration Correction for Measurement Error in IGFBP4

E-alone vs. Placebo					
δ^2 (var e/var x)	0.1	0.2	0.3	0.4	0.5
Odds Ratios	1.57	1.55	1.52	1.45	1.01
95% CI	0.93 – 2.64	0.92 – 2.61	0.90 – 2.58	0.83 – 2.52	0.39 – 3.02
E+P vs. Placebo					
δ^2	0.4	0.5	0.6	0.7	0.75
Odds Ratios	1.37	1.36	1.34	1.29	0.98
95% CI	0.82 – 2.27	0.80 – 2.42	0.74 – 2.43	0.49 – 3.36	0.09 – 13.70

- Difficult to assess mediation via 'noisy' variables
- Requires large numbers of disease events; biomarker data at more than 2 time points for parameter identification; careful modeling and interpretation
- Similar issues if treatment/exposure is noisy, while potential mediator is comparatively stable (e.g., energy consumption and BMI)

Summary

- Biomarkers have much potential to elucidate clinical trial findings; further development is needed on biomarker mediation methods.
- Biomarkers have potential to correct self-report data for difficult-to-measure dietary and physical activity exposures, and to strengthen related epidemiologic research.
- Biomarkers have potential to invigorate the prevention intervention development enterprise.

Hormone Therapy Odds Ratios Following Regression Calibration Correction for Measurement Error in IGFBP4

[Replace (x_0, x_1) by $E\{(x_0, x_1)|(z_0, z_1), V\}$]

Correlation (p) between x_0 and x_1	E-Alone Odds Ratio	E+P Odds Ratio
0.7	1.57	1.37
0.8	1.54	1.37
0.9	1.47	1.34
0.95	1.34	1.30
0.96	1.28	1.28
0.97	1.18	1.25
0.98	1.01	1.18
0.99	0.62	1.01

Careful modeling and additional data needed to interpret mediation analysis, if intervention produces a nearly constant shift in potential mediator.

Mediation and Hazard Ratio Analysis

$$\lambda(t; x_0, x_1, V) = \lambda_0(t)\exp(a_0x_0 + a_1x_1 + a_2V + a_3x_0V + a_1x_1V)$$

Under rare disease and classical measurement error gives

$$\delta(t, z_0, z_1, V) = \delta_0'(t)\exp(b_0z_0 + b_1z_1 + b_2V + b_3z_0V + b_4z_1V)$$

to an excellent approximation, with 1-1 correspondence between {a} and {b}.

Gives measurement error corrected estimate of (a_1, . . . a_4), and corrected estimate of (a_0, a_1,a_2) conditional on a_3=a_4=0. From simulations, it appears possible to remove most bias from estimation of a_2, but additional data typically needed to assess correlation between x_0 and x_1.