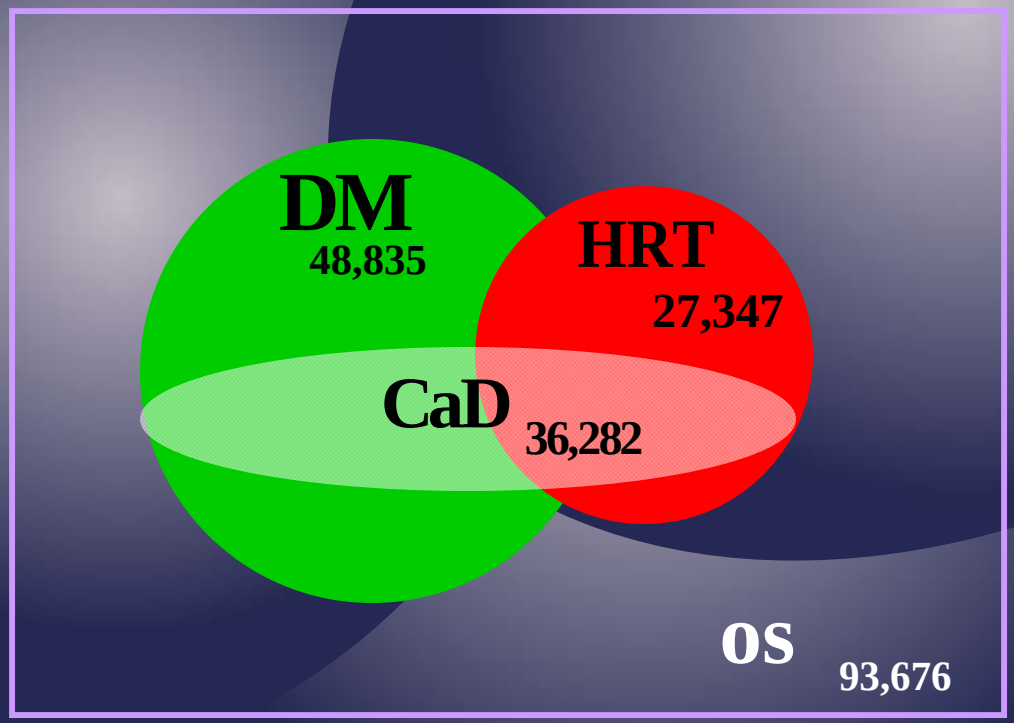
Measurement Error in Diet and Physical Activity Epidemiology

Ross Prentice, October 22, 2015

Design of WHI



CT=68,132

WHI =161,808

History: WHI Development and Research Program Phases

Pre-WHI (1984 – 1992)

- NCI-sponsored Women's Health Trial – pilot and feasibilities trial toward a full-scale trial of a low-fat diet intervention for breast cancer prevention
- Observational studies of postmenopausal estrogens (e.g., Premarin) and estrogens plus progestin (e.g., Prempro) by epidemiologic and clinical communities; NHLBI-sponsored intermediate outcomes trial- PEPI

WHI announced by NIH Director Bernedine Healy (1992) as a trans-NIH initiative.

- Contract for Clinical Coordinating Center (CCC) in 1992 and for the initial 16 of 40 Clinical Centers in 1993
- IOM Review in 1994
- WHI allowed to proceed; program office moved to NHLBI

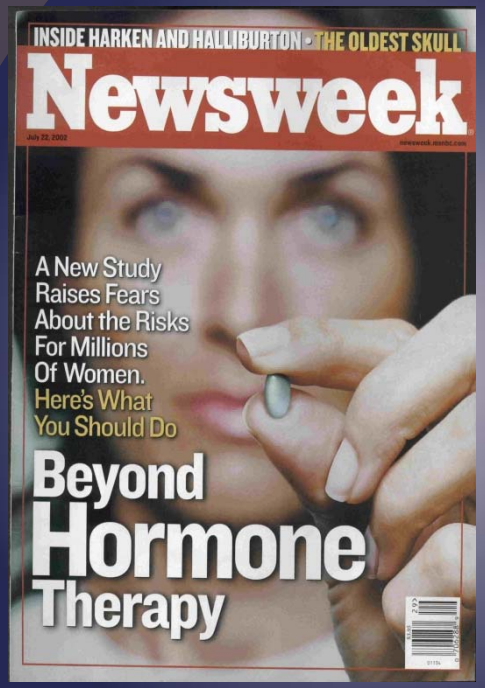
Women's Health Initiative Clinical Centers

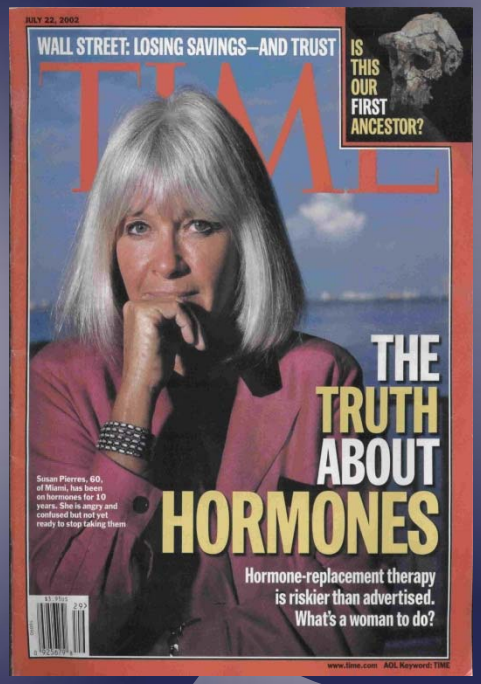
Intervention Phase (1993 – 2005)

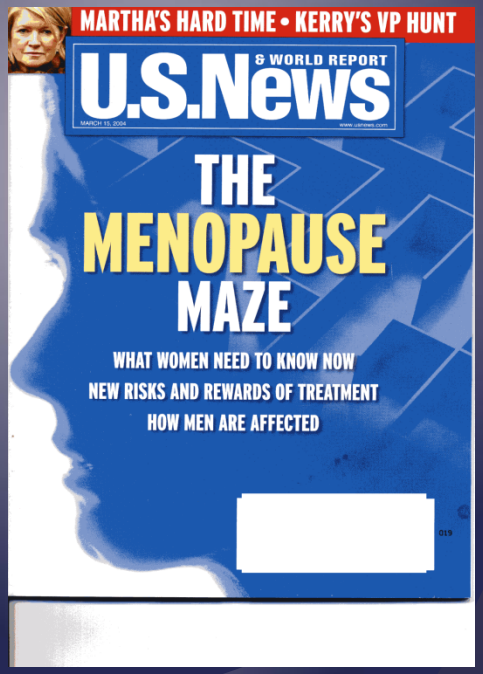
- Recruitment, protocol and procedures, intervention implementation
- Data and specimen collection; quality assurance; data and safety monitoring
- Complex organization and committee structure(s)

Intervention Phase (1993 – 2005)

- E+P trial stopped early (2002)

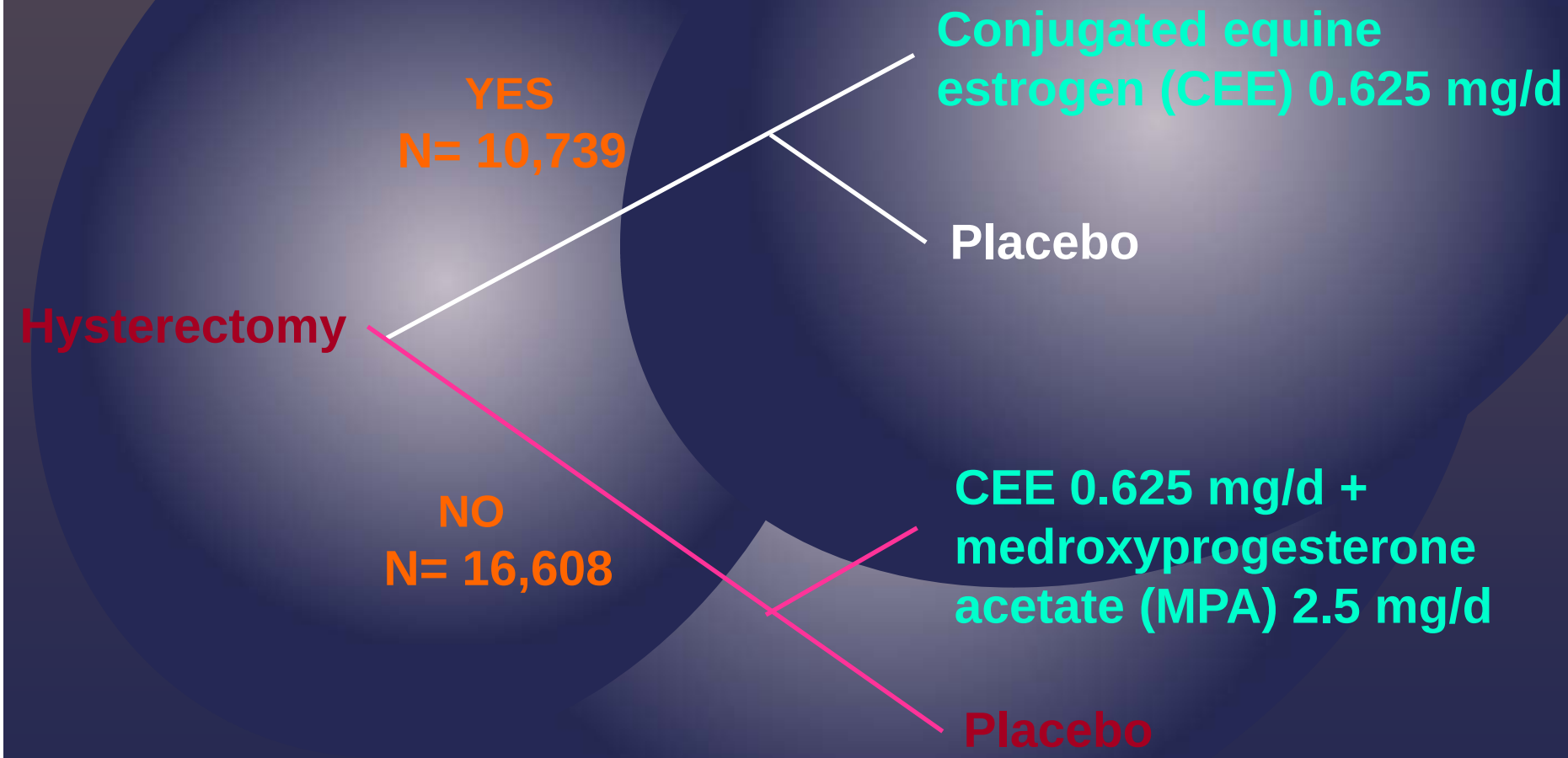






- E-alone trial stopped early (2004)
- DM and CaD interventions concluded at planned termination (2005)

Postmenopausal Hormone Therapy



Clinical Outcomes in the WHI Postmenopausal Hormone Therapy Trials

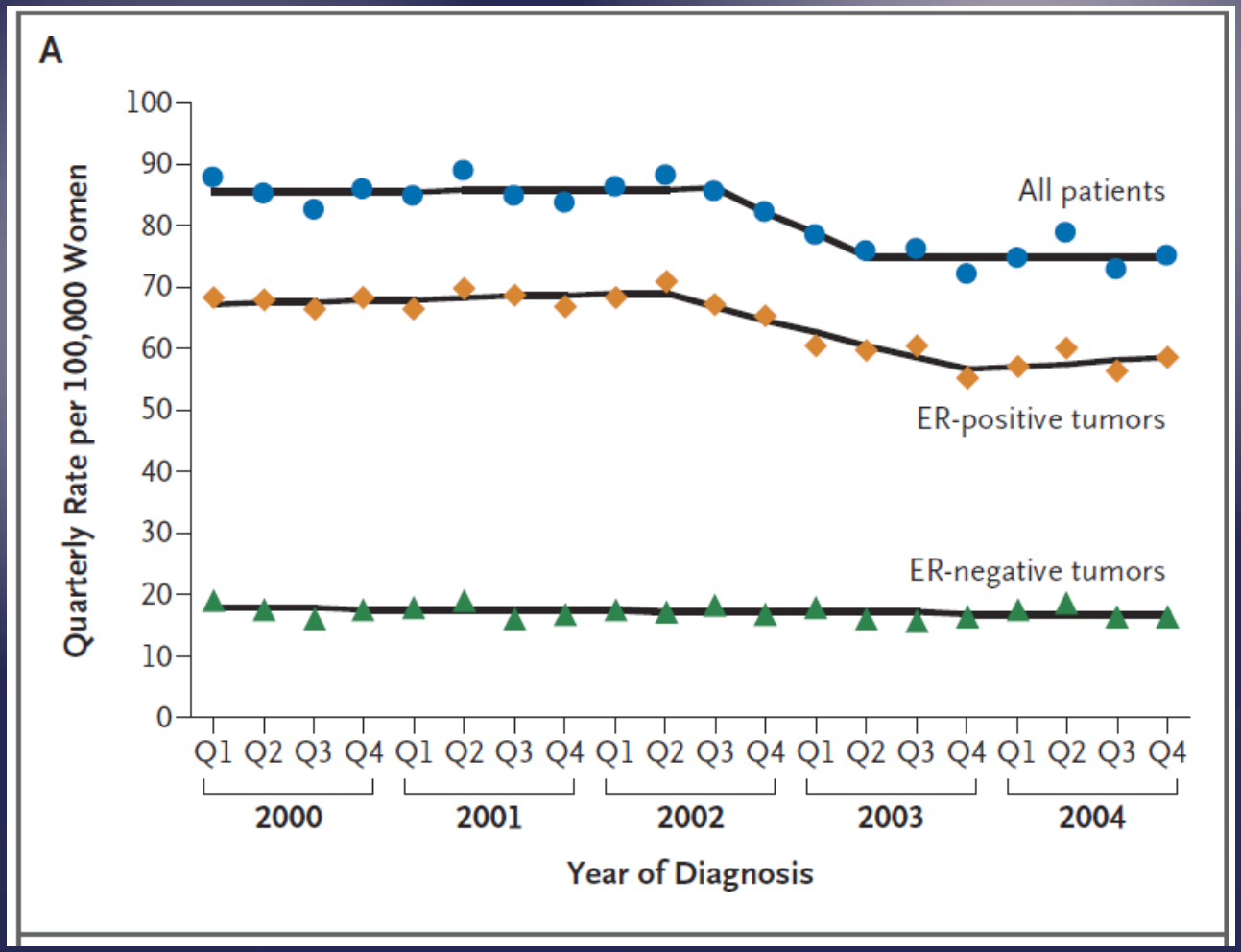
(WHI Study Group, JAMA 2002; Anderson et al, JAMA 2004)

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)

Quarterly Incidence of Breast Cancer in Women between the Ages of 50 and 69 Years, According to Estrogen-Receptor (ER) Status (2000-2004)

Data are from the NCI's SEER registries.

(Ravdin et al, NEJM, 2007)



Low-Fat Dietary Pattern

Support from international correlation analyses,

rodent feeding studies; and case-control studies concerning

dietary fat and breast cancer, but cohort studies

were generally unsupportive

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

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

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)	

Some Observations

1. Chronic disease rates tend to be highly disparate around the world, with many disease rates highly elevated in Western populations.
2. Migrant populations tend to assume rates that prevail in their new environment, within a few generations.
3. Risk prediction models, including those incorporating findings from recent GWAS, do not allow one to identify persons who will develop a specific disease, with even moderate precision.

Implication: There is still much to be learned about the determinants of chronic disease risk, including identification of important modifiable risk factors.

Where to Look?

1. Changes in incidence rates among migrants, and time trends in incidence rates, suggest that chronic disease risk depends importantly on common/ordinary habits and exposures.
2. Genetic/genomic factors may also be importantly involved in chronic disease pathogenesis, but perhaps more often as mediators than as primary 'exposures'.
3. Diet and physical activity patterns continue as likely sources of disease rate variations, but few clear associations have emerged to date, especially from nutritional epidemiology.

Measurement Error in Nutritional Epidemiology

- Cohort studies have provided the mainstay epidemiologic approach to nutritional epidemiology in recent years.
- These and other observational study designs need to contend with the usual confounding and outcome ascertainment bias issues, but also with measurement issues in dietary assessment, that may be dominant.
- Consider a hazard rate h(t;Z,V) at cohort follow-up time t, given a baseline dietary variable Z, and other potential risk factors V. A Cox model specifies

$$h(t;Z,V) = h(t) \exp(Zb + Va).$$

Measurement Error *(continued)*

- Because of measurement error, Z can be assessed only indirectly, by means of a variable Q (e.g., an FFQ assessment of Z). The hazard function that can be estimated h(t ;Q,V) is typically (rare disease) well approximated by

$$h(t;Q,V) = h^*(t) \exp\{E(Z;Q,V)b + Va\}.$$

- Hence, to address measurement error in assessment of Z, one needs to appropriately estimate E(Z;Q,V).
- Objective markers (W) adhering to a classical measurement model W=Z+u are crucial for such estimation.

Regression Calibration

- A classical measurement model entails W=Z+u with mean zero noise ‘u’ that is independent of the targeted Z, independent of all pertinent study subject variables V and, importantly, independent of the measurement error associated with the self-report assessment Q, in which case

$$E(W;Q,V)= E(Z;Q,V)$$

- And E(Z; Q,V) can be estimated by regression of the biomarker W on the self-report Q and study subject characteristics V (confounders; factors involved in systematic bias in Q; or that may help to explain the variation in W, or Z, in the study population).

Nutrient and Physical Activity Biomarkers Studies in 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 about 6 months later. (NBS; 2004-2006)
- Biomarker study among 450 women in the WHI Observational Study for evaluating and comparing measurement properties of dietary and physical activity assessment approaches (frequencies, records, and recalls); with 20% reliability subsample. (NPAAS; 2007-2009)
- Recently completed feeding study among 153 WHI women in Seattle, for development of objective markers for additional nutrients or foods. (NPAAS II; 2010-present)

Nutrition and Physical Activity Biomarker Study Participants (NBS and NPAAS)

Coordination (FHCRC):

Marian Neuhouser, Lesley Tinker, Johanna Lampe, Ross Prentice

Clinical Center PIs:

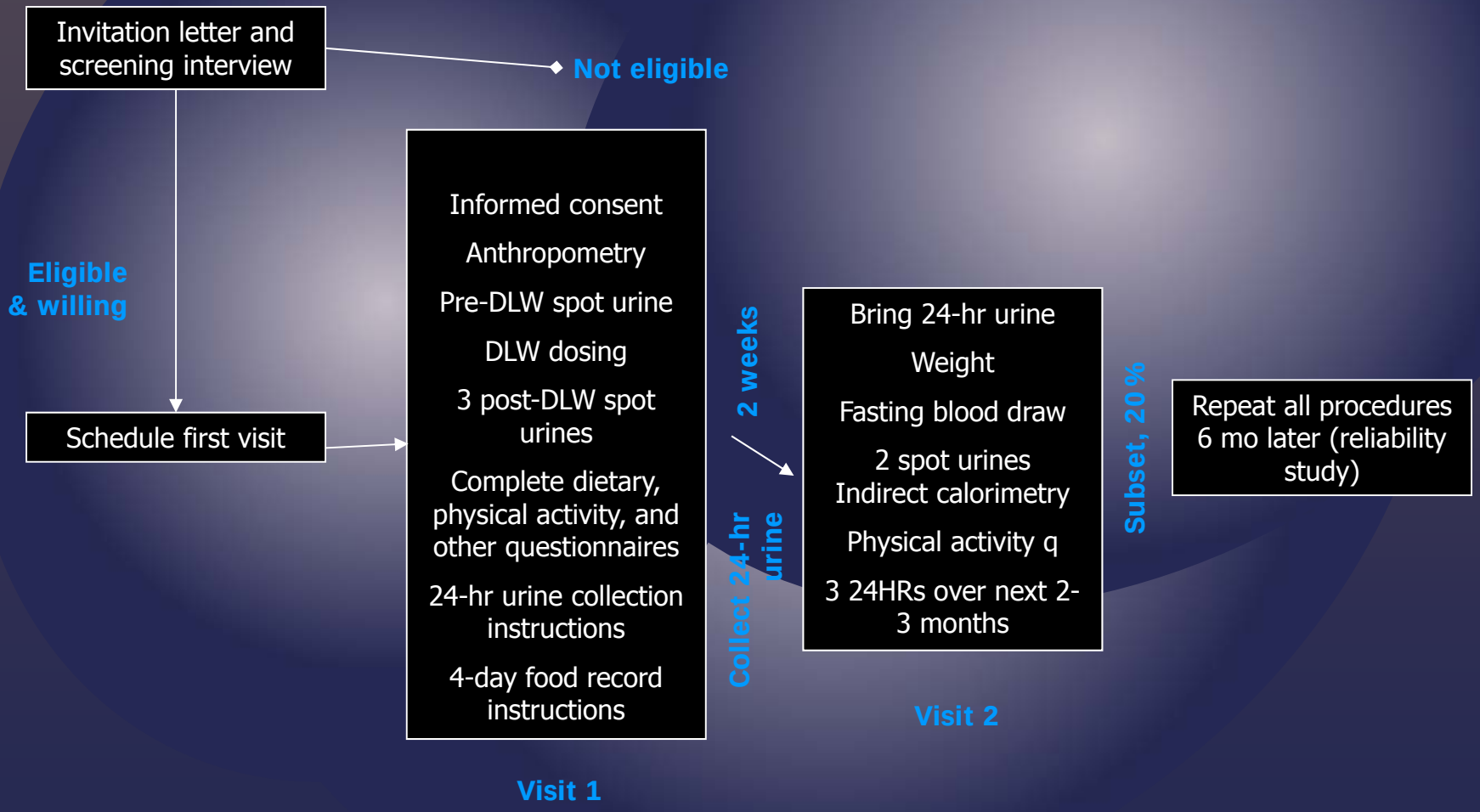
Shirley Beresford - Seattle
Bette Caan - Oakland
Linda Van Horn - Chicago
Cynthia Thomson - Arizona
Yasmin Mossavar-Rahmani† - NYC
Karen Johnson - Memphis
Gloria Sarto - Wisconsin
Judy Ockene - U Mass
Gerardo Heiss - UNC
Lewis Kuller* - Pittsburgh
Marcia Stefanick* - Stanford
Ellen Smit* - Portland
Annlouise Assaf* - Pawtucket

*NBS only †NPAAS only

Additional Key Collaborators:

Dale Schoeller - Wisconsin (DLW); Sheila Bingham – Cambridge (UN); Xiaoling Song, Ying Huang, Chongzhi Di – FHCRC; Cheng Zheng - Milwaukee; Pamela Shaw - U Penn; Jeannette Beasley - Einstein; Dan Raftery – University of Washington

WHI Nutritional Biomarkers Study Procedures (NBS and NPAAS I)



Measurement Model

$$W_{biomarker} = Z + e$$

$$Q_{self\text{-}report} = a_0 + a_1 Z + a_2 V + \varepsilon$$

Estimate disease odds ratios *(Sugar, Wang and Prentice 2007, Biometrics)*, or hazard ratios *(Shaw and Prentice, 2012, Biometrics)*, using Q from cohort and W from sub-cohort

Calibration and Hazard Ratio Estimation

Under a joint normality assumption for (Z, Ɛ) given V, the conditional expectation of Z is of the form

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

and

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

- Calibrated estimates of Z from linear regression of W on (Q,V) in the biomarker subsample
- Hazard ratios estimation by inserting calibrated consumption estimates into Cox regression, and using a bootstrap procedure for standard error estimation. Some important nuances if some elements of V (e.g., BMI) are not included in disease risk model *(Prentice and Huang, 2011, Can J Stat; Zheng et al, 2014, AJE).*

Calibration Equation Coefficients (β), Standard Errors (SE), and Percent of Biomarker Variation Explained (R^2) from Regression of Log(biomarker) on Log(self-report), and Other Factors among 450 Observational Study Women

Energy

Variable	Food Frequency				4DFR				24HR			
	β	SE	R^2	Adj R^2	β	SE	R^2	Adj R^2	β	SE	R^2	Adj R^2
						ENERGY						
Intercept	7.614[a]	0.009			7.597[a]	0.009			7.607[a]	0.009		
FFQ	0.054[a]	0.017	3.8	6.5								
4DFR					0.161[a]	0.028	7.8	13.3				
24HR									0.101[a]	0.026	2.8	4.8
BMI	0.013[a]	0.001	26.9	45.9	0.013[a]	0.001	27.0	46.0	0.013[a]	0.001	28.7	48.9
Age	-0.010[a]	0.001	9.7	16.5	-0.009[a]	0.001	8.4	14.3	-0.009[a]	0.001	9.1	15.5
Black	-0.023	0.019			-0.024	0.018			-0.024	0.018		
Hispanic	-0.062[a]	0.021	1.3	2.2	-0.065[a]	0.020	1.5	2.6	-0.063[a]	0.020	1.5	2.6
Other minority	-0.041	0.040			-0.039	0.038			-0.038	0.039		
(Total)[b]			41.7	71.1			44.7	76.2			42.1	71.8

Prentice et al (2011, AJE)

Calibration Equation Coefficients (β), Standard Errors (SE), and Percent of Biomarker Variation Explained (R^2) from Regression of Log(biomarker) on Log(self-report), and Other Factors among 450 Observational Study Women

Protein

Variable	Food Frequency				4DFR				24HR			
	β	SE	R^2	Adj R^2	β	SE	R^2	Adj R^2	β	SE	R^2	Adj R^2
Intercept	4.263	0.017			4.235	0.016			4.269	0.016		
FFQ	0.135[a]	0.021	**8.4**	16.4								
4DFR					0.465[a]	0.045	**22.6**	44.2				
24HR									0.404[a]	0.046	**16.2**	31.7
BMI	0.012[a]	0.002	5.8	11.4	0.012[a]	0.002	5.1	10.0	0.012[a]	0.002	5.8	11.4
Age	-0.012[a]	0.002	4.1	8.0	-0.009[a]	0.002	2.2	4.3	-0.011[a]	0.002	3.4	6.7
Black	-0.120[a]	0.038			-0.138[a]	0.034			-0.145[a]	0.035		
Hispanic	-0.078	0.040	2.0	3.9	-0.067	0.036	2.7	5.3	-0.069	0.037	3.0	5.9
Other minority	-0.018	0.076			0.012	0.070			-0.026	0.072		
(Total)[b]			**20.3**	39.7			**32.7**	63.8			**28.4**	55.6

Prentice et al (2011,AJE)

Other Regression Calibration Activities Using NBS/NPAAS I Data

- Activity-Related Energy Expenditure (AREE)
  Objective measure: DLW total energy expenditure
  minus resting energy expenditure from indirect calorimetry
  Self-report: AAFQ, 7dPAR, PHQ+ *(Neuhouser et al, 2013, AJE)*
- Sodium and Potassium and their Ratio
  Objective measure: 24-hour urinary recovery
  Self-report: FFQ, 4DFR, 24HR for these nutrients
  (Huang et al, 2014, Hypertension)
- Total Sugars
  Objective measure: 24-hour urinary recovery of fructose and sucrose
  Self-report: FFQ, 4DFR, 24HR total sugars
  (Tasevska et al, 2014, CEBP)
- Respiratory Quotient
  Objective measure: RQ from indirect calorimetry
  Self-report: FQ = 0.7 (% energy from F+A) + 0.8 (% energy from protein) + 1.0 (% energy from carbohydrate)
  (Prentice et al, 2013, CEBP)

Regression calibration β-coefficients and percent of biomarker variation explained in the WHI NPAAS *(Neuhouser et al, 2013, AJE)*

	AAFQ			PAR			PHQ		
			Adjusted			Adjusted			Adjusted
Variable	β	R^2	R^2	β	R^2	R^2	β	R^2	R^2
Intercept	4.789*			5.353*			5.632*		
Log (self-report AREE)	0.273*	**7.6**	24.0	0.184*	**4.8**	15.1	0.153*	**3.4**	10.7
Age	-0.018*	9.1	28.7	-0.024*	8.6	27.1	-0.022*	8.5	26.8
Race		2.4	7.6		2.3	7.2		2.9	9.1
Black	0.086			0.075			0.095		
Hispanic	0.108*			0.141*			0.055		
BMI	0.015*	6.1	19.2	0.012*	5.7	18.0	0.015*	7.0	22.1
TOTAL		25.2	79.4		21.5	67.8		21.8	68.7

* Regression coefficient differs from zero at $P<0.05$

- Adjusted R^2 is corrected for biomarker measurement error, i.e., R^2/log(biomarker) correlation from reliability study

Estimated Hazard Ratios (95% Confidence Intervals) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Various Cardiovascular Disease Categories in the Women's Health Initiative Observational Study (OS) from Baseline (1994-1998) Through September 30, 2010 *(Zheng et al, 2014, AJE)*

	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
Disease Category	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Total CHD	1.00	0.98,1.02	0.99	0.97,1.01	**1.57**	1.19,2.06	**0.78**	0.65,0.95
Nonfatal MI	1.00	0.98,1.03	0.99	0.97,1.01	1.49	1.13,1.97	0.80	0.67,0.97
Coronary Death	0.97	0.94,1.02	0.97	0.94,1.00	2.22	1.36,3.61	0.63	0.46,0.86
Congestive Heart Failure	1.04	1.01,1.08	0.97	0.95,1.00	**3.51**	2.12,5.82	**0.57**	0.41,0.79
CABG and PCI	1.01	0.99,1.03	1.01	0.99,1.03	1.43	1.19,1.70	0.90	0.79,1.03
Total Stroke	0.97	0.95,1.00	0.99	0.98,1.01	1.36	1.05,1.76	0.83	0.69,0.99
Ischemic Stroke	0.98	0.96,1.01	0.99	0.97,1.01	1.55	1.14,2.10	0.78	0.64,0.94
Hemorrhagic Stroke	0.94	0.88,0.99	1.03	0.99,1.08	0.47	0.21,1.07	1.37	0.85,2.20
Total CVD: CHD and Stroke	0.99	0.97,1.00	0.99	0.98,1.00	**1.49**	1.18,1.88	**0.80**	0.69,0.92
Total CVD including CABG and PCI	1.00	0.99,1.01	1.00	0.99,1.01	**1.49**	1.23,1.81	**0.83**	0.73,0.93

Estimated Hazard Ratio (95% Confidence Interval) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Various Cancer Categories, in the Women's Health Initiative Observational Study (OS) from baseline (1994-1998) Through September 30, 2010 *(Zheng et al, 2014, AJE)*

	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
Cancer Category	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Total Invasive Cancer	1.01	1.00,1.02	0.99	0.99,1.00	**1.43**	1.17,1.73	**0.84**	0.73,0.96
Obesity-related Cancer	1.02	1.00,1.03	1.00	0.99,1.01	**1.71**	1.33,2.21	**0.79**	0.65,0.94
Breast Cancer	1.01	0.99,1.02	1.00	0.99,1.01	**1.47**	1.18,1.84	**0.82**	0.71,0.96
Colon Cancer	1.00	0.96,1.03	1.00	0.97,1.03	**1.86**	1.18,2.93	0.83	0.66,1.04
Rectum Cancer	1.01	0.92,1.10	0.99	0.93,1.05	**2.75**	1.10,6.83	**0.51**	0.27,0.99
Ovary Cancer	1.00	0.95,1.05	1.01	0.98,1.05	0.85	0.43,1.68	1.12	0.73,1.71
Endometrial Cancer	1.08	1.04,1.12	1.01	0.98,1.05	**2.72**	1.44,5.13	0.77	0.49,1.21
Bladder Cancer	1.03	0.97,1.10	0.96	0.92,1.00	1.80	0.88,3.69	0.68	0.42,1.09
Kidney Cancer	1.05	0.98,1.12	1.02	0.96,1.07	**2.94**	1.37,6.28	0.62	0.35,1.12
Pancreas Cancer	0.95	0.89,1.01	0.97	0.92,1.01	2.06	0.98,4.33	**0.61**	0.37,1.00
Lung Cancer	0.99	0.96,1.01	0.97	0.95,1.00	1.14	0.74,1.76	0.79	0.60,1.03
Lymphoma	1.08	1.03,1.13	1.00	0.96,1.03	0.99	0.48,2.07	1.16	0.69,1.94
Leukemia	1.01	0.95,1.07	0.98	0.93,1.02	1.48	0.70,3.12	0.74	0.47,1.18

Estimated Hazard Ratio (95% Confidence Interval) for 20% Increments in Total Energy (TE) Consumption and in Activity-related Energy Expenditure (AREE), With and Without Calibration to Correct for Measurement Error, for Diabetes Incidence, in the Women's Health Initiative Observational Study (OS), from Baseline (1994-1998) Through September 17, 2012 *(Zheng et al, 2014, AJE)*

	Uncalibrated				Calibrated			
	Energy		AREE		Energy		AREE	
Outcome Category	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
Diabetes Mellitus	1.06	1.04,1.07	1.01	1.00,1.02	**4.17**	2.68,6.49	**0.60**	0.44,0.83

Biomarker Development for Additional Nutrients/Dietary Components

Human Feeding Study for Biomarker Development (NPAAS II)

- 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 is log-provided nutrient.
  X is comprised of log- urine or blood nutritional measures (including metabolomic profile assessments - Raftery).
  V is a vector of study subject characteristics.
- A potentially useful biomarker should be highly correlated with W (e.g., correlation ≥ 50%), and avoid important omissions to (X,V).
- Feeding study among 153 WHI women in Seattle completed in 2013.

Raftery Lab Platforms

Quantitated metabolites generated from the NPAAS				
Batch	Method	Specimen	Metabolites Detected	CV*
1	Targeted LC-MS/MS	Serum	131 aqueous metabolites	7.9%
1	Global LC-QTOF-MS	Serum	494 lipids	6.7%
1	NMR	Urine	48 metabolites	3.9%
1	Global GC-MS	Urine	122 metabolites	15%
2	Targeted LC-MS/MS	Serum	105 aqueous metabolites	8.0%
2	Global LC-QTOF-MS	Serum	231 lipids	15.7%
2	NMR	24hr Urine	53 metabolites	5.0%
2	NMR	Spot Urine	53 metabolites	1.6%

*** Median CV for blinded duplicate samples. Final Batch 2 GCMS reliability data still under development.**

Correlations of Provided Intake with Metabolomic Profile Estimated Intake for Several Nutritional Variables in NBAAS II

Correlation* between mean daily provided (consumed) nutrient intake and corresponding metabolite-based cross-validated estimated intake for specific nutrition variables.**

Nutritional Variable		Blood+24 hr urine		Blood Only		24-hr urine only		Blood +spot urine	Spot urine only
	IQR intake ranges	Batch 1	Batch 2	Batch 1	Batch 2	Batch 1	Batch 2	Batch 2	Batch 2
Energy, kcal	1744-2084	0.7	0.51	0.72	0.51	0.56	0.46	0.56	0.31
Protein, g	68.7-89.9	0.63	0.55	0.59	0.6	0.46	0.46	0.61	0.4
Nitrogen, g	11.1-14.7	0.63	0.56	0.57	0.6	0.47	0.47	0.63	0.38
Fat, g	70.3-89.4	0.68	0.6	0.65	0.57	0.62	0.29	0.7	0.32
Saturated fat, g	20.6-31.6	0.79	0.56	0.7	0.08	0.61	0.25	0.57	0.24
Monounsaturated fat, g	26.0-35.0	0.71	0.52	0.69	0.5	0.63	0.38	0.54	0.39
Polyunsaturated fat, g	13.1-19.6	0.59	0.72	0.6	0.61	0.49	0.6	0.58	0.4
Linoleic acid,g	10.8-17.0	0.58	0.68	0.63	0.69	0.49	0.59	0.59	0.42
Linolenic acid, g	1.4-2.3	0.67	0.48	0.65	0.45	0.56	0.48	0.48	0.33
EPA, g	0.04-0.16	0.77	0.65	0.65	0.69	0.63	0.54	0.72	0.43
DHA, g	0.08-0.25	0.8	0.64	0.75	0.64	0.61	0.39	0.68	0.42
Carbohydrate, g	188.1-241.9	0.79	0.6	0.64	0.6	0.73	0.57	0.59	0.41
Added sugars, g	34.9-63.0	0.69	0.57	0.49	0.52	0.7	0.45	0.6	0.37
Soluble fiber, g	6.1-9.3	0.49	0.66	0.53	0.61	0.47	0.64	0.67	0.5
Insoluble fiber, g	13.6-20.0	0.55	0.53	0.6	0.46	0.58	0.5	0.49	0.31
Alcohol, g	0.25-15.9	0.64	0.63	0.71	0.54	0.62	0.54	0.51	0.38
Caffeine, mg	57.6-225.4	0.64	0.58	0.48	0.57	0.6	0.5	0.54	0.47
3-Methyhistidine, mg	11.8-19.9	0.73	0.6	0.78	0.61	0.71	0.46	0.63	0.36
Vitamin C, mg	88.3-154.3	0.8	0.67	0.68	0.48	0.73	0.68	0.52	0.59
Vitamin D, mcg	3.6-7.1	0.67	0.56	0.62	0.59	0.42	0.29	0.59	0.28
Sodium, mg	2040-2918	0.74	0.49	0.54	0.48	0.68	0.49	0.45	0.35
Potassium, mg	2589-3509	0.71	0.62	0.66	0.5	0.49	0.57	0.59	0.44
Folate (natural), mcg	223.2-320.9	0.73	0.59	0.7	0.37	0.58	0.6	0.56	0.45

*Tabular entries are, more precisely, the square roots of the % (log-transformed) provided nutrient intake for the test sample explained by the applying the training sample metabolite model to estimate (log-transformed) intake in the (20%) test sample. These 'R' values may be slightly less than the correlation between (log-transformed) provided and estimated intake since regression coefficients are not re-estimated in application to the test sample.

**Batch 1 includes blood and 24-hr urine for the first 50 women, and Batch 2 includes these specimens for the subsequent 103 women, along with spot urine for all 153 feeding study women

Summary/Additional Comments

- The use of objective measures in nutritional and physical activity epidemiology provides a practical means of enhancing association study reliability in these important research areas.

- Additional objective measures forthcoming from accelerometer data for detailed physical activity patterns (Andrea LaCroix, OPACH); could also consider using DXA data in BMD sub-cohort for calibrated body composition.

Looking Ahead

- Metabolomic profiling using stored blood and urine (targeted MS, lipidomics; NMR) provides an approach to identifying many additional dietary biomarkers (including total energy biomarkers from stored specimens).
- Novel biomarkers can be applied in a regression calibration mode using self-report data, or applied directly to case-control data, bypassing any use of dietary self-report data (NPAAS III).
- Addressing ‘exposure’ measurement error is key to obtaining reliable information in these crucial public health areas.