What Data Sources Can Help Learn About Diet, Nutrition and Chronic Disease Risk

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

- Some axes to consider for priorities in 'learning from health care data to improve patient outcomes and public health'
- Energy intake and activity-related energy expenditure in relation to chronic disease risk
- Corrected disease associations combining objective measures with self-reported data
- Role of human feeding studies in nutrition biomarker development
- Example of micronutrient biomarkers in relation to chronic disease incidence
- Data sources and diet, nutrition and chronic disease research agenda

Axes to Consider for Health Care Analytics Priorities

- Nature (genetics) and Nurture (environment)
- Infectious and Chronic Disease
- Mechanistic (‘omics) and Black Box Studies (RCTs, cohort studies)
- Treatment and Prevention/Health Maintenance
- Statistical Collaborator and Substantive Biomedical Scientist

Some Observations on Chronic Disease Risk

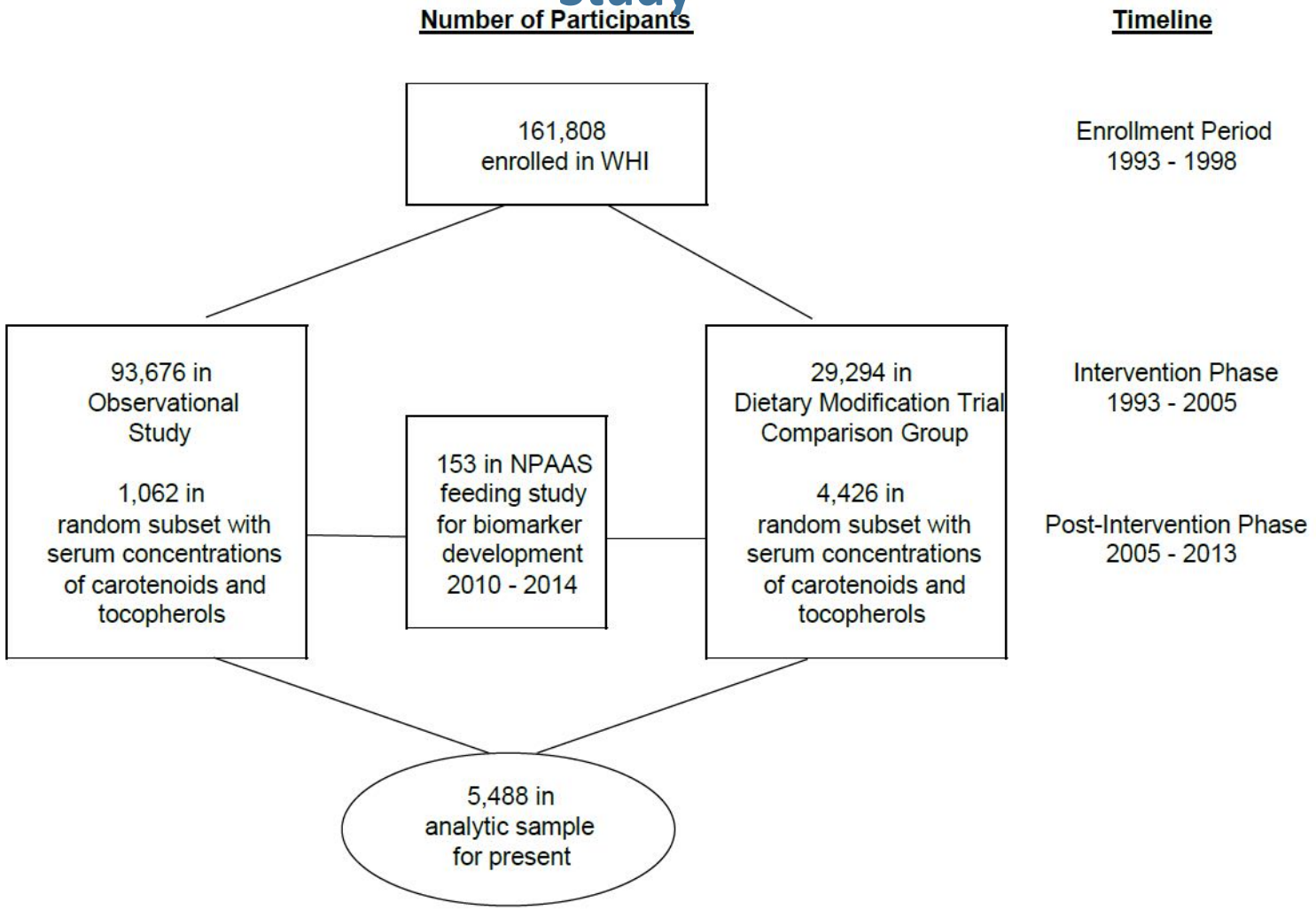
1. Chronic disease rates tend to be highly variable around the world, with rates for many diseases substantially elevated in Western populations.
2. Migrant populations tend to assume rates that prevail in their new environment, within a few generations.
3. Risk prediction models do not allow one to identify persons who will develop a specific disease, with even moderate precision, for any chronic disease.

Implication: There is still much to be learned about chronic disease risk determinants, and about modifiable risk factors for health maintenance and disease prevention

Where to Look?

1. Changes in incidence rates among migrants suggest that chronic disease risk depends importantly on commonplace habits and exposures.
2. Genetic/genomic factors also importantly involved in chronic disease pathogenesis, but perhaps more often as mechanistic factors/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 from nutritional epidemiology studies to date.

Study Cohort and Participant Flow in the Women's Health Initiative (WHI) and in its Nutrition and Physical Activity Assessment Study (NPAAS) Feeding Study



Estimated Hazard Ratio for 20% Increments in Self-Reported Total Energy Intake and in Self-Reported Activity-related Energy Expenditure (AREE) for Various Chronic Diseases in the OS from baseline (1994-1998) Through September 30, 2010

(Zheng et al, 2014, AJE)

	Uncalibrated				Calibrated			
	Energy		AREE					
Outcome 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				
Heart Failure	1.04	1.01,1.08	0.97	0.95,1.00				
Total CVD including CABG and PCI	1.00	0.99,1.01	1.00	0.99,1.01				
Total Invasive Cancer	1.01	1.00,1.02	0.99	0.99,1.00				
Invasive Breast Cancer	1.01	0.99,1.02	1.00	0.99,1.01				
Obesity-related Cancer	1.02	1.00,1.03	1.00	0.99,1.01				
Diabetes Mellitus	1.06	1.04,1.07	1.01	1.00,1.02				

Nutrient and Physical Activity Assessment Studies (NPAAS) in WHI

544 DM Trial women completed two-week DLW protocol with urine and blood collection and FFQ (50% intervention, 50% control). A 20% reliability subsample repeated protocol. (NBS; 2004-2006)

Biomarker study among 450 women in the OS for evaluating measurement properties of dietary and physical activity assessment approaches (frequencies, records, and recalls). With 20% reliability subsample. (NPAAS I; 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)

Calibration Equation Coefficients (β), Standard Errors (SE), and Percent of Biomarker Variation Explained (R^2) from Regression of Log(DLW energy biomarker) on Log(self-reported energy), 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
	7.614	0.009			7.597	0.009			7.607	0.009		
FFQ energy	0.054	0.017	3.8	6.5								
4DFR energy					0.161	0.028	7.8	13.3				
24HR energy									0.101	0.026	2.8	4.8
BMI	0.013	0.001	26.9	45.9	0.013	0.001	27.0	46.0	0.013	0.001	28.7	48.9
Age	-0.010	0.001	9.7	16.5	-0.009	0.001	8.4	14.3	-0.009	0.001	9.1	15.5
Black	-0.023	0.019			-0.024	0.018			-0.024	0.018		
Hispanic	-0.062	0.021	1.3	2.2	-0.065	0.020	1.5	2.6	-0.063	0.020	1.5	2.6
Other minatory	-0.041	0.040			-0.039	0.038			-0.038	0.039		
(Total)			41.7	71.1			44.7	76.2			42.1	71.8

Prentice et al (2001, AJE)

APPENDIX TABLE. Estimates of energy intake (kcal/day) obtained by self-reported food frequency questionnaire, a biomarker (total energy expenditure), and a calibrated food frequency questionnaire, according to body mass index category, Women's Health Initiative Nutritional Biomarkers Study, 2004–2005*

Body mass index† category	Self-reported FFQ‡		Total energy expenditure		Calibrated FFQ	
	Geometric mean	IQR‡	Geometric mean	IQR	Geometric mean	IQR
Normal (<25.0)	1,407	1,157–1,759	1,894	1,714–2,083	1,912	1,853–1,980
Overweight (25.0–29.9)	1,462	1,196–1,837	2,043	1,904–2,232	2,028	1,962–2,103
Obese (≥30)	1,454	1,161–1,897	2,213	2,034–2,415	2,247	2,156–2,338

* Note that the difference between FFQ energy intake (self-report) and total energy expenditure (biomarker) increases as body mass index increases. The biomarker-calibrated estimates, for the same women, correct for the measurement error using the model shown in table 4.

† Weight (kg)/height $(m)^2$.

‡ FFQ, food frequency questionnaire; IQR, interquartile range (25th–75th percentiles).

Calibrated Estimates of Energy Intake and Activity-Related Energy Expenditure

Activity-Related Energy Expenditure (AREE)
(Neuhouser et al, 2013, AJE)
Objective measure – DLW energy minus resting energy expenditure using indirect calorimetry

Energy and AREE in relation to cardiovascular disease, cancer and diabetes
(Zheng et al, 2014, AJE)

Estimated Hazard Ratio for 20% Increments in Total Energy and in Activity-related Energy Expenditure (AREE), Without and With Calibration to Correct for Measurement Error, for Various Chronic Diseases, in the OS from baseline (1994-1998) through September 30, 2010 *(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
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
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
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
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
Invasive 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
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
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

There is an Acute Need for Additional Intake Biomarkers for use in Chronic Disease Risk Association Analyses!

- Intake biomarker objective is to obtain consistent results when studying the biomarker association with chronic disease risk, as would be obtained if actual intakes were available on study cohort, even if with reduced precision

Chronic Disease Association Model (Cox model)

$$\lambda(t;z)=\lambda 0(t)ez\beta$$

where z is typically log-transformed average daily intake of a food or nutrient over a specified (short) period of time.

Intake Biomarker Requirement in Chronic Disease Context

Principal Biomarker (x) Requirement

$$z = x + e$$

where x is the intake biomarker, and the error term e is independent of x and of factors that may confound the relationship between z and the study disease. For biomarker plausibility and efficiency, x should explain much of the variability in z, in the study population.

Induced Hazard Model

To a typically excellent approximation (rare disease, normality)

$$\lambda(t;x) = \lambda 0(t) e x \beta$$

with the same hazard ratio parameter β.

Human Feeding Studies for Intake Biomarker Identification

$$z = x + e$$

z – intake during feeding period

x – pertinent blood or urine measured, study subject characteristics that may need to be considered in rescaling blood or urine measures to reflect intake, and potential confounding factors for nutritional variable in relation to the chronic disease under study

Identify x through linear regression of feeding study estimated intake on blood or urine measures and other factors

Biomarker equations *(Lampe et al, 2017, AJCN)*:

log(α-carotene) = 6.326 + 1.241 x log(serum α-carotene) + 0.082 x BMI - 0.325 x spring season indicator - 0.534 x summer season indicator - 0.258 x fall season indicator;

log(β-carotene) = 8.478 + 0.624 x log(serum β-carotene) + 0.050 x BMI;

log(L+Z) = 7.426 + 1.101 x log(serum L+Z) - 0.028 x age + 0.049 x BMI + 0.593 x white race indicator; and

log(α-tocopherol) = 2.885 + 2.077 x log(serum α-tocopherol) + 0.510 x dietary supplement use indicator.

Units are μg/d for the carotenoids and mg/d for α-tocopherol.

$R^2 \geq 36\%$ criterion for biomarker plausibility and for estimation efficiency.

Summary: Data Sources, and Diet and Health Research Agenda

- Diet (and physical activity) may be principal drivers of chronic diseases that are highly elevated in the US/ Western societies.
- Objective, reliable measures of key exposures are needed to learn about diet, nutrition and chronic disease risk.
- Research to identify intake biomarkers, and methodology to apply such, is multidisciplinary with statistics as one of the core disciplines.
- There is an urgent need to further evaluate the role of total energy intake, and of the absolute intakes of specific nutrients and foods, in determining the risk of major chronic diseases in US and elsewhere.