

# **Statistical Aspects of the Use of Biomarkers in Nutritional Epidemiology Research**

Ross L. Prentice<sup>1</sup>, Ying Huang<sup>1</sup>, Lesley F. Tinker<sup>1</sup>, Shirley A.A. Beresford<sup>2</sup>,  
Johanna W. Lampe<sup>1</sup>, Marian L. Neuhouser<sup>1</sup>

<sup>1</sup>Fred Hutchinson Cancer Research Center, Seattle, WA, USA

<sup>2</sup>University of Washington, Seattle, WA, USA

Correspondence to: Ross L. Prentice, Ph.D., Division of Public Health Sciences, Fred  
Hutchinson Cancer Research Center, 1100 Fairview Avenue North, P.O. Box 19024,  
Seattle, WA 98109-1024; phone: 206-667-4264; fax: 206-667-4142;  
e-mail: [rprentic@fhcrc.org](mailto:rprentic@fhcrc.org).

Abstract word count: 208

Text word count: 4808

Running title: Biomarkers in nutritional epidemiology

## **Abstract**

Few strong and consistent associations have arisen from observational studies of dietary consumption in relation to chronic disease risk. Measurement error in self-reported dietary assessment may be obscuring many such associations. Attempts to correct for measurement error have mostly used a second self-report assessment in a subset of a study cohort to calibrate the self-report assessment used throughout the cohort, under the dubious assumption of uncorrelated measurement errors between the two assessments. The use, instead, of objective biomarkers of nutrient consumption to produce calibrated consumption estimates provides a promising approach to enhance study reliability. As summarized here, we have recently applied this nutrient biomarker approach to examine energy, protein, and percent of energy from protein, in relation to disease incidence in Women's Health Initiative cohorts, and find strong associations that are not evident without biomarker calibration. A major bottleneck for the broader use of a biomarker-calibration approach is the rather few nutrients for which a suitable biomarker has been developed. Some methodologic approaches to the development of additional pertinent biomarkers, including the possible use of a respiratory quotient from indirect calorimetry for macronutrient biomarker development, and the potential of human feeding studies for the evaluation of a range of urine- and blood-based potential biomarkers, will briefly be described.

Keywords: Biological markers, epidemiologic methods, human feeding study, indirect calorimetry, nutrition, measurement error, regression calibration

## 1. Introduction

There is a pressing need to obtain reliable information on desirable dietary and physical activity patterns for body weight maintenance and chronic disease risk reduction. While sensible guidelines and recommendations on these important public health topics have been provided by various pertinent organizations, the underlying research often lacks the force and credibility to favorably influence the nutrition and physical activity choices of individuals, the advice given by primary care providers, agricultural policies, food production and processing choices, environmental design, educational programs, or food fortification and regulation activities, as may be required to achieve public health goals.

Consider the association of nutrition and physical activity to cancer, as but one example. A positive association between overweight, obesity and the incidence and mortality of cancer of the breast, prostate, colon, endometrium and kidney, among other prominent cancers, is well established (WCRF/AICR, 1997; Adams et al, 2006). However, a recent international review (WCRF/AICR, 2007) found few dietary or physical activity associations, the key elements of energy balance that drives body fat accumulation, that were judged to be 'convincing' or 'probable'. Similarly, the authors of the earlier international review (WCRF/AICR, 1997) wrote 'the significance of the data on energy intake and cancer risk in humans remains unclear' and 'In the view of the panel, the effect of energy intake on cancer is best assessed by examining the factors: rate of growth, body mass, and physical activity'. This state of affairs reflects considerable uncertainty in the nutrient consumption estimates from self-report data,

upon which association studies are typically based, for various clinical outcomes, including cancer, cardiovascular disease, diabetes, and frailty.

The effects of simple, classical measurement error, where the measured exposure equals the targeted exposure plus measurement error that is independent of the targeted exposure and of other study subject characteristics, are well recognized in nutrition epidemiology: Such measurement error primarily results in attenuation of odds ratios or hazard ratios, which can be overcome substantially by increasing sample size, at least as far as testing the null hypothesis (of no association) is concerned. However, the effects of measurement error can be far more insidious if there are systematic elements to measurement biases. For example, if a subset of a study cohort has a relatively high consumption of a nutrient and an elevated disease risk, but the self-reported exposure by this subset tends to be in the center of the nutrient distribution, then odds ratio as a function of self-reported exposure may increase as exposure increases to a peak, and drop off thereafter. This type of odds ratio pattern is rather common in nutritional epidemiology reports, and it will not be corrected by increasing sample size, or by studying the association in multiple populations, if due to systematic measurement error in self-reported consumption. We provide below examples of systematic bias in dietary self-report, and illustrate its impact on disease association analyses.

A group of researchers met in 2003 to discuss this measurement error problem, as well as the research agenda in nutrition and physical activity epidemiology more generally. This group argued that changes in diet and physical activity patterns could reverse the obesity epidemic in the United States and elsewhere, and could reduce the

risk of several prominent age-related chronic diseases, and they assessed that the conduct of the needed research is a 'demanding task that is now becoming scientifically achievable' (Prentice et al, 2004). A major approach recommended to strengthen dietary and physical activity association studies was the use of biomarkers for exposure assessment. Fortunately, good quality biomarkers had been developed and validated for some elements of diet, including total energy consumption among weight-stable persons using a doubly-labeled water (DLW) technique (Schoeller, 1999) and protein consumption using urinary nitrogen (UN) (Bingham, 2003). These biomarkers involve the urinary recovery of metabolites formed when the 'nutrient' is expended. The logarithms of these recovery biomarkers plausibly adhere to the classical measurement model described above, and these biomarkers have been applied to several hundreds of study subjects in various settings, including the National Cancer Institute's (U.S.A.) Observing Protein and Energy (OPEN) study (Subar et al, 2003; Kipnis et al, 2003).

Soon thereafter Women's Health Initiative (WHI) investigators were funded to conduct a Nutrient Biomarker Study (NBS) within a subcohort of the WHI Dietary Modification (DM) randomized controlled trial of a low-fat dietary pattern (48,835 postmenopausal women), with a goal of calibrating the WHI food frequency questionnaire (FFQ) data on energy and protein, for use in dietary association studies and in elucidating results on principal outcomes in the DM trial (Prentice et al, 2006; Beresford et al, 2006; Prentice et al, 2007). The NBS involved 544 women, half in the intervention group and half in the comparison group, of the DM trial. The findings to date from the NBS will be described below. For the moment, note that strong evidence of systematic bias was found for the FFQ assessments of energy and protein. For

example, FFQ energy underreporting was greater among overweight and obese women, and younger women, while that for proteins was in the same direction but weaker (Neuhouser et al, 2008). These data were used to develop calibrated energy consumption estimates in the WHI cohorts. While associations were generally absent without calibration, calibrated energy was positively related to total cancer risk as well as several specific cancers (Prentice et al, 2009).

WHI investigators built upon the NBS experience by proposing to bring in a biomarker of activity-related energy expenditure (AREE), defined as total energy expenditure (TEE) from DLW minus resting energy expenditure (REE) from indirect calorimetry, and by proposing to develop calibration equations, not just for the FFQ, but rather for frequencies, records and recalls of both dietary consumption and physical activity. At the time of this writing we are mid-way through the third and final year of this Nutrition and Physical Activity Assessment Study (NPAAS), which includes biomarker studies among 450 postmenopausal women in the WHI Observational Study (93,676 postmenopausal women). Laboratory analyses are still underway from NPAAS and only preliminary data analyses have been carried out at the time of this writing.

The most striking limitation of the nutrition biomarker data collected to date is the limited range of nutrients for which a suitable biomarker has been developed. Some approaches to the development and evaluation of biomarkers for additional nutrients and dietary components will be described below.

## **2. Biomarker-Calibrated Nutrient Consumption Association Studies**

Consider a biomarker assessment  $W$  of an (unmeasured) nutrient consumption variable  $Z$ , along with a corresponding self-report assessment  $Q$ . For example,  $W$  may be a (precise) estimate of the logarithm of short-term total daily energy (calories) consumed by a study subject as assessed using the doubly-labeled water method mentioned above, while  $Z$  is the logarithm of actual average daily energy consumption over a longer period of time (e.g., 6-month period), and  $Q$  is the logarithm of daily energy consumption over this longer time period as assessed using a FFQ. A plausible statistical model (Prentice et al, 2002) assumes a classical measurement model for the biomarker, and a more complex measurement model for the self-report data:

$$\begin{aligned} W &= Z + u \\ Q &= Z^* + e \end{aligned} \tag{1}$$

where  $Z^* = a_0 + a_1Z + a_2^T V + a_3^T V \otimes Z + r$ , and  $V$  is a vector of study subject characteristics (e.g., body mass index, age, ethnicity, ‘social desirability’ factors) that may influence dietary self-report measurement error,  $r$  is a person-specific random effect, and  $a_0, \dots, a_3$  are constants. Random variables on the right side of (1) are assumed to be statistically independent, given  $V$ . An important feature of (1) is the allowance for systematic bias in the self-report data, a modeling feature also proposed in earlier papers in this research area (Prentice, 1996; Carroll et al, 1998; Kipnis et al, 2001). While measurement error without such systematic components primarily attenuates odds ratios or hazard ratios in nutritional epidemiology, systematic biases can more severely distort dose-response associations. The key features of the model (1) are that the ‘reference’ assessment  $W$ , here a biomarker, has been properly calibrated so that the measured assessment plausibly adheres to a classical

measurement model; the ability to allow for sources of systematic bias in the self-report through the vector  $V$ ; and the independence of the error terms,  $u$  and  $r + e$ , for the biomarker and self-report. This independence assumption is crucial to the development of reliable epidemiologic association information, and is an unlikely assumption if  $W$  is instead based on a second self-report assessment. For example, if  $Q$  is based on a FFQ, while  $W$  is based on a food record from the same study subject, then positive correlations between  $W$  and  $Q$  could arise in part or in whole from correlated measurement errors, rather than from the self-reports reflecting the underlying nutrient consumption  $Z$ . A biomarker assessment  $W$ , on the other hand, has the advantage of objectivity and freedom from self-report biases. However, a biomarker assessment, perhaps based on a urine- or blood-based nutrient consumption, needs to be properly calibrated and the classical measurement model in (1) needs to be applicable.

Specialized statistical methods are needed to apply the measurement model (1) to data  $(Q, V)$  on a study cohort, and data  $(Q, V, W)$  on a biomarker subcohort. For example, Sugar et al (2007) developed regression calibration, refined regression calibration and conditional scores procedures for odds ratio estimation, and provided pertinent asymptotic distribution theory. Extensive simulation studies showed ordinary regression calibration to yield log-odds ratio parameter estimates having good efficiency and robustness, and minimal bias in configurations of interest. Moreover, it was shown that biomarker subsamples as small as 500 could be used to effectively calibrate cohorts of the size included in WHI, even with such high incidence outcomes as breast cancer, or total invasive cancer. Note, however, that about half of the variance in log-odds ratio parameter estimates may be due to variation in calibration equation



parameter estimates at a biomarker sample size of 500, depending somewhat on disease incidence. These and other statistical procedures for hazard ratio (Cox model) parameter estimation procedures were developed and compared in an unpublished 2006 Department of Biostatistics, University of Washington, doctoral dissertation by Dr. Pamela Shaw. Once again, ordinary regression calibration proved to provide an efficient approach to association parameter (here log-hazard ratio) estimation that involved negligible bias in simulation studies that emulate applications to WHI cohorts. The regression calibration approach involves replacing a self-report nutrient consumption (e.g., log-energy consumption) by a (nearly) unbiased estimate of actual nutrient consumption (e.g., actual log-energy consumption), here under the measurement model (1). Under a joint normality assumption for  $(Z, r + e)$  given  $V$ , it follows that  $Z$  given  $(Q, V)$ , and hence  $W$  given  $(Q, V)$  adheres to a simple linear regression model with non-zero coefficients for  $V$  or  $V \otimes Z$  indicating systematic bias in the FFQ assessment. Linear regression of  $W$  on  $(Q, V)$  in the biomarker subsample then allows calibrated consumption estimates to be obtained, throughout the remainder of the study cohort, from each subject's  $(Q, V)$  value.

We recently reported (Neuhouser et al, 2008) calibration equations under (1) for energy, protein, and % of energy from protein derived from biomarker data from 544 weight-stable women recruited from the Nutrient Biomarker Study mentioned above. These women (50% DM trial intervention group; 50% comparison group) were recruited at a representative 12 Clinical Centers (from a total of 40 Centers in the WHI). Each woman completed a basic protocol over a two-week period that included DLW, UN, an FFQ and other questionnaires, and each provided a blood specimen. A 20% reliability

subsample repeated the entire protocol about 6 months later. Table 1 shows estimated coefficients from linear regression of W on (Q, V) with standard error estimated obtained from a ‘sandwich’ variance estimator.

**Table 1.** Regression Calibration Coefficients for Log-transformed Total Energy, Total Protein, and % of Energy from Protein (Neuhouser et al, 2008)

Characteristic	Total Energy (kcal) Coefficient (Std. Error)	Protein (gm) Coefficient (Std. Error)	% Energy from Protein Coefficient (Std. Error)
Intercept	7.61 (0.013)	4.28 (0.024)	2.66 (0.011)
(Log) FFQ	0.062 (0.018)	0.211 (0.032)	0.439 (0.058)
Body Mass Index	0.013 (0.001)	0.012 (0.002)	-0.004 (0.002)
Age (yrs.)	-0.005 (0.001)	-0.008 (0.002)	-0.005 (0.002)

Note, for energy, the rather weak signal (coefficient of 0.062) arising from FFQ log-energy, while body mass index (weight in kg/height in meters squared) and age provide more highly significant predictors of biomarker-derived log-energy consumption. The full regression models fitted (Neuhouser et al, 2008) also included some moderate dependencies on ethnicity and socioeconomic factors, but there was little evidence of a dependence of systematic bias on actual consumption (i.e., of  $a_3 \neq 0$  in (1)). Log-protein consumption tends to show those same patterns, but with a larger regression coefficient (0.211) for log (FFQ) protein. The coefficient for log (FFQ) % of energy from protein was considerably larger (0.439), indicating better FFQ properties for this nutrient density measure, while there was an inverse dependence on body mass index, suggesting that energy underreporting among overweight and obese women derives primarily from fat and/or carbohydrate underreporting.

These calibration equations were applied to FFQ data obtained early in WHI to develop calibrated estimates of energy, protein, and % of energy from protein for individual women in the DM trial, as well as for women in the companion WHI Observational Study (OS), which included 93,676 postmenopausal women. Table 2, from Prentice et al (2009), shows Cox model hazard ratio parameter estimates for a 20% increment in nutrient consumption with and without biomarker calibration of the nutrient, based on the analysis of combined data from the DM trial comparison group and the OS that includes 5041 women who developed invasive cancer during WHI follow-up. The log-hazard ratio was modeled as a linear function of Z in these analyses. A regression calibration procedure was used to estimate log-hazard ratio parameters, and a bootstrap procedure was used to estimate standard errors for these parameters. The following variables were included in the hazard ratio model for total invasive cancer (to control confounding) (Prentice et al, 2009): race/ethnicity, education, exercise, current or past cigarette smoking, alcohol consumption, unopposed estrogen use, estrogen plus progestin use, history of diabetes, and hypertension.

**Table 2.** Total Invasive Cancer Hazard Ratio Estimates for a 20% Increment in Nutrient Consumption (Prentice et al, 2009)

	Hazard Ratio (95% Confidence Interval)		
	Energy	Protein	% of Energy from Protein
Without Biomarker Calibration	1.01 (0.99, 1.02)	1.00 (0.99, 1.02)	0.98 (0.96, 1.01)
With Biomarker Calibration	1.18 (1.10, 1.27)	1.06 (1.01, 1.12)	0.92 (0.85, 0.99)

Note, from Table 2, that following biomarker calibration there is a noteworthy positive association of total invasive cancer risk with energy consumption, a weaker positive association with protein consumption, and an inverse association with % of energy from protein, whereas these associations are not evident in the absence of biomarker calibration. The confidence intervals are considerably wider for the calibrated compared to the uncalibrated hazard ratio 'estimates', reflecting both the attenuation of coefficients and standard error estimates that attends the hazard ratio estimates without calibration, and the random variation in the coefficient estimates in Table 1 and hence in the calibrated consumption estimates, with a biomarker sample of only 544 women. The positive association with energy was also evident for several site-specific cancers including breast, colon, endometrium, and kidney, in alignment with the obesity associations mentioned in the introduction, whereas these associations were not evident without biomarker calibration (Prentice et al, 2009). The inverse association of total cancer with % of energy from protein points to fat, alcohol and carbohydrate collectively, as nutrients responsible for the positive energy association. Corresponding analyses have been carried out for other clinical outcomes, including cardiovascular diseases, diabetes, and frailty, with equally interesting, but yet to be published, results.

The analyses just summarized do not control for body mass index in the disease risk model, and the association between energy consumption and disease risk tend to be reduced or to disappear if body mass index is added to this model. Similarly, the association between body mass index and disease risk also tends to be reduced or disappear when calibrated energy consumption is included in the disease risk model (Prentice et al, 2009). Basically available data are not extensive enough to reliably

establish separate roles for the two cancer risk factors. Note that years of consuming a high calorie diet could readily lead to body fat accumulation, so that including body mass index in the disease risk model may lead to 'overcontrol'. On the other hand, a high body mass implies greater energy requirements, and analyses that do not control for body mass could include some confounding. Further study of this issue, with longitudinal data on body mass and calibrated energy consumption, is needed to sort out the joint association of energy consumption and body mass to the risk of these diseases.

It will also be important for biomarker data of these types to be assembled in additional epidemiologic cohorts, to study the consistency of emerging associations, and to study the transferability of calibration equations from one study population to another. Even then, the fact that suitable biomarkers, plausibly adhering to (1), have been developed only for a few nutrients will remain as an important nutritional epidemiology research barrier.

### **3. Opportunities for the Development of Biomarkers for Additional Nutrients**

#### **3.1 Possible use of the respiratory quotient**

While it is highly desirable that a biomarker yield a precise estimate of the consumption of a nutrient, at least over the short term, for nutritional epidemiology purposes, it may be possible to tolerate moderate imprecision as long as the classical measurement model in (1) applies. Even though the DLW method yields precise total energy expenditure precisely within a short protocol time period, the correlation of DLW-based log-energy estimates for the same women, separated by six months, among 101

women in the reliability study mentioned above was only 0.72 (Neuhouser et al, 2008), pointing to considerable temporal variation in actual consumption. A biomarker that assesses short-term consumption with, say, random measurement error of a similar magnitude to a study subject's actual temporal variation in nutrient consumption may have practical utility for nutrient association study purposes. Such a random component to measurement error, however, adds to the variance of  $u$  in (1), and results in reduced precision in hazard ratio parameter estimates.

For macronutrient associations with clinical outcomes, one possible biomarker approach is to augment total energy consumption from DLW, and protein consumption from UN, by the respiratory quotient (RQ) that attends indirect calorimetry. The RQ is defined as the ratio of carbon dioxide ( $\text{CO}_2$ ) produced to oxygen ( $\text{O}_2$ ) consumed during a testing period with  $\text{CO}_2$  and  $\text{O}_2$  in the same units. Many metabolized substances include only the elements carbon, hydrogen, and oxygen, and have a characteristic RQ value for a person in metabolic balance, ranging from 1.0 for carbohydrate to about 0.7 for fat or alcohol. Protein may produce an RQ in the vicinity of 0.8, depending on which amino acids are metabolized. A mixed diet may produce an RQ that approximates 0.7 times the percent of energy from fat (plus alcohol), plus 0.8 times the percent of energy from protein, plus 1.0 times the percent of energy from carbohydrate. There may, however, be substantial noise in the estimated RQ due to variation in the application of the indirect calorimetry protocol, the specific fats, proteins, or carbohydrates metabolized, or departure from precise metabolic balance.

We are currently implementing the following statistical approach in an attempt to develop potential biomarkers for fat (plus alcohol), carbohydrate, and protein from

measurements (T, P, R) where T is log(total energy) from DLW, P is log(protein) from UN, and R is the RQ from indirect calorimetry. We postulate a measurement model,

$$\begin{aligned} \log T &= \log p - \log(1-f-c) + e_T \\ \log P &= \log p + e_P \\ \log R &= \log \{\Theta_1 f + \Theta_2 c + \Theta_3(1-f-c)\} + e_R, \end{aligned} \tag{2}$$

where  $e_T$ ,  $e_P$ , and  $e_R$  are measurement errors associated with  $\log T$ ,  $\log P$ , and  $\log R$  respectively,  $f$  and  $c$  are the (short-term) percent of energy from fat and carbohydrate consumed by a study subject, and  $\Theta_1$ ,  $\Theta_2$ , and  $\Theta_3$  are RQ values for fat, carbohydrate, and protein for the study cohort. Regarding  $(p, f, c)$  as parameters to be estimated, the error terms on the right side of (2) are assumed to be statistically independent. Our goal is to specify  $(p, f, c)$  for each woman from which biomarkers for total protein, total fat, and total carbohydrate are  $p$ ,  $pf(1-f-c)^{-1}$  and  $pc(1-f-c)^{-1}$  respectively.

For this specification we will assume a joint normal model for  $(\log T, \log P, \log R)$  and maximize the corresponding likelihood with respect to  $(p, f, c)$ , at a specified  $\Theta = (\Theta_1, \Theta_2, \Theta_3)$ . This would be a trivial maximization were it not for positivity constraints on  $f$ ,  $c$ , and  $1-f-c$  that may be violated by corresponding empirical estimates, because of the measurement errors  $(e_T, e_P, e_R)$ . In fact, the specification of  $(p, f, c)$  may be tightened by imposing stronger constraints, such as  $f \geq 0.15$ ,  $c \geq 0.25$ ,  $1-f-c \geq 0.10$  on the fraction of calories from fat, carbohydrate, and protein in the woman's diet. A maximized likelihood  $L(\underline{\Theta})$  can then be calculated for the entire biomarker sample, and this profile likelihood can be scanned to yield potential biomarker values  $(p(\Theta^*), f(\Theta^*), c(\Theta^*))$  for each study subject, where  $\Theta^*$  is a maximizer (not necessarily unique) of the profile likelihood.

The procedure outlined above will be applied to data from the 450 women in the OS biomarker subsample mentioned above, and corresponding association studies of absolute and relative fat (plus alcohol) and carbohydrate consumption with clinical outcomes will be carried out. The independence of the error term 'u' that attends these objective assessments and the self-report measurement error ( $r + e$ ) will be highly plausible with these objective assessments. However, it will be necessary to examine the magnitude of the measurement error variance for these potential new biomarkers, and available data will not allow one to establish that these potential biomarkers are themselves properly calibrated. For example, the biomarker-derived estimates of % of energy from fat or % of energy from protein could be relocated and rescaled relative to the underlying consumption fractions. A human feeding study could be used to directly calibrate these potential biomarkers, and could be used for the development and evaluation of biomarkers for a range of other nutrients.

### **3.2 Role of a human feeding study for biomarker development**

A human feeding study provides the possibility of directly assessing short-term nutrient consumption by providing food and drink over a feeding period for each participant. Nutritional assessments in blood and urine samples obtained at the end of the feeding period can then be used to estimate the provided nutrient consumption. A key requirement is that blood and urine measures of interest stabilize to the provided diet by the end of the feeding period, as may be able to be accomplished with a feeding period as short as two weeks, by choosing an individualized dietary pattern that aims to approximate the usual dietary pattern for each study subject. Let  $W$  denote the



logarithm of provided average daily nutrient during the feeding period. Under this individualized approach the distribution of  $W$  is expected to approximate the distribution of actual consumption in the study cohort. A linear regression model

$$W = b_0 + b_1^T X + b_2^T V + e_w \quad (3)$$

may be used to predict the provided nutrient from corresponding blood or urine measures,  $X$ , and study subject characteristics,  $V$ . For example,  $X$  may include the logarithm of the blood concentration of the nutrient in question, or functions thereof, along with other nutritional measures, such as the DLW estimate of total energy consumption, or blood concentrations of related nutrients, that may help to explain variation in  $W$  values among the set of feeding study participants. Similarly,  $V$  may include study subject characteristics, such as body mass index, that may, for example, play a role in determining blood nutrient concentrations or urinary metabolic recovery under a given diet. Product terms between elements of  $X$  and  $V$  could also be included in (3). The estimated coefficients ( $b_0^*$ ,  $b_1^*$ ,  $b_2^*$ ) then allow a potential biomarker,  $W^* = b_0^* + (b_1^*)^T X + (b_2^*)^T V$ , to be defined for all cohort members, based on their  $X$  and  $V$  values. Under (3) this biomarker plausibly adheres to (1) with error term 'u' comprised of temporal variations in nutrient consumption plus departure of  $W^*$  from  $W$ .

How can  $W^*$  be evaluated as a biomarker for  $Z$  in (1)? First, it is important that (3) be comprehensive in the sense that factors that influence  $W$  (e.g., genetic factors) that may be omitted from the right side of (3) should not confound the corresponding disease risk association. Specifically, such factors should be independent of  $X$ , given  $V$ , so that any influence on  $W$  can be incorporated into  $e_w$ . Biological judgment concerning the blood or urine measures, and other factors included in  $(X, V)$  will be needed to

assess the adequacy of the choice of regression variables in (3). For example, urine- or blood-derived measures having strong postprandial or diurnal variations would seem less attractive for explaining consumption variations among individuals. Similarly, the dependence of the expectation of  $W$  given  $(X, V)$  needs to be adequately modeled in (3). Under these conditions the biomarker  $W^*$  can be expected to lead to approximately unbiased associations between  $Z$  and the clinical outcome, though the utility of the biomarker will depend on its precision as an estimate of  $W$ . The estimated percent of variation in  $W$  explained by the regression model (3) is a sensible measure of precision. For example, if the modeled regression variables explain 50% of the variation in  $W$ , then the variance in estimating  $Z$  derives equally from temporal variation in  $W$  and from variation of  $W^*$  about  $W$ . This may translate to sufficient precision for a range of useful association studies. Note, however, that the variance of association parameter estimates will reflect precision in  $\underline{b}^*$  estimation and the feeding study sample size, as well as precision in the estimation of coefficients of calibration equations that relate self-report data to biomarker values ( $W^*$ ) in (1).

At the time of this writing the authors have proposed a sizeable human feeding study, involving about 150 women in the Seattle component of the WHI. A number of urine- and blood-derived potential biomarkers will be entertained related to the consumption of sugars, whole grains, meat, fruits and vegetables, fats and oils, macronutrients and micronutrients, with evaluation based in part on the % of variation in provided dietary factor  $W$  explained by the putative biomarker in conjunction with other measurements.

#### **4. Discussion**

Data are beginning to emerge from various research groups suggesting that measurement error in dietary assessment may be dominating the results of nutritional epidemiology studies, which to date have mostly relied either on an assumption of self-report validity, or on an assumption of uncorrelated measurement errors between self-report assessment procedures. A consumption biomarker, capable of accurately assessing short-term intake, is available for a small number of nutritional factors. Biomarker data for these factors suggest strong systematic components to self-report measurement error, and correlated measurement errors among various self-report methodologies, implying that self-report data alone are unlikely to be sufficient for reliable association studies in nutritional epidemiology.

Some key biomarkers, such as the DLW assessment of energy consumption, are too expensive to be practical in an entire cohort of perhaps tens of thousands of study subjects, but biomarker substudies in a few hundred study subjects can be used to calibrate self-report data in a manner that has potential to reduce or eliminate the systematic error component.

For this biomarker approach to provide a fresh look at nutritional epidemiology associations broadly, it is necessary to develop biomarkers for additional nutrients and dietary components. A human feeding study of sufficient size could potentially do much to close this research gap. It will be interesting, for example, to compare the ability of urinary recovery and blood concentration nutritional measures (e.g., Kaaks et al, 2002) in their ability to explain variations in provided nutrient consumption. Each type of potential biomarker has its challenges. Available urinary metabolites may reflect only a

subset of an expended nutrient, while blood nutrient concentrations may be influenced by various factors other than consumption of the nutrient of interest, including body mass and the consumption of other nutrients. Hence, this is an important research pathway, but the extent to which it will satisfy the need for nutritional biomarkers cannot currently be predicted.

There are strong statistical components to measurement modeling, and to calibration substudy and feeding study design and analysis that merit the attention of additional members of the statistical community. This is truly an interdisciplinary research venue requiring the efforts of basic nutritional scientists, nutritional epidemiologists, and biostatisticians to help move this research area forward, with its major implications for public health and disease prevention.

## **ACKNOWLEDGMENTS**

This work was partially supported by CA119171 and CA53996 from the National Cancer Institute, and by contract N01-WH22110 from the National Heart, Lung and Blood Institute.

## References

- Adams, K. F., Schatzkin, A., Harris, T. B., Kipnis, V., Morris, T., Ballard-Barbash, R., Hollenbaeck, A., and Leizmann, M. F. (2006). Overweight, obesity and mortality in a large prospective cohort of persons 50 to 71 years old. *N. Engl. J. Med.* **355**, 763-778.
- Beresford, S. A., Johnson, K. C., Ritenbaugh, C., Lasser, N. L., Snetselaar, L. G., Black, H. R., Anderson, G. L., Assaf, A. R., Bassford, T., Bowen, D., Brunner, R. L., Brzyski, R. G., Caan, B., Chlebowski, R. T., Gass, M., Harrigan, R. C., Hays, J., Heber, D., Heiss, G., Hendrix, S. L., Howard, B. V., Hsia, J., Hubbell, F. A., Jackson, R. D., Kotchen, J. M., Kuller, L. H., LaCroix, A. Z., Lane, D. S., Langer, R. D., Lewis, C. E., Manson, J. E., Margolis, K. L., Mossavar-Rahmani, Y., Ockene, J. K., Parker, L. M., Perri, M. G., Phillips, L., Prentice, R. L., Robbins, J., Rossouw, J. E., Sarto, G. E., Stefanick, M. L., Van Horn, L., Vitolins, M. Z., Wactawski-Wende, J., Wallace, R. B., and Whitlock, E. (2006). Low-fat dietary pattern and risk of colorectal cancer: The Women's Health Initiative Randomized Controlled Dietary Modification Trial. *J. Am. Med. Assoc.* **295**, 643-654.
- Bingham, S. A. (2003). Urine nitrogen as a biomarker for the validation of dietary protein intake. *J. Nutr.* **133**, 921S-924S.
- Carroll, R. J., Freedman, L., Kipnis, V., and Li, L. (1998). A new class of measurement error models, with application to dietary data. *Canadian J. Statist.* **26**, 467-477.
- Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. (2002). Uses and limitations of random error correlations, in the validation of dietary questionnaire assessments. *Pub Health Nutr.* **5**, 969-976.

- Kipnis, V., Midthune, D., Freedman, L. S., Bingham, S., Schatzkin, A., Subar, A., and Carroll, R. J. (2001). Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am. J. Epidemiol.* **153**, 394-403.
- Kipnis, V., Subar, A. F., Midthune, D., Freedman, L. S., Ballard-Barbash, R., Troiano, R. P., Bingham, S., Schoeller, D. A., Schatzkin, A., and Carroll, R. J. (2003). Structure of dietary measurement error: results of the OPEN biomarker study. *Am. J. Epidemiol.* **158**, 14-21; discussion 22-26.
- Neuhouser, M. L., Tinker, L., Shaw, P. A., Schoeller, D., Bingham, S. A., Van Horn, L., Beresford, S. A. A., Caan, B., Thomson, C., Satterfield, S., Kuller, L., Heiss, G., Smit, E., Sarto, G., Ockene, J., Stefanick, M. L., Assaf, A., Runswick, S., and Prentice, R. L. (2008). Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *Am. J. Epidemiol.* **167**, 1247-1259.
- Prentice, R. L., Caan, B., Chlebowski, R. T., Patterson, R., Kuller, L. H., Ockene, J. K., Margolis, K. L., Limacher, M. C., Manson, J. E., Parker, L. M., Paskett, E., Phillips, L., Robbins, J., Rossouw, J. E., Sarto, G. E., Shikany, J. M., Stefanick, M. L., Thomson, C. A., Van Horn, L., Vitolins, M. Z., Wactawski-Wende, J., Wallace, R. B., Wassertheil-Smoller, S., Whitlock, E., Yano, K., Adams-Campbell, L., Anderson, G. L., Assaf, A. R., Beresford, S. A., Black, H. R., Brunner, R. L., Brzyski, R. G., Ford, L., Gass, M., Hays, J., Heber, D., Heiss, G., Hendrix, S. L., Hsia, J., Hubbell, F. A., Jackson, R. D., Johnson, K. C., Kotchen, J. M., LaCroix, A. Z., Lane, D. S., Langer, R. D., Lasser, N. L., and Henderson, M. M. (2006). Low-fat dietary pattern and risk of invasive breast cancer: The Women's Health

- Initiative Randomized Controlled Dietary Modification Trial. *J. Am. Med. Assoc.* **295**, 629-642.
- Prentice, R. L., Shaw, P. A., Bingham, S. A., Beresford, S. A. A, Caan, B., Neuhouser, M. L., Patterson, R. E., Stefanick, M. L., Satterfield, S., Thomson, C. A., Snetselaar, L., Thomas, A., and Tinker, L. F. (2009). Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women In press, *Am. J. Epidemiol.*
- Prentice, R. L., Sugar, E., Wang, C. Y., Neuhouser, M., and Peterson, R. (2002). Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Pub. Health Nutr.* **5**, 977-984.
- Prentice, R. L., Thomson, C. A., Caan, B., Hubbell, F. A., Anderson, G. L., Beresford, S. A. A., Pettinger, M., Lane, D. S., Lessin, L., Yasmeen, S., Singh, B., Khandekar, J., Shikany, J. M., Satterfield, S., and Chlebowski, R. T. (2007). Low-fat dietary pattern and cancer incidence in the Women's Health Initiative dietary modification randomized controlled trial. *J. Natl. Cancer Inst.* **99**, 1534-1543.
- Prentice, R. L., Willett, W. C., Greenwald, P., Alberts, D., Bernstein, L., Boyd, N. F., Byers, T., Clinton, S. K., Fraser, G., Freedman, L., Hunter, D., Kipnis, V., Kolonel, L. N., Kristal, B. S., Kristal, A., Lampe, J. W., McTiernan, A., Milner, J., Patterson, R. E., Potter, J. D., Riboli, E., Schatzkin, A., Yates, A., and Yetley, E. (2004). Nutrition and physical activity and chronic disease prevention: research strategies and recommendations. *J. Natl. Cancer Inst.* **96**, 1276-1287.
- Schoeller DA. (1999). Recent advances from application of doubly-labeled water to measurement of human energy expenditure. *J. Nutr.* **129**, 1765-1768.



Subar, A. F., Kipnis, V., Troiano, R. P., Midthune, D., Schoeller, D. A., Bingham, S., Sharbaugh, C. O., Trabulsi, J., Runswick, S., Ballard-Barbash, R., Sunshine, J., and Schatzkin, A. (2003). Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. *Am. J. Epidemiol.* **158**, 1-13.

Sugar EA, Wang CY, Prentice RL. (2007). Logistic regression with exposure biomarkers and flexible measurement error. *Biometrics.* **63**, 143-151.

World Cancer Research Fund/American Institute for Cancer Research(WCRF/AICR). (1997). *Food, Nutrition and the Prevention of Cancer: A Global Perspective*. Washington, DC: American Institute for Cancer Research.

World Cancer Research Fund/American Institute for Cancer Research(WCRF/AICR). (2007). *Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective*. Washington, DC: American Institute for Cancer Research.