



Published in final edited form as:

Can J Stat. 2011 September 1; 39(3): 498–509. doi:10.1002/cjs.10116.

Measurement error modeling and nutritional epidemiology association analyses

Ross L. PRENTICE and Ying HUANG

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington, USA 98109

Abstract

This paper summarizes the results of a Nutrient Biomarker Study in the Women's Health Initiative, and its application to studies of the association between energy and protein consumption and the risk of major cancers and cardiovascular diseases. The presentation emphasizes measurement error modeling and related data analysis methods, since addressing measurement issues appears to be central to these topics and to progress in nutritional epidemiology more generally. The manner in which body mass index is modeled in disease association analysis is particularly challenging, since it could serve as a mediator or as a confounder of the association, and at the same time contributes valuably to energy and protein consumption assessment. A hazard ratio parameter estimation procedure that acknowledges body mass index as a possible mediating variable is described and applied. Some aspects of the future nutritional epidemiology research agenda are briefly discussed, including an ongoing human feeding study to develop biomarkers for additional dietary components.

Keywords

cancer; cardiovascular disease; diet; epidemiology; failure time data; measurement error

1. INTRODUCTION

It is a pleasure to contribute a paper in honour of our esteemed colleagues Drs. Jack Kalbfleisch and Jerry Lawless. Jack and Jerry have each had a tremendous impact on statistical and biostatistical research methods, and on the analysis of failure time and life history analyses in particular. Their theoretical and applied contributions have been recognized by many Canadian and international awards. I was fortunate to co-author two editions of a book on failure time methods with Jack (Kalbfleisch & Prentice, 2002), which summarized the literature on a broad range of failure time topics, with emphasis on hazard ratio regression methods emerging from the seminal paper Cox (1972). One of those topics was that of estimating hazard ratio coefficients when there is measurement error in some elements of the regression variable. This topic is particularly central in the important public health area of nutritional epidemiology, which has depended almost exclusively on self-reported dietary consumptions, which may be affected by complex and influential assessment errors.

There is a pressing need for reliable information on recommended dietary and physical activity patterns for body weight maintenance and for chronic disease risk reduction. While sensible guidelines and recommendations are available from various organizations, these

may lack enough specificity and force to influence individual dietary choices and the type of societal changes that may be needed to begin to reverse the obesity epidemic in Western societies and to achieve related public health goals.

An international expert consultation that summarized the world literature on diet, nutrition and the prevention of chronic diseases (Diet, Nutrition and the Prevention of Chronic Diseases 2003) does not list energy consumption among the factors that are convincingly, probably or possibly associated with cardiovascular disease risk, though overweight is described as convincingly associated with increased risk. Similarly an expert panel reviewing the world literature on nutrition and cancer prevention (World Cancer Research Fund/American Institutes of Cancer Research, 1997) writes that ‘In the view of the panel, the effect of energy on cancer is best assessed by examining the data on related factors: rate of growth, body mass, and physical activity’, and that the ‘significance of the data on energy intake and cancer risk in humans is unclear’.

These summaries reflect considerable uncertainty about the reliability of nutrient consumption estimates from self-report dietary data, upon which dietary association studies are typically based. The food frequency questionnaire (FFQ) has been ubiquitous in nutritional epidemiology for the past 25 years, in part because its self-administered and machine-readable features make it practical for application to large study cohorts. The FFQ is believed to have better measurement properties for nutrient densities than for absolute nutrient consumption (Willett et al. 1985), and most epidemiologic studies only report associations for nutrients having some form of energy consumption adjustment.

There are a small number of nutrients for which a well established biomarker of short-term consumption has been developed, including a doubly-labeled water (DLW) assessment of energy consumption over a 14 day period (Schoeller 1999), and a urinary nitrogen (UN) assessment of protein consumption from a 24 hour urine collection (Bingham 2003). These urinary recovery biomarkers (Kaaks et al. 2002), provide objective estimates of short-term consumption among persons in energy balance, with measurement error that is plausibly independent of study subject characteristics such as body mass index (BMI), defined as weight in kg divided by the square of height in meters, age, and gender, and importantly is plausibly independent also of dietary self-report measurement error. However, it is not practical to obtain these biomarkers for the tens of thousands of persons in a typical epidemiology cohort study, and application in advance of disease diagnosis is essential. It follows that the practical study design entails biomarker application to a randomly selected subcohort of a study cohort, in conjunction with concurrent self-report data, followed by use of the biomarker data to produce measurement error corrected consumption estimates throughout study cohorts for use in disease association analyses.

These estimates can be obtained by simple linear regression of log-transformed biomarker values on corresponding log-FFQ values and other study subject characteristics, including BMI and age. They arise from a measurement model (Prentice et al. 2002)

$$\begin{aligned} W &= Z + e \\ Q &= a_0 + a_1 Z + a_2^T V + \varepsilon, \end{aligned} \quad (1)$$

where the biomarker W (e.g. log DLW energy) is assumed to adhere to a classical measurement model with error e , where Z is the targeted consumption (e.g., logarithm of average daily energy consumption over a certain time period); Q is the corresponding log FFQ energy consumption; V is a vector of characteristics that may also relate to Q , for example through systematic biases in the FFQ assessment; a_0 , a_1 , a_2 are parameters to be

estimated, and ϵ is a random error term that could include a random effect component. A joint normality assumption for (Z, V, ϵ) then gives

$$E(Z|Q, V) = b_0 + b_1 Q + b_2^T V,$$

and therefore

$$E(W|Q, V) = b_0 + b_1 Q + b_2^T V,$$

under the crucial assumption that the biomarker and self-report errors e and ϵ are statistically independent. It follows that regression of W on the self-report Q and pertinent characteristics V , as in Table 1, yields an estimate of Z that has been corrected for assessment error under model 1. This ‘calibrated’ consumption estimate is readily calculated from (Q, V) values for all members of the study cohort. Note from Table 1 that the FFQ provides only a weak signal for energy estimation, with a coefficient of 0.062 rather than a coefficient near 1 as would be expected from an accurate and precise consumption estimate. Note also that these calibrated estimates can be obtained under the less restrictive assumption of joint normality of (Z, ϵ) given V .

2. NUTRITIONAL BIOMARKER STUDY IN THE WOMEN’S HEALTH INITIATIVE

We pursued this strategy by conducting a Nutrient Biomarker Study (NBS) among 544 postmenopausal women enrolled in the Women’s Health Initiative (WHI) Dietary Modification Trial, during 2004–2005 (WHI Study Group 1998). The NBS involved DLW and UN biomarkers, a concurrent FFQ and some additional questionnaire information. The randomized controlled DM trial was conducted among 48,835 postmenopausal women in the age range 50–79 when enrolled during 1993–1998. It tested whether a low-fat dietary pattern (40% of women) would reduce cancer risk compared to a usual diet comparison group (60% of women). Weight-stable women without a diagnosis of cancer or cardiovascular disease during trial follow-up were randomly selected for NBS participation at a representative 12 of the 40 participating WHI clinical centers, 50% from the intervention and 50% from the comparison group. A 20% reliability subsample repeated the entire NBS protocol about 6 months after the original data collection.

Additional detail on the NBS is given in Neuhouser et al. (2008), where calibration equations for the assessment of energy, protein, and protein density (percent of energy from protein) are also given, a reduced form of which are provided here in Table 1.

In fact, FFQ energy consumption estimates exhibit strong systematic biases, with overweight and obese women underestimating to a much greater extent than normal weight women, and with younger postmenopausal women underestimating to a greater extent than older postmenopausal women. The full calibration equations included also some moderate dependencies on income and socioeconomic factors (Neuhouser et al. 2008). Note also from Table 1 that the FFQ assessment appears to have better measurement properties for protein density than for absolute protein or energy, as expected.

3. ENERGY AND PROTEIN ASSOCIATION WITH CANCER AND CARDIOVASCULAR DISEASE

These calibration equations were applied to FFQ assessments and other study subject characteristics (Q , V) obtained early in WHI follow-up, and calibrated consumption estimates were associated with the subsequent incidence of invasive cancers and cardiovascular diseases in WHI cohorts. The WHI cohorts included the comparison group in the Dietary Modification Trial, and the companion WHI Observational Study, a prospective cohort study among 93,676 postmenopausal women drawn from the same catchment population with much commonality in data collection and outcome ascertainment with the clinical trial (WHI Study Group 1998). Table 2 shows estimated hazard ratios for a 20% consumption increment, with and without biomarker calibration of the FFQ consumption estimates, for several chronic disease categories, from Prentice et al. (2009a) and Prentice et al. (2011). These hazard ratio estimates arise from Cox models with log-hazard ratios that are linear in the estimated log-nutrient consumption, so that the hazard ratio for a fractional increase in consumption is assumed to be constant across the consumption distribution. A bootstrap procedure was used to estimate variances and confidence intervals for corresponding hazard ratio parameter estimates. Without calibration there was little evidence of association of any of the cancer or cardiovascular disease categories listed with energy, protein, or protein density. In contrast, calibrated energy consumption was positively associated with breast, colon, and total invasive cancer risk and with coronary heart disease, and was marginally inversely associated with stroke incidence. Calibrated protein consumption was also positively associated with breast and total cancer, while there was some evidence of an inverse association between calibrated protein density and total cancer incidence. These analyses did not include BMI in the disease risk model, and BMI was highly correlated (0.81) with biomarker log-energy consumption in the NBS, less so with log-protein consumption (0.46) and only weakly correlated (0.12) with log-protein density. Table 2 also gives hazard ratio estimates for calibrated energy in analyses that included BMI in the log-hazard ratio regression model ('BMI adjusted' analyses). The positive energy associations essentially disappear with this addition, while an inverse association with stroke becomes more pronounced.

4. BODY MASS INDEX & THE INTERPRETATION OF ENERGY & PROTEIN ASSOCIATION ANALYSES

The analyses of Table 2 present a rare look at absolute energy and absolute protein consumption in relation to the risk of important chronic diseases. However the interpretation of these associations is complicated by the multiple possible roles and influences of BMI. First, dietary consumption patterns tend to track over many years, and body fat deposition and an increasing BMI is an expected consequence of an excessive energy consumption. Hence a disease association with energy may be substantially mediated by body fat accumulation, and it would represent over-control to include BMI in the hazard ratio regression analyses in assessing the full dietary association. On the other hand overweight and obese persons expend relatively more energy carrying out routine tasks of daily living and it is possible that there is some confounding associated with BMI, and its exclusion from the disease risk model may yield hazard ratio estimates that are biased away from the null. These mediating and confounding possibilities would exist even if energy consumption was measured without error. Here, however, BMI also serves as an important component of the energy calibration procedure, with possible further complexity in the interpretation of the calibrated energy association analysis.

Dr. Laurence Freedman (Gertner Institute, Tel Aviv, Israel) has pointed out to us (personal communication) that calibrated association analyses of the type shown in Table 2 without BMI adjustment may incorporate some bias due to the inclusion of BMI in the calibration equations. Furthermore, in his ongoing work with collaborators in the US National Cancer Institute it seemed that biases of this type could be substantially mitigated by obtaining association parameter estimates for the dietary variables from the BMI-adjusted analyses.

To examine this issue in the present context consider the BMI adjusted analysis of the type shown in Table 2. These arise from an underlying Cox model hazard rate model that can be written

$$\lambda(t; U, X) = \lambda_0(t) \exp(\beta_1 U + \beta_2^T X) \quad (2)$$

where U is BMI and X is comprised of the unmeasured dietary variable Z of interest (first element) along with other variables V involved in developing the calibrated estimate (\widehat{Z}) of Z , and other variables included in the disease risk model to control confounding (the elements of β_2 for factors involved in V that are not included in the disease rate model are set to zero).

Data on (U, X) and disease incidence times, which may be subject to independent right censoring (e.g., Kalbfleisch & Prentice 2002), do not allow one to distinguish mediating from confounding roles for U ($\beta_1 \neq 0$). The induced hazard ratio model for X alone can be written

$$\lambda(t, X) = \lambda_0(t) e^{\beta_2^T X} E \{ e^{\beta_1 U} | X, T \geq t \}$$

which under a rare disease assumption and a joint normality assumption for (U, X) is well approximated by

$$\lambda(t, X) = \lambda_0^*(t) \exp(\beta_*^T X) \quad (3)$$

where $\beta_* = \beta_2 + (\text{var} X)^{-1} \text{cov}(X, U) \beta_1$. Hence we may consider the first element of β_* as our target of estimation for the full dietary effect of Z on the hazard ratio. If Z , and hence X , were directly measurable one could reliably estimate β_* from estimates of β_1 and β_2 using the sample covariance of X and U and the sample variance for X . This idea has been proposed by Freedman and colleagues in a draft paper they shared with us entitled ‘Using regression calibration equations that combine self-reported intake and biomarker measures to obtain unbiased estimates and more powerful tests of dietary associations.’

Now suppose that Z is measured by W and Q with error according to (1), including possible systematic bias related to V . One can write

$$W = Z + e = \widehat{Z} + \widehat{\varepsilon} \quad (4)$$

where \widehat{Z} is the calibrated estimate of Z . Under the natural assumption that \widehat{Z} is unrelated to disease risk given (X, U) one has that (t, U, X, \widehat{Z}) equals (2) from which

$$\lambda(t; U, \widehat{X}) = \lambda_0(t) \exp(\beta_1 U + \beta_2^T \widehat{X}) E \{ \exp \beta_{21} (\widehat{\varepsilon} - e) | T \geq t, U, \widehat{X} \},$$

where \widehat{X} equals X except that \widehat{Z} is substituted for Z , and β_{21} is the corresponding (first) element of β_2 . Under the measurement model (1) e is independent of (U, \widehat{Z}) and the residuals $\widehat{\varepsilon}$ are uncorrelated with (U, \widehat{X}) (assuming that all pertinent confounding factors are included in the calibration equation that generates \widehat{Z}). Hence, under a rare disease assumption

$$\lambda(t; U, \widehat{X}) = \tilde{\lambda}_0(t) \exp(\beta_1 U + \beta_2^T \widehat{X}) \quad (5)$$

to a good approximation. It follows that the hazard ratio estimators from analyses of the calibrated estimator that include BMI should be approximately unbiased for the parameters of model (2) under our modeling assumptions.

The induced hazard ratio model for \widehat{X} from (5) is approximately

$$\lambda(t; \widehat{X}) = \widehat{\lambda}_0(t) \exp(\tilde{\beta}^T \widehat{X}) \quad (6)$$

where $\tilde{\beta} = \beta_2 + (\text{var} \widehat{X})^{-1} \text{cov}(\widehat{X}, U) \beta_1$. This parameter $\tilde{\beta}$ is expected to agree closely with that based on application of the Cox model directly to \widehat{Z} and potential confounding factors, as led to the calibrated consumption hazard ratio estimators (without BMI adjustment) shown in Table 2.

However $\tilde{\beta}$ may differ somewhat from the desired target β_* in (3). Specifically, even though $\text{cov}(\widehat{X}, U)$ would differ trivially from $\text{cov}(X, U)$ for a sufficiently large biomarker sample, and is readily estimated by the sample covariances with W in place of Z , the variance matrix $\text{var} \widehat{X}$ may differ from $\text{var} X$ in (3), depending on the comparative magnitudes of the variances of e and $\widehat{\varepsilon}$ in (4). Specifically, one can estimate $\text{var} X$ by using sample variances and covariances for all elements except those in the first row and column. Except for the (1,1) element, $\text{var} Z$, once can estimate the first row and column of $\text{var} X$ by biomarker study sample covariances between W and X . The estimation of $\text{var} Z$, however, is a more delicate issue that forces us to be specific about the time period that is intended to be covered by the dietary variable.

As noted above the NBS included a reliability subsample among 111 women, with primary and reliability subsamples separated by about 6 months. If one defines the targeted dietary variables as that pertaining to the recent diet, say over the past 6 months, then it will be plausible for the measurement errors e_1 and e_2 that attend the two biomarker assessments W_1 and W_2 for women in the reliability subsample to be approximately uncorrelated in which case $\text{var} Z$ can be estimated by the sample variance of W in the biomarker study, minus 0.5 times the sample variance of $(W_1 - W_2)$ in the reliability subsample. On the other hand, relative to a dietary variable that pertains to consumption over the preceding years or decades that may be relevant to disease risk and to body fat accumulation, the reliability biomarker measurement errors can be expected to have a correlation ρ that is positive, in which case $\text{var} Z$ can be estimated by the sample variance of W minus $0.5(1 - \rho)^{-1}$ times the sample variance of $(W_1 - W_2)$. Replicate biomarker data over the long period of time that may be relevant to those dietary associations would be needed to directly assess the temporal patterns of biomarker measurement error correlation. Here, instead, we provide sensitivity analyses for some choices of ρ ranging from $\rho = 0$, as may be appropriate for recent diet, to positive values as may be relevant to long-term dietary patterns. The sensitivity analyses target the coefficient of Z (first element of X) in β_* in (3).

5. APPLICATION TO CARDIOVASCULAR DISEASE ASSOCIATION ANALYSES

In Table 3 we provide energy and protein analyses using the methods of the previous section for the full range of cardiovascular disease outcomes considered in Prentice et al. (2011) where details of the variables included for confounding control for each outcome category and of the stratification of the Cox model can be found. The left column of Table 3 shows hazard ratios for a 20% consumption increment and estimated 95% confidence intervals from 500 bootstrap samples, as given in Prentice et al. (2011). The second column shows corresponding estimates based on the approximation (6). In applying this approximation $\widehat{\text{var}}\hat{X}$ was estimated as described in the preceding section, with $\widehat{\text{var}}\hat{Z}$ estimated by its sample variance. The good agreement between the two columns for each of energy, protein, and protein density attests to the adequacy of this approximation in this context, where about 5% of study subjects experienced the most global cardiovascular disease outcome (total CHD plus total stroke plus coronary artery bypass graft plus percutaneous coronary intervention). The following Table 3 columns present corresponding hazard ratio estimates based on the estimated β_* in (3), for certain specific biomarker measurement error correlations in the NBS reliability subsample, with confidence intervals again based on 500 bootstrap samples.

The fact that the underlying dietary variable has positive variance limits the possible values of the biomarker measurement error correlation to be <0.69 for energy, <0.59 for protein, and <0.53 for protein density. In fact the substantial local variation in protein over time presumably leads to measurement error correlations that are relatively small for both protein and protein density, even if Z is defined as the logarithm of dietary consumption over a period of time as short as 6 months or a year. The local temporal variations in energy consumption are likely to be much smaller, however. Energy biomarker measurement error correlations may be close to zero, if (log) consumption over the short biomarker ascertainment period of two weeks is compared to average (log) consumption over a time period as short as a few months. In contrast biomarker measurement error correlations relative to a long-term average consumption over some years or decades are likely to be positive and sizeable, since the reliability sample assessments at two close points in time can be expected to have commonalities arising from age-related changes in energy consumption and to other dietary changes during adulthood. With these thoughts in mind we present results in Table 3 for $\rho = 0$ and $\rho = 0.375$ for all three dietary variables, and also for $\rho = 0.5$ for energy. The interpretation of the protein and protein density analyses is rather insensitive to biomarker measurement error correlation. The energy association analyses are more sensitive. At a value of $\rho = 0.5$, that we think to be plausible for energy in this setting for long-term energy consumption, the hazard ratio estimates agree closely with our published values given in the first column, though confidence intervals are somewhat wider. Hence, these analyses provide some support for positive associations between long-term energy consumption and the risk of coronary heart disease, and the risk of total cardiovascular disease including coronary artery bypass graft and percutaneous coronary intervention, with those associations largely attributable to related temporal increases in BMI; and supportive of an inverse association between protein consumption, over the short or long term, and the risks of stroke.

6. ASPECTS OF THE FUTURE RESEARCH AGENDA FOR NUTRITIONAL EPIDEMIOLOGY

Developments of the type described above can help to define the future research agenda in nutritional epidemiology. It is clear that association studies of energy and protein consumption require a careful account of measurement error in dietary assessment for

reliable inferences. Studies using suitable biomarkers seem to provide the logical next step to enhancing the reliability of dietary association studies, especially for studies involving the absolute intake of nutrients or foods. The issues described in the preceding section imply a need for biomarker reliability studies that take place over some months or years, along with concurrent dietary and physical activity assessment data, to elucidate the complex issues related to diet, physical activity, energy balance, body fat deposition, and chronic disease risk.

A major limitation in pursuing the biomarker approach to nutritional epidemiology arises from the rather few nutrients or dietary components for which there is an established biomarker that plausibly adheres to a classical measurement model, as in (1). We have recently initiated a human feeding study among 150 WHI participants in Seattle as an attempt to develop and evaluate additional such biomarkers. A human feeding study provides the possibility of directly assessing short-term nutrient consumption by providing food and drink to each participant over the feeding period. We will employ a two week feeding period in which each woman will be provided a diet that approximates her usual diet so that blood and urine measures stabilize quickly, and so that the dietary variation in the study population is retained. Food and drink having well characterized nutrient composition will be used in diet formulation. Our plan is to examine the extent to which variation in the provided diet can be ‘explained’ by relevant urine and blood measures and individual characteristics, with regression equations that may explain 50% or more of the actual variation considered as potential biomarkers. Both candidate biomarkers and biomarker discovery efforts will be included. Some additional detail on these plans is given in Prentice et al. (2009b).

Acknowledgments

This work was partially supported by grants from the US National Cancer Institute, and by contract from the National Heart, Lung and Blood Institute for the Women’s Health Initiative.

REFERENCES

- Bingham SA. Urine Nitrogen as a Biomarker for the Validation of Dietary Protein Intake. *Journal of Nutrition*. 2003; 133:921S–924S. [PubMed: 12612177]
- Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*. 1972; 34:187–220.
- Joint WHO/FAO Expert Consultation on Diet, Nutrition and the Prevention of Chronic Diseases. *Diet, Nutrition and the Prevention of Chronic Disease: Report of a joint WHO/FAO expert consultation. WHO Technical Report Series*. 2003; 2003; 916:88.
- Kaaks R, Ferrari P, Ciampi A, Plummer M, Riboli E. Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public Health Nutrition*. 2002; 5:969–976. [PubMed: 12638598]
- Kalbfleisch, JD.; Prentice, RL. *The statistical analysis of failure time data*. Second edition. Wiley; New York: 2002.
- Neuhouser ML, Tinker L, Shaw PA, Schoeller D, Bingham SA, Horn LV, Beresford SA, Caan B, Thomson C, Satterfield S, Kuller L, Heiss G, Smit E, Sarto G, Ockene J, Stefanick ML, Assaf A, Runswick S, Prentice RL. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women’s Health Initiative. *American Journal of Epidemiology*. 2008; 167:1247–1259. [PubMed: 18344516]
- Prentice RL, Huang Y, Kuller LH, Tinker LF, Van Horn L, Stefanick ML, Sarto G, Ockene J, Johnson KC. Biomarker-calibrated energy and protein consumption and cardiovascular disease risk among postmenopausal women. *Epidemiology*. 2011; 22:170–179. [PubMed: 21206366]

- Prentice RL, Huang Y, Tinker LF, Beresford SA, Lampe JW, Neuhaus ML. Statistical Aspects of the Use of Biomarkers in Nutritional Epidemiology Research. *Statistics in BioSciences*. 2009b; 1:112–123. [PubMed: 19841649]
- Prentice RL, Shaw PA, Bingham SA, Beresford SA, Caan B, Neuhaus ML, Patterson RE, Stefanick ML, Satterfield S, Thomson CA, Snetselaar L, Thomas A, Tinker LF. Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *American Journal of Epidemiology*. 2009a; 169:977–989. [PubMed: 19258487]
- Prentice RL, Sugar E, Wang CY, Neuhaus M, Patterson R. Research strategies and the use of nutrient biomarkers in studies of diet and chronic disease. *Public Health Nutrition*. 2002; 5:977–984. [PubMed: 12633522]
- Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. *Journal of Nutrition*. 1999; 129:1765–1768. [PubMed: 10498745]
- Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. *American Journal of Epidemiology*. 1985; 122:51–65. [PubMed: 4014201]
- The Women’s Health Initiative Study Group. Design of the Women’s Health Initiative clinical trial and observational study. *Controlled Clinical Trials*. 1998; 19:61–109. [PubMed: 9492970]
- World Cancer Research Fund & American Institute for Cancer Research. *Food, Nutrition and the Prevention of Cancer: a global perspective*. American Institute for Cancer Research; Washington, DC.: 1997. p. 371

Table 1

Regression calibration coefficients for log-transformed total energy, total protein, and protein density (Neuhouser et al. 2008).

Characteristic	Total Energy (kcal) Coefficient (Std. Error)	Protein (gm) Coefficient (Std. Error)	Protein Density 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)

Table 2

Hazard ratio estimates for a 20% increment in nutrient consumption in Women’s Health Initiative cohort study of over 80,000 women (Prentice et al. 2009, Prentice et al. 2010).

Disease (Number of cases)	Consumption Calibration	BMI Adjustment	Hazard Ratio (95% Confidence Interval)		
			Energy	Protein	Protein Density
Total cancer (5041)	No	No	1.01(0.99,1.02)	1.00(0.99,1.02)	0.98(0.96,1.01)
	Yes	No	1.18(1.10,1.27)	1.06(1.01,1.12)	0.92(0.85,0.99)
	Yes	Yes	0.90(0.76,1.06)		
Breast Cancer (1703)	No	No	1.00(0.98,1.03)	1.00(0.98,1.02)	0.99(0.94,1.03)
	Yes	No	1.24(1.11,1.38)	1.09(1.01,1.19)	0.92(0.852,1.04)
	Yes	Yes	1.11(0.81,1.53)		
Colon cancer (363)	No	No	0.95(0.91,0.99)	0.97(0.93,1.01)	1.03(0.94,1.03)
	Yes	No	1.35(1.06,1.71)	1.02(0.87,1.20)	0.96(0.75,1.22)
	Yes	Yes	0.70(0.41,1.18)		
Coronary Heart Disease (1516)	No	No	1.00(0.98,1.03)	0.99(0.97,1.01)	0.95(0.91,1.00)
	Yes	No	1.18(1.04,1.33)	1.01(0.92,1.10)	0.85(0.75,0.97)
	Yes	Yes	0.89(0.62,1.27)		
Stroke (1224)	No	No	0.99(0.97,1.02)	0.99(0.97,1.01)	0.98(0.43,1.03)
	Yes	No	0.86(0.75,1.00)	0.89(0.82,0.98)	0.94(0.84,1.06)
	Yes	Yes	0.61(0.41,0.91)		

Table 3

Hazard ratio estimates for a 20% increment in calibrated FFQ energy, protein and protein density in relation to various cardiovascular disease outcomes, from 80,330 women enrolled in the WHI Dietary Modification Trial Comparison Group or Observational Study.

CVD	No. Incident Cases	From Prentice (2010)	Based on $\tilde{\beta}$ in (6)	Based on β in (3)		
				Biomarker	Measurement Error	Correlation
			0	0.375	0.50	
Energy						
Total CHD	1516	1.18(1.04,1.33)	1.20(1.06,1.36)	1.00(0.81,1.24)	1.06(0.86,1.26)	1.16(0.94,1.41)
Nonfatal MI	1140	1.13(0.98,1.30)	1.15(1.00,1.33)	0.99(0.76,1.27)	1.04(0.86,1.27)	1.11(0.90,1.38)
Coronary death	432	1.29(1.07,1.56)	1.32(1.08,1.63)	1.06(0.76,1.47)	1.14(0.88,1.47)	1.26(0.95,1.68)
CABG and PCI	2055	1.18(1.08,1.29)	1.24(1.11,1.37)	0.96(0.81,1.14)	1.04(0.89,1.21)	1.17(0.93,1.48)
Total stroke	1224	0.86(0.75,1.00)	0.90(0.78,1.04)	0.71(0.53,0.95)	0.76(0.60,0.98)	0.85(0.62,1.17)
Ischemic stroke	741	0.88(0.74,1.05)	0.93(0.78,1.10)	0.67(0.48,0.94)	0.74(0.55,1.00)	0.87(0.59,1.27)
Hemorrhagic stroke	217	0.75(0.53,1.07)	0.73(0.49,1.10)	0.85(0.52,1.39)	0.81(0.53,1.23)	0.76(0.46,1.24)
Total CVD 1	2627	1.05(0.95,1.15)	1.08(0.98,1.19)	0.88(0.71,1.08)	0.94(0.79,1.12)	1.03(0.82,1.29)
Total CVD 2	3917	1.11(1.03,1.19)	1.14(1.04,1.24)	0.92(0.77,1.09)	0.98(0.84,1.14)	1.08(0.88,1.34)
Protein						
Total CHD	1516	1.01(1.92,1.13)	1.04(0.95,1.13)	0.93(0.84,1.03)	0.99(0.72,1.35)	
Nonfatal MI	1140	1.00(0.90,1.10)	1.02(0.93,1.12)	0.94(0.84,1.05)	0.98(0.76,1.27)	
Coronary death	432	1.03(0.89,1.19)	1.06(0.92,1.23)	0.92(0.78,1.09)	1.00(0.63,1.58)	
CABG and PCI	2055	1.03(0.96,1.10)	1.06(0.99,1.14)	0.95(0.86,1.03)	1.01(0.75,1.35)	
Total stroke	1224	0.89(0.82,0.98)	0.90(0.83,0.98)	0.88(0.79,0.98)	0.89(0.76,1.05)	
Ischemic stroke	741	0.91(0.81,1.02)	0.92(0.83,1.03)	0.88(0.77,1.01)	0.90(0.70,1.16)	
Hemorrhagic stroke	217	0.91(0.75,1.10)	0.87(0.69,1.10)	0.97(0.78,1.20)	0.91(0.70,1.16)	
Total CVD 1	2627	0.96(0.90,1.03)	0.99(0.93,1.05)	0.92(0.85,1.00)	0.96(0.76,1.20)	
Total CVD 2	3917	1.00(0.84,1.06)	1.02(0.97,1.08)	0.94(0.87,1.01)	0.98(0.77,1.25)	
Protein Density						
Total CHD	1516	0.85(0.75,0.97)	0.86(0.76,0.97)	0.87(0.78,0.98)	0.87(0.77,0.99)	
Nonfatal MI	1140	0.89(0.78,1.02)	0.90(0.79,1.03)	0.91(0.80,1.03)	0.91(0.80,1.04)	
Coronary death	432	0.74(0.59,0.93)	0.76(0.61,0.94)	0.77(0.63,0.95)	0.77(0.62,0.97)	
CABG and PCI	2055	0.94(0.84,1.05)	0.95(0.85,1.05)	0.96(0.88,1.06)	0.97(0.86,1.08)	
Total stroke	1224	0.94(0.84,1.06)	0.94(0.83,1.07)	0.94(0.82,1.07)	0.94(0.82,1.06)	

CVD	No. Incident Cases	From Prentice (2010)	Based on $\tilde{\beta}$ in (6)		Based on $\hat{\beta}$ in (3)	
			$\tilde{\beta}$ in (6)	Biomarker Measurement Error Correlation	$\hat{\beta}$ in (3)	0.50
Ischemic stroke	741	0.97(0.83,1.12)	0.97(0.83,1.13)	0.97(0.82,1.14)	0.97(0.82,1.14)	
Hemorrhagic stroke	217	1.16(0.87,1.56)	0.87(0.69,1.10)	1.13(0.85,1.50)	1.13(0.84,1.51)	
Total CVD 1	2627	0.89(0.81,0.98)	0.90(0.82,0.99)	0.90(0.83,0.99)	0.90(0.82,1.00)	
Total CVD 2	3917	0.94(0.87,1.02)	0.95(0.88,1.03)	0.96(0.90,1.03)	0.96(0.89,1.04)	

FFQ indicates food frequency questionnaire; WHI, Women's health Initiative; CHD, coronary heart disease; MI, myocardial infarction; Total CHD comprised of nonfatal MI plus coronary death; total stroke comprised of ischemic plus hemorrhagic stroke; CVD, cardiovascular disease; Total CVD 1 comprised of total CHD plus total stroke; total CVD 2 comprised of total CVD 1 plus coronary artery bypass graft plus percutaneous coronary intervention.