A statistical model for measurement error that incorporates variation over time in the target measure, with application to nutritional epidemiology

Laurence S. Freedman,[a,b*†] Douglas Midthune,[c] Kevin W. Dodd,[c] Raymond J. Carroll[d] and Victor Kipnis[c]

Most statistical methods that adjust analyses for measurement error assume that the target exposure T is a fixed quantity for each individual. However, in many applications, the value of T for an individual varies with time. We develop a model that accounts for such variation, describing the model within the framework of a meta-analysis of validation studies of dietary self-report instruments, where the reference instruments are biomarkers. We demonstrate that in this application, the estimates of the attenuation factor and correlation with true intake, key parameters quantifying the accuracy of the self-report instrument, are sometimes substantially modified under the time-varying exposure model compared with estimates obtained under a traditional fixed-exposure model. We conclude that accounting for the time element in measurement error problems is potentially important. Copyright © 2015 John Wiley & Sons, Ltd.

Keywords: 24-hour recall; attenuation factor; calibration equations; food frequency questionnaire; recovery biomarker

1. Introduction

There is now an extensive literature on statistical methods for dealing with measurement error in a target measure (denoted T). Most of the methods assume that T is a fixed quantity for each individual and model the relationship between an error-prone measurement (denoted W) and T [1]. However, in many applications, T really varies with time.

In these circumstances, statisticians usually re-define T as the long-term average and treat the variation over time as part of the measurement error. Correlation over time between repeated values is either ignored or sometimes modeled as correlation between error terms. When the number of repeat measurements is limited, the correlation is most commonly ignored. Potentially important information is thereby lost.

Rosner *et al.* [2] did explicitly model data from a nutritional cohort study with two sets of measurements taken 4 years apart and allowed the target measure (usual intake of vitamin C) to differ at the two time points. Keogh *et al.* [3] also modeled true intake that varies with time. However, neither work distinguishes between shorter-term and longer-term instruments, nor attempts to estimate the relationship between a dietary report and the longer-term average intake.

In this paper, we model a target measure T that can change from one short-term period to the next over a limited number of periods. We aim to estimate the relationship between dietary self-reports of different types and longer-term true average intake. We describe the model within a meta-analysis framework,

[a]*Information Management Services, Inc., Rockville, MD, U.S.A.*
[b]*Gertner Institute, Biostatistics Unit, Tel Hashomer, Israel*
[c]*National Cancer Institute, Division of Cancer Prevention, Bethesda, MD, U.S.A.*
[d]*Texas A&M University, Department of Statistics, College Station, TX, U.S.A.*
*Correspondence to: Professor Laurence S. Freedman, Biostatistics Unit, Gertner Institute for Epidemiology and Health Policy Research, Sheba Medical Center, Tel Hashomer 52621, Israel.
†E-mail: lsf@actcom.co.il

so as to apply it to data from a pooling project, described in Section 2. In Section 3, we describe the statistical model, theory, and methods. In Section 4, we give a simplified example, showing the bias of estimates based on the usual model in which T is fixed, using theory and simulations. In Section 5, we describe the results of applying the method to a pooling project of validation studies. In Section 6, we discuss limitations of our approach and extensions.

Our studies span a relatively short time period (at most 2 years) and do not include health outcome data. Thus, we do not tackle here the important problem discussed by Frost and White [4] in which an error-prone time-varying exposure's relationship with a health outcome is of interest.

2. Application

The Validation Studies Pooling Project (VSPP) [5] combines data from five validation studies of dietary self-report instruments for assessing individuals' dietary intake. Three studies include men and women, while two include only women. Two types of self-report instrument, 24-hour recalls (24HRs) and food frequency questionnaires (FFQs), were validated against recovery biomarkers, that is, biomarkers known to provide nearly-unbiased measurements of dietary intake. We focus here on two such biomarkers, doubly labeled water (DLW) for energy intake and 24-hour urinary nitrogen (UN) for protein intake, and investigate reported intake of protein, energy, and their ratio, known as protein density.

The 24HRs query intake on a single day, the one previous to its administration; the FFQs enquire about average intake over the recent past, usually the previous 12 months. We wish to estimate the measurement error properties of these self-report instruments, when viewed as targeting longer-term average intake, which we will define here as the average over 12 months. Note that the biomarker measurements, like the 24HR, measure short-term intake, either the previous day (UN) or average intake over the past 10–14 days (DLW).

Each of the studies in VSPP included between 250 and 550 individuals who completed at least one FFQ, 24HR, and DLW assessment and UN assessment. The majority completed more than one 24HR and more than one UN. Each study also incorporated a sub-study where a subgroup of individuals repeated main study assessments, at varying times after the main study: 3 weeks in one study, approximately 6 months in three studies, and 10–23 months in one study.

3. Statistical Methods

3.1. Aim

We focus on estimating the attenuation factor, λ, and the correlation with truth, ρ_{WT}. They are defined as follows. If an outcome variable is related to an exposure T in a linear regression model and the coefficient of T is β_T, and we substitute for T an error-prone measure, W, then the coefficient for W, β_W, equals $\lambda\beta_T$. When W and T are normally distributed, λ can be shown to equal cov(W,T)/var(W). The correlation with truth, ρ_{WT}, is the correlation of W with T. We estimate these quantities for 24HRs and FFQs, for men and women, separately.

In addition, we also estimate calibration (prediction) equations [6] for T that include both W and personal characteristics Z related to T. In our application, Z comprises age group (<40y, 40–49, 50–59, 60–69, 70–79, $\geqslant$ 80), log of body mass index, race (African-American vs. other), and education (high school, college, and postgraduate). These covariates serve as an illustrative example of our method. For comments on choosing the set of covariates in practice, see Section 6 .

3.2. Time

In all studies but one, participants completed the FFQ at the beginning of the study; in the remaining study they completed the FFQ at the end. We set the completion of the FFQ as the common time point. Relative to this time, other instruments were completed from between 450 days beforehand to 450 days afterward. We divided this period into 10 sub-periods of 90 days each, labeled by subscript j (j=1,…,10), with the FFQ placed at the beginning of the sixth period.

3.3. Statistical models

Instead of a single error-prone measurement W, three error-prone measurements, 24HR (denoted R), FFQ (denoted Q), and biomarker (denoted M), are modeled separately in relation to true intake T. The models

for R, Q, and M form a multi-part measurement error model, commonly used in nutrition applications. The same model is used for energy, protein, and protein density, but the modeling is performed separately for each. All dietary variables are modeled on the logarithmic scale, including the unobserved true intake, as follows.

(a) *Biomarker*

Denote the kth repeat biomarker assessment in the jth sub-period ($j=1,\ldots,10$) on the ith individual of study h ($h=1,..,5$) by M_{hijk}. Then

$$M_{hijk} = T_{hijk} + \delta_{hijk} \tag{1}$$

where T_{hijk} is true intake on the day of the marker assessment and δ_{hijk} is random error independent of T. Correlation between repeat biomarker values is thus assumed to occur through the correlation between true intakes at those times, and biomarker measurements are assumed to have classical measurement error.

(b) *24-hour recall*

For the same (h,i,j,k), denote the 24HR assessment by R_{hijk}. Then

$$R_{hijk} = \beta_{R0h} + \beta_{R1h}T_{hijk} + \beta_{R2h}^{t}Z_{Rhijk} + u_{Rhi} + \varepsilon_{Rhijk} \tag{2}$$

where T_{hijk} is true intake on the day corresponding to the 24HR assessment, Z_{Rhijk} is a vector of explanatory variables, u_{Rhi} is a random error term representing subject-specific bias [7], and ε_{Rhijk} is independent random error not correlated over time. The correlation between repeat 24HRs thus occurs through the correlation between true intakes at those times and through the subject-specific bias. The covariates Z_R are included for estimating calibration equations, but are omitted for estimating attenuation factors and correlations with truth, since traditionally these measures of an instrument's quality are reported without covariate adjustment. Covariates Z_R have subscripts j and k and may vary with time, but in our application, they do not.

(c) *Food frequency questionnaire*

Denote the single FFQ assessment of individual i in study h by Q_{hi}. Then

$$Q_{hi} = \beta_{Q0h} + \beta_{Q1h}T_{hi} + \beta_{Q2h}^{t}Z_{Qhi} + u_{Qhi} + \varepsilon_{Qhi} \tag{3}$$

where T_{hi} is the average true intake over the sub-period of the FFQ administration and the three previous sub-periods, the assumed target of the FFQ. Covariates Z_Q may differ from Z_R, but in our application, they are the same; u_{Qhi} and ε_{Qhi} are random error terms representing subject-specific bias and independent random error, respectively. Model (3) is not completely identifiable when each individual has only a single FFQ, because the variances of u_{Qhi} and ε_{Qhi} cannot be estimated separately; however, the variance of their sum can be estimated.

In models (1)–(3), the error terms (δ, ε) and the subject-specific biases (u) are assumed to be mutually independent, except for u_{Rhi} and u_{Qhi}. The correlation between u_{Rhi} and u_{Qhi} and the variances of u_{Qhi},

Table I. Bias in the estimated attenuation factor and correlation with true usual intake based on a fixed-intake model when a time-varying intake model pertains – a simple example[a]

ω_ϕ	0.1			0.5			2.0		
ρ_{12}	0.9	0.5	0.1	0.9	0.5	0.1	0.9	0.5	0.1
True λ_R	0.32	0.25	0.19	0.30	0.24	0.17	0.24	0.19	0.14
Bias of $\hat{\lambda}_{RF}$ (%)[b]	2.6	3.7	4.4	12.9	16.6	22.2	52.5	66.8	90.9
Asymptotic relative bias (%)	2.6	3.3	4.5	13.2	16.7	22.7	52.6	66.7	90.9
True ρ_R	0.40	0.36	0.31	0.39	0.35	0.30	0.35	0.31	0.27
Bias of $\hat{\rho}_{RF}$ (%)[b]	0.0	−10.2	−22.4	10.0	0.9	−9.5	48.8	44.5	41.7
Asymptotic relative bias (%)	0.0	−10.5	−22.5	10.3	1.0	−9.0	48.8	44.3	41.6

[a] For details of the example, see Section 4.
[b] Relative bias: [(Mean of estimate – true value)/true value] ×100%.

ε_{Qhi}, u_{Rhi}, and ε_{Rhijk} may differ by study. In our application, the variance of δ_{hijk} is assumed constant, because insufficient replications of DLW biomarker measurements prevented study-specific estimation.

(d) *True intake*

Denote true average intake in sub-period j by T_{hij}. The distribution of T_{hij}, conditional on covariates Z_{Thij}, has first and second moments:

$$E\left(T_{hij}|Z_{Thij}\right) = \gamma_{T0h} + \gamma_{T1h}^{t} Z_{Thij}$$

$$\text{var}\left(T_{hij}|Z_{Thij}\right) = \sigma_{Thj}^{2} \qquad (4)$$

$$\text{corr}\left(T_{hij}, T_{hij'}|Z_{Thij}\right) = \rho_{jj'}$$

Furthermore, for a single day's intake within a sub-period j:

$$T_{hijk} = T_{hij} + \phi_{hijk}$$

where $\text{var}(\phi_{hijk}) = \omega_{\phi}\sigma_{Thj}^{2}$ and ϕ_{hijk} are independent of each other and of T_{hij} and Z_{Thij}. The covariates Z_T may differ from Z_R or Z_Q, but in our application, they are the same. The mean and variance of T may vary with time (i.e., with subscript j), but in our application do not; we write σ_{Thj}^{2} as σ_{Th}^{2} and Z_{Thij} as Z_{Thi}. The correlations $\rho_{jj'}$ have one of the following structures: banded Toeplitz, autoregressive model order 1 (AR(1)), compound symmetry, or degenerate (all 1s), whichever provides the best fit to the model as judged by the Akaike Information Criterion (AIC). The degenerate structure option, together with setting $\omega_{\phi} = 0$, corresponds to the fixed-intake model, and we use it to compare estimates under this model with estimates under a time-varying intake model.

Table II. Estimates of random effects (standard errors in parentheses) and goodness of fit for the time-varying and fixed-intake models without covariates.

		Protein		Protein Density	
		Time-varying		Time-varying	
Gender	Parameter	intake	Fixed-intake	intake	Fixed-intake
Men	Correlation structure	Compound symmetry	–	Compound symmetry	–
	AIC	2045.8	2070.0	−65.8	−34.1
	σ_{T1}^{2}	0.043 (0.004)	0.043 (0.006)	0.032 (0.004)	0.029 (0.004)
	σ_{T2}^{2}	0.082 (0.012)	0.065 (0.012)	0.074 (0.011)	0.066 (0.013)
	σ_{T3}^{2}	0.054 (0.006)	0.052 (0.006)	0.054 (0.005)	0.053 (0.006)
	σ_{δ}^{2}	0.015 (0.005)	0.034 (0.002)	0.009 (0.006)	0.032 (0.002)
	ω_{ϕ}	0.301 (0.090)	0	0.438 (0.128)	0
	ρ_{12}	0.873 (0.073)	1	0.885 (0.079)	1
	ρ_{13}	0.873	1	0.885	1
	ρ_{14}	0.873	1	0.885	1
	ρ_{15}	0.873	1	0.885	1
	ρ_{16}	0.873	1	0.885	1
Women	Correlation Structure	AR(1)	–	AR(1)	–
	AIC	5525.4	5603.2	682.3	747.9
	σ_{T1}^{2}	0.058 (0.006)	0.042 (0.006)	0.052 (0.006)	0.038 (0.006)
	σ_{T2}^{2}	0.101 (0.011)	0.086 (0.012)	0.080 (0.009)	0.058 (0.009)
	σ_{T3}^{2}	0.050 (0.004)	0.055 (0.007)	0.055 (0.005)	0.061 (0.008)
	σ_{T4}^{2}	0.051 (0.004)	0.038 (0.005)	0.041 (0.004)	0.023 (0.005)
	σ_{T5}^{2}	0.057 (0.004)	0.050 (0.006)	0.047 (0.005)	0.028 (0.005)
	σ_{δ}^{2}	0.007 (0.006)	0.045 (0.002)	0.009 (0.010)	0.049 (0.002)
	ω_{ϕ}	0.495 (0.109)	0	0.499 (0.162)	0
	ρ_{12}	0.918 (0.029)	1	0.792 (0.052)	1
	ρ_{13}	0.843	1	0.627	1
	ρ_{14}	0.774	1	0.497	1
	ρ_{15}	0.711	1	0.394	1
	ρ_{16}	0.652	1	0.312	1

Under model (4), the correlation between true intakes on different days in the same sub-period is $1/(1+\omega_\phi)$ and between true intakes on days in periods j and j' is $\rho_{jj'}/(1+\omega_\phi)$. Also, 24HRs and biomarkers measuring the same day's intake have covariance (conditional on covariates Z) $\beta_{R1h}\sigma^2_{Th}(1+\omega_\phi)$, compared

Table III. Estimates of regression slopes (standard errors in parentheses) and partitions of the variance of self-reports[a] for the time-varying and fixed-intake models without covariates.

Gender	Parameter	Study	Protein		Protein density	
			Time-varying intake	Fixed-intake	Time-varying intake	Fixed-intake
Men	β_{R1}	1	0.75 (0.14)	0.69 (0.11)	0.69 (0.12)	0.72 (0.11)
		2	0.24 (0.15)	0.32 (0.18)	0.32 (0.08)	0.36 (0.09)
		3	0.70 (0.07)	0.81 (0.08)	0.35 (0.05)	0.42 (0.05)
	β_{Q1}	1	0.72 (0.16)	0.68 (0.15)	0.49 (0.09)	0.52 (0.09)
		2	0.39 (0.20)	0.40 (0.19)	0.30 (0.11)	0.27 (0.09)
		3	0.41 (0.12)	0.40 (0.11)	0.28 (0.07)	0.26 (0.06)
	$\sigma^2_{uR}/\beta^2_{R1}\sigma^2_T$	1	1.5	1.9	0.7	0.6
		2	20.5	14.0	0.9	0.7
		3	0.8	0.4	1.7	1.1
	$\sigma^2_{\varepsilon R}/\beta^2_{R1}\sigma^2_T$	1	3.5	4.6	3.2	4.0
		2	43.5	30.0	13.0	11.9
		3	3.2	2.7	6.8	5.2
	$\left(\sigma^2_{uQ}+\sigma^2_{\varepsilon Q}\right)/\beta^2_{Q1}\sigma^2_T$	1	7.0	7.8	3.6	3.5
		2	13.8	16.4	5.5	7.7
		3	11.6	12.6	6.9	8.1
Woman	β_{R1}	1	0.42 (0.12)	0.47 (0.13)	0.34 (0.10)	0.35 (0.11)
		2	0.53 (0.08)	0.60 (0.11)	0.32 (0.06)	0.46 (0.09)
		3	0.61 (0.07)	0.77 (0.08)	0.36 (0.05)	0.46 (0.06)
		4	0.91 (0.15)	1.12 (0.19)	0.43 (0.14)	0.78 (0.23)
		5	0.55 (0.07)	0.63 (0.08)	0.37 (0.06)	0.62 (0.12)
	β_{Q1}	1	0.57 (0.17)	0.61 (0.18)	0.42 (0.11)	0.40 (0.10)
		2	0.12 (0.18)	0.10 (0.16)	0.56 (0.12)	0.46 (0.10)
		3	0.49 (0.13)	0.46 (0.12)	0.37 (0.08)	0.27 (0.06)
		4	0.73 (0.11)	0.83 (0.13)	0.51 (0.08)	0.68 (0.14)
		5	0.84 (0.13)	0.85 (0.14)	0.43 (0.08)	0.49 (0.11)
	$\sigma^2_{uR}/\beta^2_{R1}\sigma^2_T$	1	3.1	3.3	2.4	2.7
		2	2.3	2.0	2.9	1.7
		3	1.4	0.6	1.7	0.6
		4	0.4	0.1	2.2	0.7
		5	1.5	1.2	2.0	0.9
	$\sigma^2_{\varepsilon R}/\beta^2_{R1}\sigma^2_T$	1	11.6	13.4	10.7	14.5
		2	5.9	5.8	9.7	6.7
		3	4.9	3.0	7.6	4.5
		4	1.5	1.8	7.2	4.2
		5	5.0	4.7	8.0	5.2
	$\left(\sigma^2_{uQ}+\sigma^2_{\varepsilon Q}\right)/\beta^2_{Q1}\sigma^2_T$	1	9.0	10.8	4.1	6.6
		2	188.0	310.3	1.9	4.5
		3	10.0	10.7	3.8	7.0
		4	4.5	4.6	2.6	2.5
		5	4.8	5.3	3.4	4.4

$\sigma^2_{uR}/\beta^2_{R1}\sigma^2_T$ = ratio of subject-specific variance to "true signal" variance in 24HR.
$\sigma^2_{\varepsilon R}/\beta^2_{R1}\sigma^2_T$ = ratio of within-person variance to "true signal" variance in 24HR.
$(\sigma^2_{uQ}+\sigma^2_{\varepsilon Q})/\beta^2_{Q1}\sigma^2_T$ = ratio of subject-specific variance plus within-person variance to "true signal" variance in FFQ.

with the smaller $\beta_{R1h}\sigma^2_{Th}$ if they measure different days in the same sub-period, and the even smaller $\beta_{R1h}\sigma^2_{Th}\rho_{jj'}$ if measured in different sub-periods, j and j'.

3.4. Estimating attenuation factor λ and correlation with truth ρ_{WT}

The target exposure is defined as average true intake over 12 months. However, the timing of these 12 months differs according to self-report instrument. A 24HR reports current intake, so the target exposure is

$$T_{hi} = \sum_{j'=j-2}^{j+1} T_{hij'}/4, \tag{5}$$

the average intake during and surrounding sub-period j. An FFQ reports past intake, so the target is

$$T_{hi} = \sum_{j'=j-3}^{j} T_{hij'}/4, \tag{6}$$

the average intake during and before sub-period j. These definitions impact the estimation of attenuation factors and correlations with true intake, as indicated in the succeeding text. Other definitions of the target exposure may be used, each resulting in a different estimate of attenuation and correlation.

(a) 24-hour recall

The attenuation factor, λ_{Rh}, for a single 24HR in study h is $\mathrm{cov}(R_{hijk}, T_{hi})/\mathrm{var}(R_{hijk})$, with T_{hi} defined in (5). From model (2) without covariates Z and model (4),

$$\lambda_{Rh} = \frac{\beta_{R1h}\sigma^2_{Th}\left(1 + \rho_{j,j-2} + \rho_{j,j-1} + \rho_{j,j+1}\right)}{4\left(\beta^2_{R1h}\sigma^2_{Th}(1 + \omega_\phi) + \sigma^2_{uRh} + \sigma^2_{\varepsilon Rh}\right)}.$$

Table IV. Estimates of attenuation factors and correlations with truth for the time-varying and fixed-intake models without covariates (standard errors in parentheses): protein intake among men.

Instrument	Study	Time-varying intake model		Fixed-intake model	
		Attenuation factor	Correlation with truth	Attenuation factor	Correlation with truth
FFQ [a]	1	0.158 (0.033)	0.338 (0.069)	0.166 (0.035)	0.337 (0.067)
	2	0.160 (0.082)	0.249 (0.126)	0.146 (0.071)	0.240 (0.114)
	3	0.178 (0.050)	0.269 (0.074)	0.184 (0.049)	0.271 (0.069)
	Weighted mean	0.163 (0.026)	0.298 (0.047)	0.168 (0.026)	0.295 (0.045)
	p [d]	0.944	0.721	0.899	0.693
1 × 24HR [b]	1	0.192 (0.032)	0.380 (0.058)	0.196 (0.032)	0.367 (0.055)
	2	0.058 (0.037)	0.118 (0.074)	0.069 (0.039)	0.149 (0.084)
	3	0.246 (0.030)	0.415 (0.046)	0.302 (0.031)	0.494 (0.039)
	Weighted mean	0.182 (0.019)	0.355 (0.032)	0.207 (0.019)	0.413 (0.030)
	p	< 0.001	0.001	< 0.001	< 0.001
4 × 24HR [c]	1	0.327 (0.047)	0.495 (0.067)	0.364 (0.059)	0.500 (0.071)
	2	0.115 (0.072)	0.166 (0.103)	0.138 (0.078)	0.211 (0.117)
	3	0.452 (0.047)	0.562 (0.054)	0.593 (0.052)	0.692 (0.046)
	Weighted mean	0.344 (0.033)	0.491 (0.047)	0.423 (0.035)	0.594 (0.037)
	p	< 0.001	0.002	< 0.001	< 0.001

[a] Food frequency questionnaire.
[b] Single 24-hour recall.
[c] Average of 4 repeats of a 24-hour recall.
[d] *p*-value for heterogeneity across studies.

Similarly, for the correlation with true average intake,

$$\rho_{Rh} = \frac{\beta_{R1h}\sigma_{Th}\left(1+\rho_{j,j-2}+\rho_{j,j-1}+\rho_{j,j+1}\right)}{\sqrt{\left(\beta_{R1h}^2\sigma_{Th}^2(1+\omega_\phi)+\sigma_{uRh}^2+\sigma_{\varepsilon Rh}^2\right)\left(4+2\rho_{j-1,j-2}+2\rho_{j,j-2}+2\rho_{j+1,j-2}+2\rho_{j,j-1}+2\rho_{j+1,j-1}+2\rho_{j+1,j}\right)}}.$$

To compute these quantities for the mean of m 24HRs (m > 1), $\sigma_{\varepsilon Rh}^2$ in the aforementioned equations is replaced by $\sigma_{\varepsilon Rh}^2/m$.

(b) *Food Frequency Questionnaire*

The attenuation factor for an FFQ, λ_{Qh}, may be computed from model (3) without covariates and T_{hi} defined in (6) as follows:

$$\lambda_{Qh} = \frac{\beta_{Q1h}\sigma_{Th}^2\left(4+2\rho_{j-2,j-3}+2\rho_{j-1,j-3}+2\rho_{j,j-3}+2\rho_{j-1,j-2}+2\rho_{j,j-2}+2\rho_{j,j-1}\right)}{\beta_{Q1h}^2\sigma_{Th}^2\left(4+2\rho_{j-2,j-3}+2\rho_{j-1,j-3}+2\rho_{j,j-3}+2\rho_{j-1,j-2}+2\rho_{j,j-2}+2\rho_{j,j-1}\right)+16\sigma_{uQh}^2+16\sigma_{\varepsilon Qh}^2},$$

and the correlation with truth, ρ_{Qh}, as follows:

$$\rho_{Qh} = \frac{\sqrt{\beta_{Q1h}^2\sigma_{Th}^2\left(4+2\rho_{j-2,j-3}+2\rho_{j-1,j-3}+2\rho_{j,j-3}+2\rho_{j-1,j-2}+2\rho_{j,j-2}+2\rho_{j,j-1}\right)}}{\sqrt{\beta_{Q1h}^2\sigma_{Th}^2\left(4+2\rho_{j-2,j-3}+2\rho_{j-1,j-3}+2\rho_{j,j-3}+2\rho_{j-1,j-2}+2\rho_{j,j-2}+2\rho_{j,j-1}\right)+16\sigma_{uQh}^2+16\sigma_{\varepsilon Qh}^2}}.$$

They were estimated under the time-varying intake model or the fixed-intake model by substituting the maximum likelihood estimates of the parameters into the aforementioned equations. Variances of the estimates were obtained using the nonparametric bootstrap. Overall attenuation factors and correlations were calculated by taking weighted averages of the estimates across the five studies, with weights inversely proportional to the estimates' variances.

3.5. *Estimating the model parameters*

Parameters were estimated using maximum likelihood (conditional likelihood given fixed covariates Z), assuming that Q_{hi}, R_{hijk}, and M_{hijk} have a conditional multivariate normal distribution given Z_{hi} (we drop the subscript Q, R, or T assuming the three sets of covariates coincide). The conditional means, variances and covariances of Q_{hi}, R_{hijk}, and M_{hijk}, given Z_{hi}, are provided in Appendix A. Actually, if the first and second moments are correctly specified, all parameters are consistently estimated, even if the normality assumption fails. The use of the nonparametric bootstrap for inference also provides robustness to non-normality. Estimation was performed using a custom-built program written in SAS [8].

Model fit was investigated in two ways. First, AIC was used to choose the most appropriate correlation structure for true intake and to distinguish between models. Second, empirical correlations between biomarker values obtained in different sub-periods were compared with those predicted from the model (Appendix B).

3.6. *Calibration Equations*

Calibration equations predict the target average intake, T_{hi}. In a fixed-intake model, they are most naturally obtained by regressing the mean biomarker value on the reported intake and covariates Z. However, when intake varies with time, one needs to fit models (1)–(4), estimate the parameters and then derive the equations. Some technical details of the derivation for the FFQ follow.

The conditional expectation of target average intake given Q is

$$E(T_{hi}|Q_{hi}, Z_{hi}) = \lambda_{Qh}\left\{Q_{hi} - E(Q_{hi}|Z_{hi})\right\} + \gamma_{T0h} + \gamma_{T1h}^t Z_{hi}, \tag{7}$$

where

$$E(Q_{hi}|Z_{hi}) = \beta_{Q0h} + \beta_{Q1h}\left(\gamma_{T0h} + \gamma_{T1h}^t Z_{hi}\right) + \beta_{Q2h}^t Z_{hi}.$$

Formula (7) can be used to calculate the calibration equation for any single study h.

If an overall calibration equation is required for use in new studies employing a similar FFQ in a similar population to the ones used in the five studies, then one may use in place of λ_{Qh} and γ_{T0h} their weighted averages over the five studies, and the estimate of γ_{T1} from a model that sets γ_{T1h}^t equal for all studies, after testing for between-study heterogeneity. The only quantity in (7) that would remain to be determined is $E(Q_{hi}|Z_{hi})$, which could be estimated from the new study, because both Q and Z are observed.

4. A simple example

Consider the case where there is only one study (h=1), two sub-periods (j=1,2), and each individual (i=1,..,n) has two 24HRs, one in each period, and two markers that are performed in the first period, one on the same day as the 24HR. We show in Appendix C that under both the time-varying intake model and the fixed-intake model the maximum likelihood estimators of the attenuation and correlation with truth for a single 24HR can be obtained from explicit functions of the second moments, whenever those estimates fall within the parameter space, and that the estimates of attenuation and correlation with truth for the 24HR can then be estimated under each model.

To demonstrate the potential bias in the estimate based on the fixed-intake model, we present the relative bias, based on its asymptotic expectation (see Appendix C for the expression) and on simulations of finite sample estimates with parameters taking the following values: $\sigma_T^2 = 0.05; \beta_{R1} = 0.5; \sigma_u^2 = 0.01; \sigma_\varepsilon^2 = 0.05, \sigma_\delta^2 = 0.0005, \rho_{12} = 0.1, 0.5, 0.9; \omega_\phi = 0.1, 0.5, 2.0$. The parameters β_{R0} and γ_{T0} have no influence on the estimates of attenuation or correlation and are assigned arbitrarily to zero. Data were simulated under the time-varying intake model described in the previous section with the nine possible combinations of parameters for 500 individuals, and estimates of attenuation and correlation were obtained using the method described in Appendix C. Each scenario was simulated 1000 times.

Table V. Estimates of attenuation factors and correlations with truth for the time-varying and fixed-intake models without covariates (standard errors in parentheses): protein intake among women.

Instrument	Study	Time-varying intake model		Fixed-intake model	
		Attenuation factor	Correlation with truth	Attenuation factor	Correlation with truth
FFQ [a]	1	0.161 (0.045)	0.302 (0.085)	0.139 (0.039)	0.291 (0.080)
	2	0.039 (0.056)	0.069 (0.099)	0.031 (0.047)	0.057 (0.086)
	3	0.168 (0.043)	0.288 (0.072)	0.187 (0.047)	0.292 (0.071)
	4	0.231 (0.032)	0.410 (0.054)	0.218 (0.031)	0.424 (0.056)
	5	0.188 (0.028)	0.397 (0.058)	0.186 (0.028)	0.398 (0.057)
	Weighted mean	0.180 (0.017)	0.336 (0.032)	0.169 (0.016)	0.329 (0.030)
	p [d]	0.052	0.029	0.015	0.005
1 × 24HR [b]	1	0.137 (0.038)	0.241 (0.065)	0.121 (0.034)	0.238 (0.063)
	2	0.179 (0.032)	0.311 (0.049)	0.189 (0.035)	0.337 (0.056)
	3	0.194 (0.024)	0.348 (0.040)	0.283 (0.032)	0.466 (0.042)
	4	0.304 (0.039)	0.531 (0.071)	0.304 (0.040)	0.584 (0.076)
	5	0.207 (0.025)	0.341 (0.039)	0.229 (0.027)	0.381 (0.040)
	Weighted Mean	0.201 (0.015)	0.340 (0.025)	0.222 (0.015)	0.397 (0.023)
	p	0.029	0.036	0.002	0.002
4 × 24HR [c]	1	0.295 (0.072)	0.354 (0.087)	0.282 (0.078)	0.363 (0.094)
	2	0.330 (0.053)	0.422 (0.061)	0.375 (0.065)	0.475 (0.074)
	3	0.365 (0.040)	0.477 (0.049)	0.561 (0.057)	0.657 (0.051)
	4	0.456 (0.053)	0.650 (0.064)	0.565 (0.079)	0.797 (0.090)
	5	0.389 (0.039)	0.468 (0.047)	0.472 (0.053)	0.547 (0.055)
	Weighted mean	0.376 (0.026)	0.476 (0.031)	0.463 (0.029)	0.580 (0.030)
	p	0.264	0.022	0.019	0.003

FFQ, food frequency questionnaire.
[a] Food frequency questionnaire.
[b] Single 24-hour recall.
[c] Average of four repeats of a 24-hour recall.
[d] *p*-value for heterogeneity across studies.

Results are shown in Table I. Biases can be appreciable depending on the values of ρ_{12} and ω_ϕ. In general, there are two sources of the bias. The first is that $\text{Cov}(\bar{R}_{i\cdot}, \bar{M}_{i1\cdot})$ overestimates the covariance between the self-report and true usual intake, because of the first self-report and marker being performed on the same day. The second is that $\text{Cov}(M_{i11}, M_{i12})$ overestimates the variance of true usual (average) intake, because of both markers being performed in the same sub-period. The attenuation factor $\hat{\lambda}_{RF}$ is overestimated if $\omega_\phi > 0$. The correlation with true usual intake $\hat{\rho}_{RF}$ may be overestimated or underestimated depending on the values of ω_ϕ ($\omega_\phi > 0$) and $\rho_{12}(0 < \rho_{12} < 1)$.

5. Application to Validation Studies Pooling Project

5.1. Model fit and parameter estimates

We present results of applying our method to intakes of protein and protein density for men and women separately. Models (1)–(4) were first fitted without covariates Z. The correlation structure that led to the lowest AIC was compound symmetry for men and AR(1) for women. The fixed-intake model ($\rho_{j,j'} = 1, \omega_\phi = 0$) was also fit. Table II compares estimates of parameters related to random effects from the two models.

The time-varying intake model provided a better fit than the fixed-intake model, as seen from the differences in AIC between the two models. The estimate of biomarker within-person variance was considerably smaller under the time-varying model compared with the fixed-intake model, because under the former model, this variance is estimated only from biomarker values repeated within the same sub-period. Estimated correlations between average intakes in different sub-periods were quite high, except for protein density among women more than 6 months apart.

Table III displays estimates of the regression coefficients under the same two models. The estimated coefficients of the 24HR variable were mostly smaller under the time-varying model than under the fixed-intake model, especially for women reporting protein density. However, this trend was not seen with coefficients of the FFQ. Table III also shows ratios of error components of variance to the 'signal' component in self-reported intake. The larger are these ratios, the smaller will be the attenuation factors and correlations with truth. In most cases, the ratios for the 24HR increased when changing from the fixed-intake to the time-varying intake model, whereas for the FFQ, the ratios decreased. The increases

Table VI. Estimates of attenuation factors and correlations with truth for the time-varying and fixed-intake models without covariates (standard errors in parentheses): protein density intake among men.

Instrument	Study	Time-varying intake model		Fixed-intake model	
		Attenuation factor	Correlation with truth	Attenuation factor	Correlation with truth
FFQ [a]	1	0.411 (0.067)	0.449 (0.069)	0.425 (0.070)	0.470 (0.070)
	2	0.479 (0.160)	0.378 (0.124)	0.427 (0.148)	0.339 (0.111)
	3	0.418 (0.094)	0.342 (0.075)	0.418 (0.093)	0.331 (0.069)
	Weighted mean	0.420 (0.052)	0.396 (0.048)	0.423 (0.052)	0.390 (0.045)
	p [d]	0.927	0.551	0.998	0.321
1 X 24HR [b]	1	0.248 (0.038)	0.414 (0.057)	0.252 (0.037)	0.425 (0.055)
	2	0.187 (0.051)	0.244 (0.062)	0.206 (0.055)	0.271 (0.064)
	3	0.258 (0.042)	0.303 (0.045)	0.328 (0.045)	0.372 (0.044)
	Weighted mean	0.240 (0.025)	0.328 (0.032)	0.266 (0.026)	0.365 (0.031)
	p	0.438	0.132	0.206	0.187
4 X 24HR [c]	1	0.457 (0.058)	0.562 (0.063)	0.538 (0.080)	0.621 (0.076)2
	2	0.514 (0.121)	0.404 (0.091)	0.597 (0.145)	0.461 (0.101)
	3	0.529 (0.078)	0.433 (0.058)	0.709 (0.089)	0.546 (0.058)
	Weighted mean	0.481 (0.048)	0.483 (0.042)	0.612 (0.055)	0.554 (0.042)
	p	0.711	0.229	0.362	0.435

FFQ, food frequency questionnaire.
[a] Food frequency questionnaire.
[b] Single 24-hour recall.
[c] Average of four repeats of a 24-hour recall.
[d] *p*-value for heterogeneity across studies.

		Time-varying intake model		Fixed-intake model	
Instrument	Study	Attenuation factor	Correlation with truth	Attenuation factor	Correlation with truth
FFQ [a]	1	0.370 (0.094)	0.396 (0.097)	0.330 (0.082)	0.362 (0.085)
	2	0.509 (0.095)	0.535 (0.098)	0.398 (0.081)	0.428 (0.079)
	3	0.454 (0.100)	0.408 (0.086)	0.470 (0.098)	0.355 (0.069)
	4	0.442 (0.061)	0.476 (0.065)	0.417 (0.056)	0.532 (0.072)
	5	0.426 (0.076)	0.428 (0.074)	0.375 (0.068)	0.430 (0.074)
	Weighted Mean	0.439 (0.036)	0.450 (0.039)	0.396 (0.033)	0.422 (0.035)
	p [d]	0.890	0.805	0.836	0.420
$1 \times 24HR$ [b]	1	0.164 (0.051)	0.240 (0.073)	0.156 (0.045)	0.235 (0.068)
	2	0.179 (0.040)	0.245 (0.052)	0.232 (0.047)	0.327 (0.057)
	3	0.205 (0.037)	0.279 (0.046)	0.356 (0.040)	0.405 (0.045)
	4	0.173 (0.055)	0.278 (0.089)	0.215 (0.057)	0.411 (0.102)
	5	0.186 (0.032)	0.270 (0.046)	0.226 (0.035)	0.375 (0.052)
	Weighted Mean	0.184 (0.021)	0.263 (0.031)	0.235 (0.020)	0.356 (0.027)
	p	0.955	0.970	0.055	0.279
$4 \times 24HR$ [c]	1	0.367 (0.094)	0.359 (0.096)	0.387 (0.118)	0.369 (0.104)
	2	0.370 (0.077)	0.353 (0.069)	0.497 (0.091)	0.479 (0.078)
	3	0.433 (0.073)	0.405 (0.063)	0.789 (0.099)	0.604 (0.061)
	4	0.344 (0.095)	0.392 (0.113)	0.459 (0.117)	0.600 (0.143)
	5	0.387 (0.057)	0.390 (0.060)	0.505 (0.072)	0.561 (0.074)
	Weighted Mean	0.384 (0.041)	0.382 (0.043)	0.534 (0.042)	0.537 (0.037)
	p	0.914	0.961	0.063	0.325

Table VII. Estimates of attenuation factors and correlations with truth for the time-varying and fixed-intake models without covariates (standard errors in parentheses): protein density intake among women.

FFQ, food frequency questionnaire.
[a] Food frequency questionnaire.
[b] Single 24-hour recall.
[c] Average of four repeats of a 24-hour recall.
[d] *p*-value for heterogeneity across studies.

were particularly large for women reporting protein density on a 24HR and were governed largely by the corresponding decrease in the regression coefficients.

5.2. Attenuation factors and correlations with truth

Tables IV–VII compare estimated attenuation factors and correlations with truth for the time-varying and fixed-intake models for protein and protein density intakes in men and women. The same trends are seen in each table. Under the time-varying model, FFQ estimates were slightly larger and 24HR estimates were smaller. For protein intake (Tables IV and V), despite this trend, under both models, a single 24HR had slightly higher estimated attenuation factors and correlations than an FFQ, and the mean of four 24HRs had markedly higher estimates than an FFQ. However, for protein density (Tables VI and VII), the fixed-intake model gave estimates for the mean of four 24HRs that were higher than that for an FFQ, while the time-varying intake model gave estimates that were comparable for the two instruments.

5.3. Calibration equations

Models (1)–(4) with covariates Z were fitted with the restriction that γ^t_{T1h} did not differ across studies. Table VIII displays an example of the results, with estimates of γ_{T1}, λ_{Qh} for each study, and the weighted average of γ_{T0h}, for FFQ-reported intakes of protein and protein density for women. Values of λ_{Qh} appeared quite homogeneous across studies, lending support to using the weighted average value in new studies. Also, estimates of λ_{Qh} were similar to those obtained without covariates Z in the model (compare with Tables V and VII). Age, race, and body mass index strongly predicted protein intake, but only age and race strongly predicted protein density. Tests for between-study heterogeneity of γ^t_{T1h} were not formally significant (although there was an indication of heterogeneity for protein density). Overall,

Table VIII. Calibration equations[a] for usual intakes of protein and protein density in women, using FFQ-reported protein and protein density, respectively.

Covariate	Protein Regression coefficient (SE) p		Protein density Regression coefficient (SE) p	
Intercept (weighted average of γ_{T0h})	4.441 (0.115)		2.596 (0.118)	
λ_{Qh}: Study 1	0.132 (0.041)		0.419 (0.092)	
Study 2	0.039 (0.050)		0.528 (0.094)	
Study 3	0.149 (0.040)		0.425 (0.102)	
Study 4	0.191 (0.029)		0.426 (0.061)	
Study 5	0.162 (0.025)		0.412 (0.076)	
Weighted average	0.152 (0.016)		0.436 (0.037)	
Heterogeneity[b]		0.116		0.886
γ_{T1}:				
Age (years): 40 vs. 50–59	0.059 (0.034)		0.027 (0.035)	
40–49 vs. 50–59	0.023 (0.027)		−0.033 (0.029)	
60–69 vs. 50–59	0.021 (0.030)	<0.001	0.048 (0.032)	<0.004
70–79 vs. 50–59	−0.048 (0.034)		0.037 (0.036)	
80 vs. 50–59	−0.177 (0.043)		-0.051 (0.044)	
Log BMI	0.371 (0.035)	<0.001	−0.012 (0.036)	<0.744
Race: African-American vs. other[d]	−0.116 (0.019)	<0.001	−0.072 (0.020)	<0.001
Education: High school vs. college	−0.030 (0.018)	<0.001	−0.016 (0.018)	0.220
Postgrad vs. college	0.014 (0.016)		0.018 (0.016)	
Heterogeneity[c]		0.579		0.054

[a]See Equation 7 in text.
[b]*p*-value for heterogeneity of adjusted 'attenuation coefficient' for FFQ across studies.
[c]*p*-value for heterogeneity of regression coefficients for other(γ_{T1h}).
[d]Other includes non-Hispanic white people.

the results support using these calibration equations for new studies with a similar FFQ and population to those in the VSPP.

6. Discussion

We have described a measurement error model for dietary intake that accounts for the timing of biomarker measurements in relation to self-report instruments, for the target measure of each instrument, and for the time between repeat biomarker measurements. Application to dietary validation study data showed some substantial changes in the estimates of attenuation and correlation with truth for self-report instruments. Estimates of attenuation factors and correlations with truth were slightly increased for FFQs and decreased for 24HRs.

Our method required subdividing the time into sub-periods of 90 days. The length of the sub-period should be chosen with care and should be short enough to capture the variation in dietary intakes, but not so short that measurements within a sub-period are very sparse. Ultimately, the choice depends on the study design and the available data. Alternatively, one could try to model the true intake in continuous time. Such an approach may improve our parameterization in which correlations between intakes change abruptly at discrete times. However, a continuous model would require specifying weekly cyclical variations in diet, which we preferred to avoid.

Preis *et al.* [9] have reported investigating the influence of timing of repeat biomarkers on estimates of attenuation factor and correlation with truth. Using data from two of the five VSPP studies (Observing Protein and Energy Nutrition (OPEN) and Automated Multiple-pass Method (AMPM)), they adopted a fixed-intake model and applied estimates of within-person biomarker variance found in the AMPM study to that of the OPEN study. They reported that the time between repeat biomarkers impacted materially on estimated correlations, tending to decrease them when based on biomarker repeats close in time and increase them when based on repeats spaced further apart, although this was disputed by Dodd *et al.* [10]. Our results in Tables IV–VII show a different pattern, with estimated correlations for FFQs relatively unaffected by the adoption of a fixed-intake or time-varying intake model and correlations for 24HRs

that were overestimated by the fixed-intake model. This overestimation seemed particularly strong in study 3 (the AMPM Validation Study). In this study, two of the 24HRs coincided with the day of the biomarker determination, and also the time between repeat biomarkers was 10–23 months, the longest period between repeats of the five studies. Both aspects, proximity of 24HR to biomarker determination and extended time between repeat biomarker measurements, tend to spuriously increase the fixed-intake model estimates of attenuation and correlation with truth in a fixed-intake model and receive proper treatment only in a time-varying model.

Choosing which covariates to include in a calibration equation is complex. A covariate appearing in the health outcome model should also be included in the calibration model, unless it is uncorrelated with both true intake and dietary measurement error [1]. However, there is a debate regarding whether covariates in the calibration equation should be included in the health outcome model. If the covariate included in the calibration model is known to be independent of health outcome conditional on other covariates in the health outcome model, then it can be safely omitted from the health outcome model, but otherwise not [11]. Further research is needed to establish valid modeling principles in other settings. Body mass index is particularly problematic because, while it is an important predictor of intakes such as energy, protein, and sodium, it is unclear whether it is a confounder (that should be included in the disease model) or a mediator, which requires special methods [12], [13]. Zheng *et al.* [14] provide further discussion. Our aim here is simply to show how to use the time-varying intake model for estimating calibration equations.

This work has several implications for the design of dietary validation studies. The timing of biomarker measurements relative to the administration of the self-report instruments needs more attention. The period of the targeted long-term average intake must be defined and may be considerably longer than the usual 6–9 months. Having decided on a relevant period, the biomarker measurements and short-term self-reports (such as 24HRs) should be spread out over that period, and the FFQ administered towards the end, so as to compare it with past diet measured by biomarkers.

Because correlations with truth for multiple 24HRs appear to be somewhat lower than previously estimated, there is a need to study the combination of multiple 24HRs with FFQs to achieve higher correlations. Such combination has recently been proposed by Carroll *et al.* [15].

Allowing for variation in time of a target measure could be important in other areas of epidemiology and medicine that involve quantities measured with error. These include studies of association of physical activity with health outcomes, longitudinal studies of serum cholesterol and other biological precursors of heart disease, and interventions to affect behavioral outcome variables such as exposure to second-hand smoke. The work described in this paper, together with Rosner *et al.* [2] and Keogh *et al.* [3], represents a more general move to acknowledge the role of time in the study of measurement error and its effects.

Appendix A

Conditional means, variances and covariances of Q, R, and M, given Z, and identifiabilty of parameters

Using the same notation as in Section 3 of the main text, the conditional means, variances, and covariances of Q_{hi}, R_{hijk}, and M_{hijk}, given Z_{hi}, are as follows:

$$E\left(Q_{hi}|Z_{hi}\right) = \beta_{Q0h} + \beta_{Q1h}E(T_{hi}|Z_{hi}) + \beta_{Q2h}^t Z_{hi}$$

$$E\left(R_{hijk}|Z_{hi}\right) = \beta_{R0h} + \beta_{R1h}E(T_{hijk}|Z_{hi}) + \beta_{R2h}^t Z_{hi}$$

$$E\left(M_{hijk}|Z_{hi}\right) = E(T_{hijk}|Z_{hi}) = \gamma_{T0h} + \gamma_{T1h}^t Z_{hi},$$

$$\text{Var}\left(Q_{hi}|Z_{hi}\right) = \beta_{Q1h}^2 \sigma_{Th}^2 \left(2 + \sum_{j=3}^{6}\sum_{j'=j+1}^{6} \rho_{jj'}\right)/8 \quad + \quad \sigma_{uQh}^2 \quad + \quad \sigma_{\varepsilon Qh}^2,$$

$$\text{Var}\left(R_{hijk}|Z_{hi}\right) = \beta_{R1h}^2 \sigma_{Th}^2 \left(1 + \omega_\phi\right) \quad + \quad \sigma_{uRh}^2 \quad + \quad \sigma_{\varepsilon Rh}^2,$$

$$\text{Var}\left(M_{hijk}|Z_{hi}\right) = \sigma_{Th}^2 \left(1 + \omega_\phi\right) \quad + \quad \sigma_{\delta h}^2,$$

$$\text{Cov}\left(Q_{hi}, M_{hijk}|Z_{hi}\right) = \beta_{Q1h}\sigma_{Th}^2 \left(\sum_{j'=3}^{6} \rho_{jj'}\right)/4,$$

Table A1. Comparison of empirical with model-based estimates of the correlation between repeat biomarker measurements as a function of the time between the repeats: protein density in men and women.

| Gender | Difference in time periods | Empirical | | | | Model-based |
		Total no. individuals	Estimated correlation	95% CI lower limit	95% CI upper limit	Estimated correlation
Men	0	500	0.591	0.531	0.645	0.626
	1	19	0.604	0.207	0.830	0.554
	2	16	0.706	0.323	0.890	0.554
	3	12	0.831	0.490	0.951	0.554
	4	9	0.880	0.521	0.975	0.554
	5	32	0530	0.222	0.741	0.554
	6	0	—	—	—	0.554
	7	8	0.480	−0.339	0.885	0.554
Women	0	550	0.604	0.548	0.654	0.602
	1	109	0.367	0.192	0.519	0.477
	2	126	0.217	0.043	0.377	0.377
	3	20	0.120	−0.341	0.534	0.299
	4	11	0.772	0.321	0.938	0.237
	5	20	0.483	0.051	0.762	0.188
	6	1	—	—	—	0.149
	7	12	−0.259	−0.725	0.370	0.118

$$\text{Cov}\left(Q_{hi}, R_{hijk} | Z_{hi}\right) = \beta_{Q1h}\beta_{R1h}\sigma_{Th}^2\left(\sum_{j'=3}^{6}\rho_{jj'}\right)/4 + \text{Cov}\left(u_{Qhi}, u_{Rhi}\right),$$

$$\text{Cov}\left(R_{hijk}, R_{hij'k'} | Z_{hi}\right) = \beta_{R1h}^2\sigma_{Th}^2\rho_{jj'} + \sigma_{uRh}^2,$$

$$\text{Cov}\left(R_{hijk}, M_{hij'k'} | Z_{hi}\right) = \beta_{R1h}\sigma_{Th}^2\left(\rho_{jj'} + I[D_{Rhijk} = D_{Mhij'k'}]\omega_\phi\right)$$

$$\text{Cov}\left(M_{hijk}, M_{hij'k'} | Z_{hi}\right) = \sigma_{Th}^2\rho_{jj'},$$

where $\rho_{jj'} = 1$ if $j = j'$, I[c] is the indicator function and D_{Rhijk} and $D_{Mhij'k'}$ are the days (time units) for which R_{hijk} and $M_{hij'k'}$ measure intake, respectively. Whether or not the parameters are identifiable depends on the data available and on the parameterization of the correlation matrix $\{\rho_{jj'}\}$. If we have repeat biomarker data, M, in all combinations of periods j and j', then we can estimate the $\rho_{jj'}$ individually. If the repeats are restricted to certain combinations of j and j', then we may still be able to estimate all $\rho_{jj'}$ if the structure of the correlation matrix is suitably parsimonious, for example, autoregressive of order 1. The term ω_ϕ can be estimated if we have recalls R and biomarkers M that are measured on the same day, and other combinations of R and M that are measured in the same period j, but not on the same day.

Appendix B

Comparison of model-based estimates with empirical estimates of correlations between repeat biomarkers

One way of checking the goodness of fit of models (1)–(4) is to compare the empirical correlations between repeated biomarker measurements and their model-based estimates.

Because there were a limited number of repeat biomarker measurements, and they were not performed in all the possible combinations of periods j and j' (j, j' = 1,....,10), we consider the pairwise correlations as a function of the time difference, |j-j'|.

As in the modeling, we assumed that these pairwise correlations did not differ across studies. Accordingly, the empirical correlations were calculated for each study and then combined in a weighted average. Times between repeats differed across studies, so one study would contribute information on |j-j'|=0,

others would contribute to |j-j'|=1, 2, or 3, and yet another study only to |j-j'|=3 up to 7. The model-based correlations were calculated according to the following formula:

$$co\hat{r}r(M_{hijk}, M_{hij'k'}) = \hat{\sigma}_T^2\hat{\rho}_{jj'} / \left\{ \hat{\sigma}_T^2 \left(1 + \hat{\omega}_\phi \right) + \hat{\sigma}_\delta^2 \right\}, \text{ where } \hat{\sigma}_T^2 \text{ is the mean of the estimates } \hat{\sigma}_{Th}^2$$

across studies.

As an example of the results, we present in Table A1 the comparisons for protein density, for men and women separately. It can be seen that the data are rather sparse except for |j-j'|=0 for men and |j-j'|=0 up to 2 for women. The model-based correlations for |j-j'|=0 agree very well with the empirical estimates. For other time differences, the model-based estimates fall within the 95% confidence intervals of the empirical estimates, except for |j-j'|=4 for women, where the empirical estimate seems to be a rogue value, larger even than the correlation for |j-j'|=0. Thus, overall, the model appears to provide a reasonable fit to the data based on the comparison of empirical with model-based estimates of correlation. A similar picture was obtained when absolute protein intake was considered.

Appendix C

A simple example

The equations for the conditional means and variances that are provided in Appendix 1 simplify to the following in the simple example presented in the Section 4 of the main text entitled "A simple example". In that example, there is only one study (h=1), 2 sub-periods (j=1,2), no covariates Z, no FFQ, and each individual has two 24HRs, one in sub-period 1 and one in sub-period 2, and two markers, both in the first sub-period, with one on the same day as the first 24HR. Since there is only one study, the subscript h is suppressed throughout; and since there are no repeat 24HRs within the same sub-period, the subscript k is suppressed for R.

$$E(\bar{R}_{i.}) = \beta_{R0} + \beta_{R1} \gamma_{T0}$$

$$E(\bar{M}_{i1.}) = \gamma_{T0}$$

$$Var(\bar{R}_{i.}) = \beta_{R1}^2 \sigma_T^2(1 + \omega_\phi) + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2/2$$

$$Var(\bar{M}_{i1.}) = \sigma_T^2(1 + \omega_\phi) + \sigma_\delta^2/2$$

$$Cov(R_{i1}, R_{i2}) = \beta_{R1}^2 \sigma_T^2 \rho_{12} + \sigma_{uR}^2$$

$$Cov(R_{i1}, M_{i11}) = \beta_{R1} \sigma_T^2(1 + \omega_\phi)$$

$$Cov(R_{i1}, M_{i12}) = \beta_{R1} \sigma_T^2$$

$$Cov(R_{i2}, \bar{M}_{i1.}) = \beta_{R1} \sigma_T^2 \rho_{12}$$

$$Cov(M_{i11}, M_{i12}) = \sigma_T^2$$

There are nine equations for the nine unknown parameters: γ_{T0}, β_{R0}, β_{R1}, σ_T^2, ω_ϕ, ρ_{12}, σ_{uR}^2, $\sigma_{\varepsilon R}^2$, and σ_δ^2. These equations provide unique solutions for the parameters and when the solutions fall within the parameter space, they coincide with the maximum likelihood estimates under the assumptions of normality stated in the main text.

The expressions for the attenuation factor and correlation with truth for the 24HR in this case are as follows:

$$\lambda_{RT} = \frac{\beta_{R1}\sigma_T^2(1 + \rho_{12})}{2(\beta_{R1}^2\sigma_T^2(1 + \omega_\phi) + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2)},$$

$$\rho_{RT} = \frac{\beta_{R1}\sigma_T \sqrt{1 + \rho_{12}}}{\sqrt{2(\beta_{R1}^2\sigma_T^2(1 + \omega_\phi) + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2)}}.$$

Their estimates are obtained by plugging in the estimates of each parameter obtained from the aforementioned equations.

These estimates may be compared with those derived from the fixed-intake model. Under that model, $\omega_\phi = 0$ and $\rho_{12} = 1$; the estimating equations are as follows:

$$E(\bar{R}_{i.}) = \beta_{R0} + \beta_{R1}\gamma_{T0}E(\bar{M}_{i1.}) = \gamma_{T0}$$
$$\text{Var}(\bar{R}_{i.}) = \beta_{R1}^2\sigma_T^2 + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2/2$$
$$\text{Var}(\bar{M}_{i1.}) = \sigma_T^2 + \sigma_\delta^2/2$$
$$\text{Cov}(R_{i1}, R_{i2}) = \beta_{R1}^2\sigma_T^2 + \sigma_{uR}^2$$
$$\text{Cov}(\bar{R}_{i.}, \bar{M}_{i1.}) = \beta_{R1}\sigma_T^2$$
$$\text{Cov}(M_{i11}, M_{i12}) = \sigma_T^2$$

Here, again, the equations provide unique solutions for the parameters and when the solutions fall within the parameter space, they coincide with the maximum likelihood estimates. The expressions for the attenuation factor and correlation with truth for the 24HR are as follows:

$$\lambda_{RF} = \frac{\beta_{R1}\sigma_T^2}{\beta_{R1}^2\sigma_T^2 + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2},$$

$$\rho_{RF} = \frac{\beta_{R1}\sigma_T}{\sqrt{\beta_{R1}^2\sigma_T^2 + \sigma_{uR}^2 + \sigma_{\varepsilon R}^2}}.$$

Their estimates are obtained by plugging in the estimates of each parameter obtained from the equations above. It can then be shown that the estimators of λ_{RF} and ρ_{RF} have asymptotic expectations that are not equal to λ_{RT} and ρ_{RT}, respectively, as one would wish, but instead $\lambda_{RT}\frac{1+\rho_{12}+\omega_\phi/2}{1+\rho_{12}}$ and $\rho_{RT}\frac{1+\rho_{12}+\omega_\phi/2}{\sqrt{2(1+\rho_{12})}}$. Note that when $\omega_\phi = 0$ and $\rho_{12} = 1$, both multiplicative factors equal 1. In general, there are two sources of the biases in these estimates. The first is that $\text{Cov}(\bar{R}_{i.}, \bar{M}_{i1.})$ overestimates the covariance between the self-report and true usual intake, because of the first self-report and marker being performed on the same day. The second is that $\text{Cov}(M_{i11}, M_{i12})$ overestimates the variance of true usual intake, because of both markers being performed in the same sub-period. The attenuation factor $\hat{\lambda}_{RF}$ is overestimated if $\omega_\phi > 0$. The correlation with true usual intake $\hat{\rho}_{RF}$ may be overestimated or underestimated depending on the values of ω_ϕ ($\omega_\phi > 0$) and ρ_{12} ($0 < \rho_{12} < 1$). Table I of the main text shows the magnitude of the bias for selected values of ω_ϕ and ρ_{12}.

References

1. Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM. *Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition.* Chapman and Hall: Boca Raton, FL, 2006.
2. Rosner BA, Michels KB, Chen YH Day NE. Measurement error correction for nutritional exposures with correlated measurement error: use of the method of triads in a longitudinal setting. *Statistics in Medicine* 2008; **27**(18):3466–3489.
3. Keogh RH, White IR, Rodwell SA. Using surrogate biomarkers to improve measurement error models in nutritional epidemiology. *Statistics in Medicine* 2013; **32**(22):3838–3861.

4. Frost C, White IR. The effect of measurement error in risk factors that change over time in cohort studies: do simple methods over-correct for regression dilution. *International Journal of Epidemiology* 2005; **34**(6):1359–1368.

5. Freedman LS, Commins JM, Moler JE, Arab L, Baer DJ, Kipnis V, Midthune D, Moshfegh AJ, Neuhouser ML, Prentice RL, Schatzkin A, Spiegelman D, Subar AF Tinker LF, Willett W. Pooled results from five validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014; **180**(2):172–188.

6. Prentice RL, Mossavar-Rahmani Y, Huang Y, Van Horn L, Beresford SAA, Caan B, Tinker L, Schoeller D, Bingham S, Eaton CB, Thomson C, Johnson KC, Ockene J, Sarto G, Heiss G, Neuhouser ML. Evaluation and comparison of food records, recalls and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology* 2011; **174**(5):591–603.

7. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard-Barbash R, Troiano R, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. The structure of dietary measurement error: results of the OPEN biomarker study. *American Journal of Epidemiology* 2003; **158**:14–21.

8. *SAS Institute Inc.: Statistical Analysis System (SAS) software*, Version 9.2. SAS Institute Inc. Cary: NC,US.

9. Preis SR, Spiegelman D, Zhao BB, Moshfegh A, Baer DJ, Willett WC. Application of a repeat-measure biomarker measurement error model to two validation studies: examination of the effect of within-person variation in biomarker measurements. *American Journal of Epidemiology* 2011; **173**(6):683–694.

10. Dodd KW, Midthune D, Kipnis V. Letter to the editor re: application of a repeat-measure biomarker measurement error model to two validation studies: examination of the effect of within-person variation in biomarker measurements. *American Journal of Epidemiology* 2012; **175**(1):84–85.

11. Kipnis V, Midthune D, Buckman DW, Dodd KW, Guenther PM, Krebs-Smith SM, Subar AF, Tooze JA, Carroll RJ, Freedman LS. Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics* 2009; **65**(4):1003–1010.

12. Freedman LS, Midthune D, Carroll RJ, Tasevska N, Schatzkin A, Mares J, Tinker L, Potischman N, Kipnis V. Using regression calibration equations that combine self-reported intake and biomarker measures to obtain unbiased estimates and more powerful tests of dietary associations. *American Journal of Epidemiology* 2011; **174**(11):1238–1245.

13. Prentice RL, Huang Y. Measurement error modeling and nutritional epidemiology association analyses. *Canadian Journal of Statistics* 2011; **39**:498-509.

14. Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhouser ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, Manini T, LaCroix AZ, Prentice RL. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *American Journal of Epidemiology* 2014; **180**(5):526–535.

15. Carroll RJ, Midthune D, Subar AF, Shumakovich M, Freedman LS, Thompson FE, Kipnis V. Taking advantage of the strengths of two different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *American Journal of Epidemiology* 2012; **175**(4):340–347.