

National Cancer Institute

measurementERRORwebinar series

**Assessing diet-health relationships using a short-term unbiased dietary instrument: focus on risk models with multiple dietary components**

Victor Kipnis, PhD  
National Cancer Institute, USA

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES  
National Institutes of Health

measurementERRORwebinar series



*This series is dedicated to the memory of Dr. Arthur Schatzkin*

In recognition of his internationally renowned contributions to the field of nutrition epidemiology and his commitment to understanding measurement error associated with dietary assessment.

**Presenters and Collaborators**

Sharon Kirkpatrick  
*Series Organizer*

Regan Bailey	Laurence Freedman	Douglas Midthune
Dennis Buckman	Patricia Guenther	Amy Subar
Raymond Carroll	Victor Kipnis	Fran Thompson
Kevin Dodd	Susan Krebs-Smith	Janet Tooze



measurementERRORwebinar series

**Learning objectives**

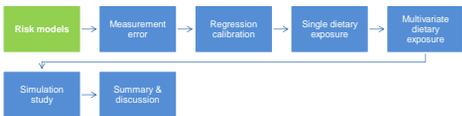
- Review statistical risk models for evaluating diet-health relationships in nutritional epidemiology
- Learn application of regression calibration to correct for measurement error in a single dietary exposure when diet is assessed by repeat administration of a short-term unbiased instrument
- Learn application of a new methodology to carry out regression calibration in risk models with multiple dietary components (some of which are episodically consumed) measured by repeat administration of a short-term unbiased instrument

measurementERRORwebinar series

**Outline**

- Risk models in nutritional epidemiology
- Dietary measurement error
- Regression calibration using repeat short-term unbiased measurements:
  - Single dietary component
  - New methodology for multivariate extension
- Simulation study
- Summary & discussion

measurementERRORwebinar series



**RISK MODELS IN NUTRITIONAL EPIDEMIOLOGY**

measurementERRORwebinar series

Risk models in nutritional epidemiology

### Types of epidemiologic studies

- Animal experiments
- Ecological studies
- Cross-sectional studies
- Case-control studies
- Cohort studies** (main focus here)
- Randomized prevention trials

7 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: exposure

- We consider studies that relate:
 

Dietary Exposure  
(adjusted for  
covariates)

→

Health Outcome
- Dietary exposure thought to be most relevant is usual (long-term average) daily dietary intake
- Health outcome examples: continuous (e.g., blood pressure), binary (event, no event), time to event (survival analysis)

8 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: general description

- Notations:
  - $Y$  - health outcome
  - $\mathbf{T} = (T_1, \dots, T_p)^T$  - vector of dietary components
  - $\mathbf{Z} = (Z_1, \dots, Z_q)^T$  - vector of adjusting covariates
  - $r(Y | \mathbf{T}, \mathbf{Z})$  - outcome risk function
  - $\eta(\mathbf{T}, \mathbf{Z}; \boldsymbol{\alpha})$  - covariate-based predictor ( $\boldsymbol{\alpha}$  is a vector of parameters)
- Risk model:  $r(Y | \mathbf{T}, \mathbf{Z}) = \eta(\mathbf{T}, \mathbf{Z}; \boldsymbol{\alpha})$

9 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: examples

- Common risk models:
  - Linear regression** for continuous outcome (e.g., blood pressure, cholesterol level)
  - Logistic regression** for binary outcome (event, no event)
  - Cox regression** for survival analysis (time to event)

10 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: risk function (1)

- Linear regression**
  - Outcome:  $Y$  - continuous variable (e.g., blood pressure, cholesterol level, etc.)
  - Risk function: conditional expected value (mean) given covariates, i.e.,

$$r(Y | \mathbf{T}, \mathbf{Z}) = \mathbb{E}(Y | \mathbf{T}, \mathbf{Z})$$

11 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: risk function (2)

- Logistic regression**
  - Outcome: binary variable  $Y = \begin{cases} 1 & \text{if event} \\ 0 & \text{if no event} \end{cases}$
  - Risk function: logit of the probability of event (log odds of event) conditional on covariates, i.e.,

$$r(Y | \mathbf{T}, \mathbf{Z}) = \log \frac{\mathbb{P}(Y = 1 | \mathbf{T}, \mathbf{Z})}{1 - \mathbb{P}(Y = 1 | \mathbf{T}, \mathbf{Z})}$$

12 measurmentERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Risk models in nutritional epidemiology

### Risk models: risk function (3)

- **Cox regression**
  - Outcome:  $Y = t$  (time to event)
  - Risk function: log of the hazard function  $h(t | \mathbf{T}, \mathbf{Z})$  conditional on covariates, i.e.,
 

$$r(Y | \mathbf{T}, \mathbf{Z}) = \log h(t | \mathbf{T}, \mathbf{Z})$$

measurementERRORwebinar series

Risk models in nutritional epidemiology

### Risk models: risk predictor (1)

- A rather flexible risk model specifies predictor as linear over transformed covariates
 

$$\eta(\mathbf{T}, \mathbf{Z}; \boldsymbol{\alpha}) = \alpha_o + \sum_{k=1}^p \alpha_{T_k} T_k^* + \sum_{l=1}^q \alpha_{Z_l} Z_l^*$$

where for any variable  $v$ ,  $v^* = g(v; \gamma_v)$  denotes its transformed value using Box-Cox family of transformations

$$g(v; \gamma_v) = \begin{cases} (v^{\gamma_v} - 1) / \gamma_v & \text{if } \gamma_v \neq 0 \\ \log(v) & \text{if } \gamma_v = 0 \end{cases}$$

measurementERRORwebinar series

Risk models in nutritional epidemiology

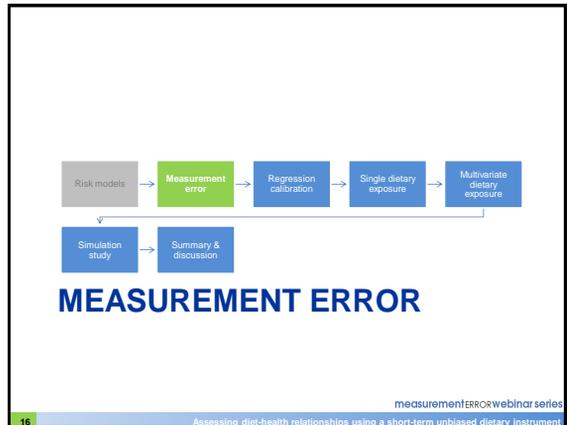
### Risk models: risk predictor (2)

- Risk model:
 

$$r(Y | \mathbf{T}, \mathbf{Z}) = \alpha_o + \sum_{k=1}^p \alpha_{T_k} T_k^* + \sum_{l=1}^q \alpha_{Z_l} Z_l^*$$
- Slope  $\alpha_{T_k}$  represents the effect of exposure  $T_k$ 
  - Due to exposure transformation, this effect depends not only on change in exposure (case of linear predictor on original scale) but also on its initial value
  - Effect of changing exposure from  $T_{k0}$  to  $T_{k1} = T_{k0} + \Delta T_k$  on risk  $r(Y | \mathbf{T}, \mathbf{Z})$  is
 

$$\alpha_{T_k} \left[ g_{T_k}(T_{k0} + \Delta T_k) - g_{T_k}(T_{k0}) \right]$$

measurementERRORwebinar series



Measurement error

### Dietary measurement error (1)

- Problem in nutritional epidemiology: true usual intakes are **unknown** and measured with error
- Fitting risk models to measured dietary exposures leads to:
  - Bias (often attenuation) of estimated exposure effect
  - Invalid significance tests in models with multiple error-prone covariates due to residual confounding
  - Reduced power to detect exposure effect

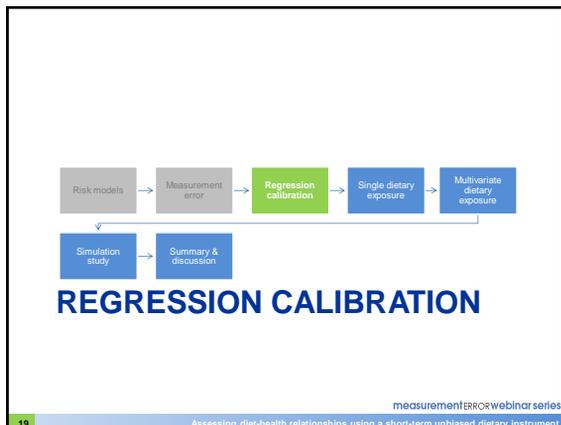
measurementERRORwebinar series

Measurement error

### Dietary measurement error (2)

- Statistical methods such as regression calibration correct biases due to measurement errors, making statistical tests valid
- These methods do not fully restore the **power** to detect a relationship which is lost due to measurement error
- It is therefore critical to use dietary assessment that, after adjustment for measurement error, leads to the minimum loss of power

measurementERRORwebinar series



### Regression calibration (1)

- Denote measured dietary intake by  $\mathbf{D} = (D_1, \dots, D_p)'$
- Assumption: measurement error in  $\mathbf{D}$  is **non-differential** with respect to health outcome  $Y$ , i.e., provides no additional information about  $Y$  beyond that in true diet
- This assumption may be justified in cohort studies where diet is usually assessed before outcome is known, but not necessarily in case-control studies due to possible recall bias when cases report their past diet differently from non-cases

measurementERRORwebinar series

### Regression calibration (2)

- Regression calibration (RC):** each error-prone covariate in a risk model is replaced with its best predictor

$$T_k^{*P}(\mathbf{D}, \mathbf{Z}) = \mathbb{E}(T_k^* | \mathbf{D}, \mathbf{Z}), \quad k = 1, \dots, p,$$

i.e., its conditional mean (expectation) given all measured dietary components  $\mathbf{D}$  and error-free covariates  $\mathbf{Z}$  in the risk model

- RC leads to (approximately) true regression slopes, i.e., true covariate effects

$$r(Y | \mathbf{D}, \mathbf{Z}) = \tilde{\alpha}_0 + \sum_{k=1}^p \alpha_{T_k} T_k^{*P}(\mathbf{D}, \mathbf{Z}) + \sum_{j=1}^q \alpha_{Z_j} Z_j^*$$

measurementERRORwebinar series

### Regression calibration (3)

- The precision of estimated slopes  $\alpha_{T_k}$  in risk model

$$r(Y | \mathbf{D}, \mathbf{Z}) = \tilde{\alpha}_0 + \sum_{k=1}^p \alpha_{T_k} T_k^{*P}(\mathbf{D}, \mathbf{Z}) + \sum_{j=1}^q \alpha_{Z_j} Z_j^*$$

and the power to detect dietary effects depend not on dietary data  $\mathbf{D}$  themselves, but on the precision of calibration predictors  $T_k^{*P}(\mathbf{D}, \mathbf{Z}) = \mathbb{E}(T_k^* | \mathbf{D}, \mathbf{Z})$  to predict true transformed intakes  $T_{k1}^*$ ,  $k = 1, \dots, p$ .

measurementERRORwebinar series

### Regression calibration (4)

- For practical reasons (relatively low cost and possible mass mailings), assessment of diet in nutritional epidemiology has been commonly done by food frequency questionnaires (FFQs)
- As you learned in webinar 10, repeat administration of more precise short-term instruments, such as 24-hour dietary recall (24HR) or food records (FR), may substantially improve the precision of the calibration predictor and the power to detect a dietary effect

measurementERRORwebinar series

### Regression calibration (5)

- Until recently, repeat application of short-term instruments as the main dietary assessment method in large studies was prohibited by a high cost of their administration and/or processing
- With advancement of new technology, repeat administration of much less expensive automated short-term instruments (e.g., web-based ASA24 developed at NCI) has become a reality

measurementERRORwebinar series

Regression calibration

### Regression calibration (6)

- Additional way to improve the precision of the calibration predictor is to consider **enhanced regression calibration**:
  - Let vector  $\mathbf{X}$  include error-free covariates  $\mathbf{Z}$  in the risk model and additional covariates  $\mathbf{C}$  that are related to true intakes but not to outcome given true intakes
  - Predictor  $\mathbb{E}(T_{ki}^* | \mathbf{D}_i, \mathbf{X}_i)$  is not only legitimate to use in regression calibration but is generally more precise than predictor  $\mathbb{E}(T_{ki}^* | \mathbf{D}_i, \mathbf{Z}_i)$

measurementERRORwebinar series

25 Assessing diet-health relationships using a short-term unbiased dietary instrument

Regression calibration

### Regression calibration (7)

- In what follows we will consider regression calibration when dietary assessment is done with repeat short-term measurements  $\mathbf{R}_{ki} = (R_{ki1}, \dots, R_{kiL})'$
- Ideally, when FFQ is also administered, it will be used in enhanced RC as part of vector  $\mathbf{C}$
- With advancement of new technology, cohort studies with repeat automated short-term instruments, alone or in combination with FFQ, are now being planned
- In what follows, we present a newly developed methodology for correcting results of such studies for measurement error

measurementERRORwebinar series

26 Assessing diet-health relationships using a short-term unbiased dietary instrument

Regression calibration

### Regression calibration (8)

- Main assumption: for person  $i$ , repeat  $j$ , short-term measurement  $R_{kij}$  is unbiased for true usual intake
- Regression calibration predictor is given by
 
$$T_{ki}^{*P} = \mathbb{E} \left[ g_{T_i} (T_{ki}) | \mathbf{R}_i, \mathbf{X}_i \right] = \mathbb{E} \left[ g_{T_i} \left\{ \mathbb{E} (R_{kij} | i) \right\} | \mathbf{R}_i, \mathbf{X}_i \right]$$
- Short of averaging an infinite number of repeat measurements, evaluation of conditional means in the above formula requires modeling of  $R_{kij}$
- Having a model, expectations can be evaluated as integrals over corresponding distributions

measurementERRORwebinar series

27 Assessing diet-health relationships using a short-term unbiased dietary instrument

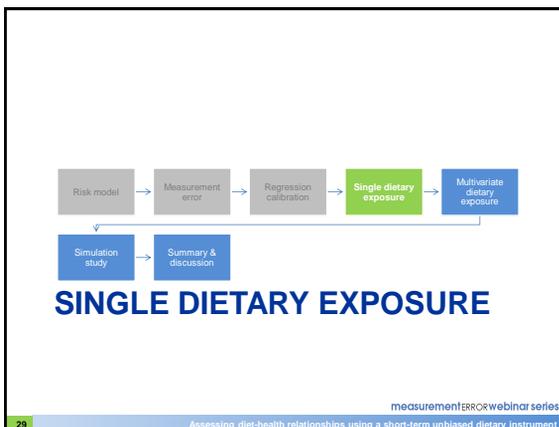
Regression calibration

### Regression calibration (9)

- Methodology below is developed for any unbiased repeat short-term measurements
- This methodology is demonstrated using 24HR
- Working assumption:** 24HR is unbiased in reporting individual's true usual dietary intake
  - Implications of possible biases in 24HR are discussed at the end

measurementERRORwebinar series

28 Assessing diet-health relationships using a short-term unbiased dietary instrument



Single dietary exposure

### Regularly-consumed dietary components (1)

- Ideal world:** the classical measurement error model
 
$$R_j = T_i + \varepsilon_j, \quad \varepsilon_j \sim N(0; \sigma_\varepsilon^2)$$
- where the regression of  $T_i$  on  $\mathbf{X}_i$  is linear, i.e.,
 
$$T_i = \beta_0 + \beta'_X \mathbf{X}_i + u_i, \quad u_i \sim N(0; \sigma_u^2)$$
- The measurement error model is thus specified as
 
$$R_j = \beta_0 + \beta'_X \mathbf{X}_i + u_i + \varepsilon_j$$

measurementERRORwebinar series

30 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (2)

- Measurement error model

$$R_{ij} = \beta_0 + \beta'_X \mathbf{X}_i + u_i + \varepsilon_{ij}$$

is a **mixed effects linear model** which includes

- fixed** (in this case linear) **effect** of covariates defined by the population-level parameters  $(\beta_0, \beta_X)$
- random effect**  $u_i$  representing part of within-person mean not explained by covariates; it is person-specific but randomly varies across people
- within-person random error**  $\varepsilon_{ij}$  representing short-term variation

measurementERRORwebinar series

31 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (3)

- Let  $\theta$  denote parameters of the measurement error model for  $\mathbf{R}$
- RC predictor of transformed true usual intake is given by

$$T_i^{*P} = \mathbb{E}[T_i^* | \mathbf{R}_i, \mathbf{X}_i] = \int g_T(\beta_0 + \beta'_X \mathbf{X}_i + u_i) f(u_i | \mathbf{R}_i, \mathbf{X}_i; \theta) du$$

- Evaluation of this expression requires evaluation of the probability density function (pdf)

$$f(u_i | \mathbf{R}_i, \mathbf{X}_i; \theta)$$

which defines the conditional distribution of  $u$

measurementERRORwebinar series

32 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (4)

- According to Bayes' theorem

$$f(u_i | \mathbf{R}_i, \mathbf{X}_i; \theta) = \frac{f(\mathbf{R}_i | \mathbf{X}_i, u_i; \theta) f(u_i | \mathbf{X}_i; \theta)}{\int f(\mathbf{R}_i | \mathbf{X}_i, u_i; \theta) f(u_i | \mathbf{X}_i; \theta) du}$$

where, given parameters  $\theta$ , conditional pdf's on the right are defined by the distributions of  $\varepsilon_{ij}$  and  $u_i$

- When  $\theta$  is estimated by fitting the measurement error model to data,  $\hat{T}_i^{*P}$  is known as the Empirical Bayes's (EB) estimator

measurementERRORwebinar series

33 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (5)

- If the risk model is fitted on original scale ( $g_T(T) \equiv T$ ) RC predictor exists in closed form and is known as the Best Linear Unbiased Predictor (BLUP) given by

$$\hat{T}_i^{*P} = \hat{w} \bar{R}_i + (1 - \hat{w}) (\hat{\beta}_0 + \hat{\beta}'_X \mathbf{X}_i)$$

where

$$\hat{w} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}_\varepsilon^2 / J_i}, \quad \bar{R}_i = \frac{1}{J_i} \sum_{j=1}^{J_i} R_{ij}$$

- In general, RC predictor  $\hat{T}_i^{*P}$  of transformed intake is not linear, does not exist in closed form, and has to be evaluated by numerical integration

measurementERRORwebinar series

34 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (6)

- Real world:** often within-person random error in  $R_{ij}$  depends on true intake and has a skewed distribution, violating classical model assumptions
- Remedy: transformation to a scale where classical model is a good approximation, i.e.,

$$g_R(R_{ij}) = \mu_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

where:

$$\mu_i = \beta_0 + \beta'_X \mathbf{X}_i + u_i, \quad u_i \sim N(0, \sigma_u^2)$$

measurementERRORwebinar series

35 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (7)

- On the transformed scale we have

$$g_R(R_{ij}) = \beta_0 + \beta'_X \mathbf{X}_i + u_i + \varepsilon_{ij}, \quad u_i \sim N(0, \sigma_u^2), \quad \varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$$

- Measurement error model is then specified as non-linear mixed effects model

$$R_{ij} = g_R^{-1}(\beta_0 + \beta'_X \mathbf{X}_i + u_i + \varepsilon_{ij})$$

- Denoting by  $\theta$  model parameters,  $R_{ij}$  is the function

$$R_{ij} = \mathfrak{R}(\mathbf{X}_i, u_i, \varepsilon_{ij}; \theta)$$

measurementERRORwebinar series

36 Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Regularly-consumed dietary components (8)

- Since  $R_{ij}$  is unbiased for true intake on original scale
 
$$T_i = \mathbb{E}(R_{ij} | i) \equiv \mathbb{E}\{\mathcal{R}(X_i, u_i, \varepsilon_{ij}, \theta) | X_i, u_i\}$$
 or
 
$$T_i = \int \mathcal{R}(X_i, u_i, \varepsilon_{ij}, \theta) f(\varepsilon_{ij} | X_i, u_i, \theta) d\varepsilon \equiv \mathcal{T}(X_i, \mathbf{u}, \theta)$$
- RC predictor of transformed intake is EB estimator
 
$$\hat{T}_i^{RP} = \int g_T(\mathcal{T}(X_i, \mathbf{u}, \hat{\theta})) f(u_i | \mathbf{R}_i, X_i, \hat{\theta}) du$$
 where, as before, pdf  $f(u_i | \mathbf{R}_i, X_i, \hat{\theta})$  is evaluated using Bayes' theorem

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Episodic dietary components (1)

- Consider now **episodically-consumed dietary components**

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Episodic dietary components (2)

- Short-term measurements for an episodically-consumed dietary component is a **semicontinuous variable** with **excess zeros** and often **skewed** to the right positive values
- Measurement error model is the result of **two** distinct, although generally correlated **processes**:
  - One specifies binary indicator variable of short-term consumption
 
$$I_{ij} = I(R_{ij} > 0) = \begin{cases} 1 & \text{if } R_{ij} > 0 \\ 0 & \text{if } R_{ij} = 0 \end{cases}$$
  - Other specifies its positive value

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Episodic dietary components (3)

- Part I: Modeling binary indicator of consumption**
- Based on modified NCI method (webinar 8), consider continuous **latent variable**

$$\tilde{R}_{ij} = \beta_{10} + \beta'_{1X} \mathbf{X}_i + u_{1i} + \varepsilon_{1ij}, u_{1i} \sim N(0, \sigma_{u_1}^2), \varepsilon_{1ij} \sim N(0, 1)$$
 which underlies fact of consumption in period  $j$ 

$$I_{ij} = 1 \Leftrightarrow \tilde{R}_{ij} > 0$$
- Consumption probability is given by the probit model
 
$$\mathbb{P}(R_{ij} > 0 | i) = \Phi(\beta_{10} + \beta'_{1X} \mathbf{X}_i + u_{F1i})$$
 where  $\Phi$  denotes standard normal distribution function

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Episodic dietary components (4)

- Part II: Modeling amount during consumption period**
- Given consumption in period  $j$ , transformed amount is specified as mixed effects linear model
 
$$g_R(R_{ij} | R_{ij} > 0) = \beta_{20} + \beta'_{2X} \mathbf{X}_i + u_{2i} + \varepsilon_{2ij}$$
 where
 
$$u_{2i} \sim N(0, \sigma_{u_2}^2), \varepsilon_{2ij} \sim N(0, \sigma_{\varepsilon_2}^2)$$
- Person-specific random effects  $u_{1i}, u_{2i}$  in parts I and II are allowed to be correlated to induce correlation between probability to consume and consumption amount

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument

Single dietary exposure

### Episodic dietary components (5)

- Measurement error model is formally specified as non-linear mixed effects model
 
$$R_{ij} = I(\beta_{10} + \beta'_{1X} \mathbf{X}_i + u_{1i} + \varepsilon_{1ij} > 0) g_R^{-1}(\beta_{20} + \beta'_{2X} \mathbf{X}_i + u_{2i} + \varepsilon_{2ij})$$
 where
 
$$\mathbf{u} = (u_{1i}, u_{2i})' \sim N(\mathbf{0}, \Sigma_u)$$

$$\boldsymbol{\varepsilon}_y = (\varepsilon_{1ij}, \varepsilon_{2ij})' \sim N\left\{\mathbf{0}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_{\varepsilon_2}^2 \end{pmatrix}\right\}$$

measurementERRORwebinar series  
Assessing diet-health relationships using a short-term unbiased dietary instrument



Multivariate dietary exposure

### Multivariate measurement error model (5)

- Denoting by  $\theta$  model parameters, we have:
 
$$\mathbf{R}_{ij} \equiv (R_{1,ij}, \dots, R_{p,ij})^T = \mathfrak{R}(\mathbf{X}_i, \mathbf{u}_i, \varepsilon_{ij}; \theta)$$
- Multivariate true usual intake is given by
 
$$\mathbf{T}_i = \int \mathfrak{R}(\mathbf{X}_i, \mathbf{u}_i, \varepsilon_{ij}; \theta) f(\varepsilon_{ij} | \mathbf{X}_i, \mathbf{u}_i; \theta) d\varepsilon \equiv \mathfrak{T}(\mathbf{X}_i, \mathbf{u}_i; \theta)$$
- RC predictor of transformed intake is EB estimator
 
$$\hat{\mathbf{T}}_i^{EB} = \int \mathbf{g}_T(\mathfrak{T}(\mathbf{X}_i, \mathbf{u}_i; \hat{\theta})) f(\mathbf{u}_i | \mathbf{R}_i, \mathbf{X}_i; \hat{\theta}) / d\mathbf{u}$$

where, as before, pdf  $f(\mathbf{u}_i | \mathbf{R}_i, \mathbf{X}_i; \hat{\theta})$  is evaluated using Bayes' theorem

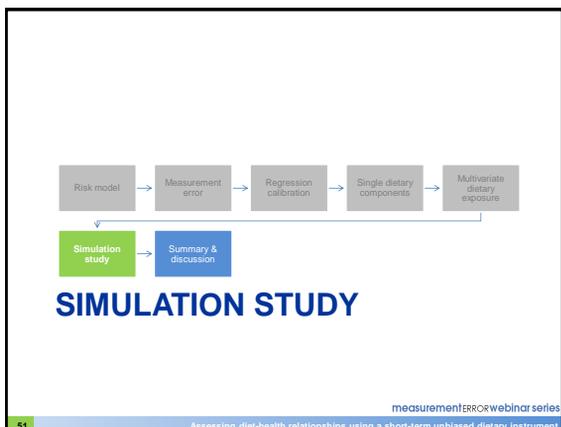
measurementERRORwebinar series

Multivariate dietary exposure

### Multivariate measurement error model (6)

- New multivariate measurement error model is a highly non-linear mixed effects model with many correlated latent variables and patterned covariance matrix  $\Sigma_\varepsilon$  with structured zeros and ones
- Currently available software for MLE or EM fitting cannot handle such models
- The model is therefore fitted using Markov Chain Monte Carlo paradigm
- Working version of SAS program has been developed by Dennis Buckman
- A user-friendly version is under construction

measurementERRORwebinar series



Simulation study

### Simulation study (1)

- Data: generated FFQ and 1000 24HRs for 2000 subjects, with distributions similar to those of red meat, white meat, total fruit, and energy in NIH-AARP calibration study of men
- True usual intakes: calculated as averages of 1000 24HRs; density intakes were calculated as ratios of true usual components to usual energy intakes
- Binary disease outcome (e.g., cancer or no cancer): generated using probability of disease defined by logistic regression based on specified odds ratios (OR) for each of the 3 true densities and energy

measurementERRORwebinar series

Simulation study

### Simulation study (2)

- Simulated cohort: 2000 subjects with 2 24HRs (took first 2 of 1000 simulated), FFQ and binary disease outcome
- Goal: estimating log OR of disease for increasing:
  - Red meat between 10 & 60 g/1000 kcal
  - White meat between 10 & 60 g/1000 kcal
  - Total fruit between 0.2 & 1.0 cups/1000 kcal
  - Total energy between 1500 and 3000 kcal
- Risk model: logistic regression with standard errors estimated by bootstrap

measurementERRORwebinar series

Simulation study

### Simulation study (3)

Reported consumption of red meat, white meat, total fruit, and total energy on two 24HRs

	Red meat (g/day)	White meat (g/day)	Total fruit (cups/day)	Total energy (kcal/day)
Mean intake (s.e.)	82.2 (0.12)	75.5 (0.12)	1.66 (0.002)	2299.6 (1.11)
Mean amount on consumption days (s.e.)	112.1 (0.13)	116.3 (0.15)	1.77 (0.002)	2299.6 (1.11)
Probability to consume	0.75	0.67	0.92	1
% who consumed:				
0 out of 2 days	14.01	18.38	2.54	0
1 out of 2 days	29.77	36.37	12.63	0
2 out of 2 days	56.22	45.25	84.83	100

measurementERRORwebinar series

Simulation study

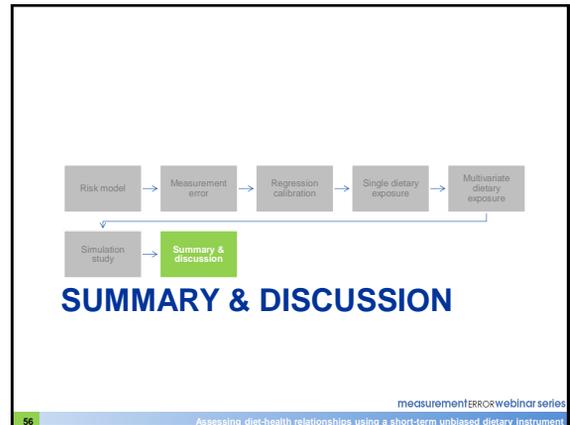
### Simulation study (4)

Mean and standard deviation of estimated log odds ratio in logistic regression of disease on red meat, white meat, total fruit, and energy

Dietary exposure	Covariates in risk model	Mean Log OR (s.e.)
Red meat density 10 – 60 g/1000 kcal	True intakes	0.4
	Mean 24HR	0.142 (0.008)
	Enhanced RC predictor	0.395 (0.023)
White meat density 10 – 60 g/1000 kcal	True intakes	0
	Mean 24HR	-0.057 (0.008)
	Enhanced RC predictor	0.006 (0.023)
Total fruit density 0.2 – 1 cups/1000 kcal	True intakes	-0.2
	Mean 24HR	-0.155 (0.007)
	Enhanced RC predictor	-0.223 (0.012)
Total energy kcal 1500 – 3000 kcal	True intakes	0.2
	Mean 24HR	0.076 (0.011)
	Enhanced RC predictor	0.224 (0.021)

measurementERRORwebinar series

55 Assessing diet-health relationships using a short-term unbiased dietary instrument



- Summary & discussion
- ### Summary (1)
- Developed methodology addresses major challenges for multivariate modeling of short-term unbiased measurements of dietary intakes by allowing
    - Excess zeros in episodically-consumed dietary components
    - Skewed distributions of positive intakes
    - Correlations among positive intakes of different dietary components
    - Correlations of facts of consumption of episodic components among themselves and with consumption amounts of other dietary components
- measurementERRORwebinar series
- 57 Assessing diet-health relationships using a short-term unbiased dietary instrument

- Summary & discussion
- ### Summary (2)
- New measurement error model is highly non-linear with multiple correlated latent variables and structured covariance matrix
  - The model is fitted using Markov Chain Monte Carlo technique implemented in SAS
  - Developed methodology allows for rigorous regression calibration correction for measurement error when repeat short-term dietary assessment methods are used as the main instrument in the study, alone or in combination with FFQ
  - New methodology allows for rather flexible risk models with covariates on transformed scales
- measurementERRORwebinar series
- 58 Assessing diet-health relationships using a short-term unbiased dietary instrument

- Summary & discussion
- ### Discussion (1)
- We considered episodically-consumed dietary components that are eventually consumed in the long run
  - What about never consumers?
    - Model could be extended to include never consumers for multiple dietary components
    - The extension is currently under development
- measurementERRORwebinar series
- 59 Assessing diet-health relationships using a short-term unbiased dietary instrument

- Summary & discussion
- ### Discussion (2)
- Developed methodology is based on the important assumption that a repeat short-term instrument is **unbiased** for true usual dietary intake
  - In considered applications, such instrument was 24HR
  - Studies with recovery biomarkers (DLW for energy, UN for protein, UK for potassium) demonstrate some bias in 24HR, suggesting possible biases in reporting of other dietary components
- measurementERRORwebinar series
- 60 Assessing diet-health relationships using a short-term unbiased dietary instrument

Summary & discussion

### Discussion (3)

- Our preliminary simulations based on OPEN biomarker study suggests that, in spite of biases, using repeat 24HRs in the developed methodology on average leads to better results than no correction for measurement error
- Using more precise short-term instruments, such as automated 24HR, in nutritional epidemiology is therefore a step forward toward better understanding of diet-health outcome relationships

measurementERRORwebinar series

51 Assessing diet-health relationships using a short-term unbiased dietary instrument

## QUESTIONS & ANSWERS

Moderator: Sharon Kirkpatrick

Please submit questions using the *Chat* function

measurementERRORwebinar series

62 Assessing diet-health relationships using a short-term unbiased dietary instrument

measurementERRORwebinar series

This concludes our Webinar series.  
Thank you for participating.

For access to series archives and supporting materials, please visit:  
[riskfactor.cancer.gov/measurementerror/sessions](http://riskfactor.cancer.gov/measurementerror/sessions)