Assessing diet-health relationships with FFQ: focus on episodically-consumed dietary components

Victor Kipnis, PhD
National Cancer Institute, USA

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 Midhune
- Dennis Buckman
- Patricia Guenther
- Amy Subar
- Raymond Carroll
- Victor Kipnis
- Fran Thompson
- Kevin Dodd
- Susan Krebs-Smith
- Janet Tooze

Learning objectives

- Review statistical risk models for evaluating diet-health relationships in nutritional epidemiology
- Learn application of regression calibration to correct for FFQ measurement error using repeat short-term reference measurements in a substudy
- With focus on episodically-consumed dietary components, learn application of a new methodology to carry out regression calibration in risk models with energy-adjusted dietary covariates

Outline

- Risk models in nutritional epidemiology
- Dietary measurement error
- Regression calibration
- Modeling episodically-consumed dietary components
  - Two-part model and its extensions
  - Three-part model for episodic component & energy
- Example: NIH-AARP Diet & Health Study
- Simulation study
- Summary & discussion

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

**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)

**Risk models: general description**

- **Notations:**
  - $Y$ - health outcome
  - $T = (T_1, \ldots, T_p)'$ - vector of dietary components
  - $Z = (Z_1, \ldots, Z_q)'$ - vector of adjusting covariates
  - $\eta(T, Z; \alpha)$ - covariate-based predictor ($\alpha$ is a vector of parameters)
  - Risk model: $r(Y | T, Z) = \eta(T, Z; \alpha)$

**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)

**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.,
  
  $$r(Y | T, Z) = E(Y | T, Z)$$

**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.,
  
  $$r(Y | T, Z) = \log\left( \frac{P(Y = 1 | T, Z)}{1 - P(Y = 1 | T, Z)} \right)$$
### Risk models: risk function (3)

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

### Risk models: risk predictor (1)

- Commonly used predictor is a linear function of covariates
  \[
  \eta(T, Z, \alpha) = \alpha_0 + \sum_{k=1}^p \alpha_k T + \sum_{l=1}^q \alpha_l Z
  \]
  - Note: \( \alpha_0 \) is a constant in linear and logistic regressions and \( \alpha_0 = h_0(t) \) (baseline hazard) in Cox regression
  - Convenient but doesn’t always provide a good fit
    - Example: orange vegetables vs. lung cancer in NIH-AARP Diet and Health Study (to be discussed later)

### Risk models: risk predictor (2)

- A more flexible risk model specifies predictor as linear over transformed covariates
  \[
  \eta(T, Z, \alpha) = \alpha_0 + \sum_{k=1}^p \alpha_k T + \sum_{l=1}^q \alpha_l Z
  \]
  where for any variable \( V \), \( V^* = g_V(V) \) denotes its transformed value
  - Popular transformations include power functions \( g_V(V) = V^\gamma \) and logarithm \( g_V(V) = \log V \)

### Risk models: risk predictor (3)

- Risk model:
  \[
  r(Y | T, Z) = \alpha_0 + \sum_{k=1}^p \alpha_k T + \sum_{l=1}^q \alpha_l Z
  \]
  - Slope \( \alpha_k \) 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 \( T_k = T_{k0} \) to \( T_k = T_{k0} + \Delta T_k \) on risk \( r(Y | T, Z) \) is
      \[
      \alpha_k \left[ g_V(T_{k0} + \Delta T_k) - g_V(T_{k0}) \right]
      \]

### Measurement error (1)

- Problem in nutritional epidemiology: true usual intakes are **unknown** and measured with error
- Assessment of diet in nutritional epidemiology is commonly done by food frequency questionnaire (FFQ) querying diet over a specified time period (usually 1 year)
- FFQ is known to contain substantial measurement error, both **random** and **systematic**
Dietary measurement error (2)

- Generally, fitting risk models to error-prone measured dietary exposures $Q$ leads to:
  - Bias (often attenuation) of estimated exposure effect
  - Reduced power to detect exposure effect
  - In theory, invalid significance test for the main exposure (multiple error-prone covariates)
- Most popular method for correcting for dietary measurement error: **regression calibration**

**Regression calibration**

- Main assumption: measurement error is nondifferential with respect to health outcome, i.e., provides no additional information about the outcome 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

**Regression calibration (1)**

In absence of true intakes, each predictor $T_{i,k}^{r}$ is estimated in a substudy (called calibration study) using (often repeat) reference measurements $R_{ij}$

**Regression calibration (2)**

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

\[
T_{i,k}^{r}(X) = E(T_{i,k}^{r} | X), \quad k = 1, \ldots, p.
\]

given vector $X$ that includes all observed error-prone covariates $Q$ and error-free covariates $Z$
- RC leads to (approximately) true regression slopes, i.e., true covariate effects

\[
r(T | X) = \tilde{a}_k + \sum_{k=1}^{p} a_k T_{i,k}^{r}(X) + \sum_{k=1}^{p} \tilde{a}_k Z_{i,k}
\]

**Regression calibration (3)**

Ideal reference measurements of dietary intakes:
- Short-term recovery biomarkers (unfortunately, only few are known: DLW for energy, UN for protein, UK for potassium)
- Reference measures in practice:
  - More detailed short-term self-reports such as multiple-day food records or repeat 24-hour dietary recalls (24HRs)
Implications of possible biases in parameters are estimated in a calibration methodology. This methodology is demonstrated using 24HR (reference instrument in many important dietary cohorts) – Short-term reference period is 1 day

**Working assumption:** 24HR is unbiased in reporting individual's true usual dietary intake – Implications of possible biases in 24HR are discussed at the end

**Methodology below is developed for any correct reference measurement.**

**Regularly-consumed dietary components**

**Ideal world:** the classical measurement error model

\[ R_{ij} = T_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0; \sigma^2) \]

where the regression of \( T_{ij} \) on \( X_i \) is linear, i.e.,

\[ T_{ij} = \beta_0 + \beta_2 X_{ij} + u_{ij} - N(0; \sigma^2) \]

**The measurement error model for reference measurements is thus specified as**

\[ R_{ij} = \beta_0 + \beta_2 X_{ij} + u_{ij} + \epsilon_{ij} \]

**Regression calibration predictor on a transformed scale is given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

For the risk model with predictor on original scale, conditional expectation above exists in closed forms, so that the regression calibration predictor is a linear function of covariates

\[ T_{ij}^* = \beta_0 + \beta_2 X_{ij} \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor on a transformed scale is given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]

**Regression calibration predictor is then given by**

\[ T_{ij}^* = E \left[ \beta_2 \left( \frac{R_{ij}}{T_{ij}} \right) | X_i \right] = E \left[ \beta_2 \left( \frac{\hat{R}_{ij}}{E(T_{ij})} \right) | X_i \right] \]
Linear regression calibration

- Linear regression calibration approximation (LRC): commonly used in nutritional epidemiology
- **Working assumptions:**
  - There is a scale where \( r^* \) is well approximated by the linear mixed effects model and the risk model's predictor, is linear i.e., \( g_T(.) = g_R(.) = g(.) \)
  - On this scale, approximately \( \mathbb{E}(R^*_i | X) \approx T_i \)
- Then \( T_i \approx \mathbb{E}(R^*_i | X) = \beta_0 + \beta_1 X \)
- We will see later that LRC may fail to provide a good approximation for nonlinear models

Episodic dietary components

- Our focus: episodically-consumed dietary components i.e., those that are not consumed daily by nearly everyone (but are eventually consumed in the long run)
- Examples:
  - Many foods (fish, red meat, whole grains, dark green or orange vegetables, etc.)
  - Some nutrients (vitamin A or B12)

Episodic dietary components (1)

Episodic dietary components (2)

- Example: typical short-term report

Episodic dietary components (3)

- Short-term reference measurements for episodically-consumed dietary components
  - Are semicontinuous variables with excess zeros and often skewed to the right positive values
  - Even if otherwise precise, contain substantial within-person measurement error due to short-term variation in intake

Two-part model

Episodic dietary components (4)
For model identifiability, need at least 2 repeat observations on at least a subsample of subjects

Two-part model (1)

- Goal: specifying a measurement error model for semicontinuous reference measurements
- Main idea: modeling a semicontinuous variable as the result of two distinct, although generally correlated processes:
  - One determines whether the variable takes positive or zero value
  - Other specifies its positive value

Two-part model (2)

- Two-part model – first proposed by Cragg (1971) and intensively studied in econometrics and (later) in biostatistics
- Part I – logit/probit regression specifying the probability of positive values
- Part II – linear regression specifying log-transformed positive values

Two-part model (3)

- Extended to longitudinal data by Olsen & Schafer (2001) and Tooze et al. (2002) by introducing mixed effects two-part model with:
  - Fixed effects that are defined by a function of covariates with population-level parameters
  - Random effects that represent part of within-subject mean not explained by covariates; it is subject-specific but randomly varies across subjects
  - Within-subject random errors in positive values representing longitudinal variation

Two-part model (4)

- Longitudinal two-part mixed effects model:
  - Part I – mixed effects logistic regression specifying the probability of positive values
  - Part II – mixed effects linear regression for log-transformed positive values
  - Both parts are linked by allowing correlated person-specific random effects and overlapping covariates
  - For model identifiability, need at least 2 repeat observations on at least a subsample of subjects

Two-part model (5)

- New methodology (NCI method) further extended the longitudinal two-part mixed effects model for short-term reference measurements of episodically-consumed dietary components by:
  - Including Box-Cox family of transformations of positive values (to allow flexibility in addressing skewness)
  - Allowing for within-subject random measurement error

Two-part model (6)

- Two-part NCI model:
  \[
  P(R > 0 | i) = H \left( \beta_1 + \beta_2 X_i + u_{i1} \right), \quad H(v) = \left( 1 + e^{-v} \right)^{-1},
  \]
  \[
  g(R > 0 | \gamma, \beta) = \beta_0 + \beta_2 X_i + u_{i2} + \varepsilon_{i2}, \quad \varepsilon_{i2} \sim N \left( 0, \sigma_{\varepsilon}^2 \right)
  \]
  where:
  \[
  u_i = (u_{i1}, u_{i2}) \sim N \left( \theta, \Sigma \right), \quad \Sigma = \begin{pmatrix}
  \sigma_{u11}^2 & \sigma_{u12}^2 \\
  \sigma_{u12} & \sigma_{u22}^2
  \end{pmatrix}
  \]
  - Part I specifies the probability of consumption & part II specifies the consumption amount; both parts are linked by allowing correlated person-specific random effects and overlapping covariates
### Two-part model (7)

- For a single episodically consumed dietary component, using two repeat 24HRs in US NHANES Survey as the main dietary-assessment instrument, NCI method was applied to:
  - Estimating the distribution of usual intake and its characteristics (Tooze et al, JADA, 2006)
  - Estimating relationships of usual intake with health outcome (Kipnis et al, Biometrics, 2009)

### Two-part model (8)

- Goal: extending NCI methodology for adjusting diet-health relationships for FFQ measurement error when the risk model includes several dietary components
  - In many cases, regression calibration can be applied to error-prone covariates in a risk model one by one
  - But there is a problem with dietary risk models due to energy adjustment

### Bivariate model (1)

- To understand effects of dietary composition, epidemiologists usually consider risk models with energy-adjusted dietary covariates such as:
  - Density, i.e., ratio of usual intake of interest to usual energy intake (focus here)
  - Residual from regressing usual intake of interest on usual energy intake
- Energy-adjusted risk models also include energy
- Since many dietary intakes are correlated with energy intake, estimation of RC predictor requires modeling episodic component and energy together

### Bivariate model (2)

- **Observed data** in calibration sub-study: for person \(i\), time period \(j\)
  - \(R_{ij}\) - short-term reference measurements of episodic dietary component \(F\) and energy \(E\)
  - \(X_i\) - vector of observed covariates, including FFQ-reported intakes \(Q_i\) and error-free covariates \(Z_i\)
  - Indicator variable of reference consumption in period \(j\) for episodic component
    
    \[
    I_{ej} = I(R_{ej} > 0) = \begin{cases} 
    1 & \text{if } R_{ej} > 0 \\
    0 & \text{if } R_{ej} = 0 
    \end{cases}
    \]

### Bivariate model (3)

- **Latent (unobserved) variables**:
  - \(T_{ej}\) - true intakes of interest in period \(j\)
  - \(T_{ei}\) - true usual intake of component of interest
  - \(T_{ei}\) - true usual energy intake
  - \(\frac{T_{ej}}{T_{ei}}\) - true density intake of interest
- Additional latent variables: person-specific random effects and within-person random errors
Note: in part I of the original NCI model the energy intake is naturally specified as part II of the model since energy is always consumed.

During any short-term period, energy intake (continuous variable) should be allowed to be correlated with the indicator of consumption of dietary component of interest (binary variable).

Additional model requirement:
- To satisfy the above requirement, need to modify part I of the two-part NCI model

**Original part I of the NCI model specifies a model for the probability to consume an episodic component but not for the indicator of short-term consumption**

**To satisfy the above requirement, need to modify part I of the two-part NCI model**

**Additional model requirement:**
- Assessing diet health relationships with FFQ: focus on episodically consumed dietary components
- To both parts I and II for episodic component by using the same covariates and allowing:
  - person-specific random effects to be correlated
  - within-person errors $e_{ij}$ to be correlated with $e_{ij}$ and $e_{2ij}$

---

**Bivariate model (4)**

- Energy intake is naturally specified as part II of the NCI model since energy is always consumed
- Allowing correlations between person-specific random effects in energy and episodic component models induces correlation between usual energy and episodic component intakes
- Allowing correlation between within-person errors in energy and part II of episodic component models induces correlation between energy and episodic component during short-term consumption period

**Bivariate model (5)**

- Additional model requirement:
  - During any short-term period, energy intake (continuous variable) should be allowed to be correlated with the indicator of consumption of dietary component of interest (binary variable)
- Original part I of the NCI model specifies a model for the probability to consume an episodic component but not for the indicator of short-term consumption
- To satisfy the above requirement, need to modify part I of the two-part NCI model

**Bivariate model (6)**

- **Modified part I:** to allow $I_{Fij}$ & $R_{ij}$ to be correlated
  - Consider continuous **latent variable** in period $j$
    \[ \hat{R}_{ij} = \beta_{Fij} + \beta_{ij} X_{ij} + u_{ij} + e_{Fij} \]
    where $u_{ij} \sim N(0, \sigma_{u_{ij}}^2), e_{Fij} \sim N(0, 1)$
  - Let $\hat{R}_{ij}$ underlie binary indicator of episodic component’s reference consumption
    \[ I_{Fij} = 1 \Leftrightarrow \hat{R}_{ij} > 0 \]
  - Allow $e_{ij}$, and within-person error in the model for reference energy intake $R_{ij}$ to be correlated

**Bivariate model (7)**

- The use of additional latent variable $\hat{R}_{ij}$ leads to the probability of consumption specified as the mixed effects **probit** model
  \[ P(R_{ij} > 0 | j) = \Phi(\beta_{Fij} + \beta_{ij} X_{ij} + u_{ij}) \]
  where $\Phi$ denotes the distribution function of the standard normal random variable
- Note: in part I of the original NCI model the probability of consumption is specified as the logit model without underlying continuous latent variable in period $j$

**Bivariate model (8)**

- **Part II:** the same as in the original NCI model, i.e., transformed consumption amount during period $j$ is specified as linear mixed effects model:
  \[ g(R_{ij}; \gamma_s) | R_{ij} > 0 = \beta_{2ij} + \beta_{s2} X_{ij} + u_{2ij} + e_{F2ij} \]
  where:
  \[ u_{2ij} \sim N(0, \sigma_{u_{2ij}}^2), e_{F2ij} \sim N(0, \sigma_{e_{F2ij}}^2) \]
- Part I and II are **linked** by using the same covariates and allowing person-specific random effects to be correlated

**Bivariate model (9)**

- **Model for Energy:** transformed energy amount for period $j$ is specified as linear mixed effects model:
  \[ g(R_{ij}; \gamma_s) = \beta_{2ij} + \beta_{s2} X_{ij} + u_{2ij} + e_{F2ij} \]
  where:
  \[ u_{2ij} \sim N(0, \sigma_{u_{2ij}}^2), e_{F2ij} \sim N(0, \sigma_{e_{F2ij}}^2) \]
- Model for energy is **linked** to both parts I and II for episodic component by using the same covariates and allowing:
  - person-specific random effects to be correlated
  - within-person errors $e_{ij}$ to be correlated with $e_{ij}$ and $e_{2ij}$
Bivariate model (10)

- Bivariate model is formally specified as:

\[
\beta_{n\theta} = \int \left[ \beta_{Xn} + \beta_{Xn} \theta + \varepsilon_{n\theta} \right] g^{-1} \left( \beta_{Xn} + \beta_{Xn} \theta + \varepsilon_{n\theta} \right)
\]

where:

\[
u_n = (u_{1n}, u_{2n}, u_{3n}) \sim MVN \left(0; \Sigma_{nu} \right)
\]

\[
e_n = (e_{1n}, e_{2n}, e_{3n}) \sim MVN \left(0; \Sigma_{ne} \right)
\]

Bivariate model (11)

- Denoting by \( \theta_{X} \), model parameters for episodic component and energy, respectively, we have:

\[
R_{ij} = \mathcal{N}_F \left( X_i, u_{ij}, \varepsilon_{ij}; \theta_F \right), \quad R_{ij} = \mathcal{N}_E \left( X_i, u_{ij}, \varepsilon_{ij}; \theta_E \right)
\]

- True usual intakes of episodic component and energy are expectations of those functions, i.e.,

\[
\begin{align*}
T_{ij} &= \mathbb{E} \left[ \mathcal{N}_F \left( X_i, u_{ij}, \varepsilon_{ij}; \theta_F \right) \right] = \mathcal{F}_E \left( X_i, u_{ij}, \theta_F \right), \\
T_{ij} &= \mathbb{E} \left[ \mathcal{N}_E \left( X_i, u_{ij}, \varepsilon_{ij}; \theta_E \right) \right] = \mathcal{E}_E \left( X_i, u_{ij}, \theta_E \right)
\end{align*}
\]

Bivariate model (12)

- True episodic component density is given by:

\[
T_{ij} = \mathcal{F}_E \left( X_i, u_{ij}, \theta_F \right)
\]

- Regression calibration predictor for \( T_{ij} \) is given by:

\[
T_{ij} = \mathbb{E} \left[ g \left( T_{ij}; \theta_F \right) \mid X_i \right] = \mathbb{E} \left[ g \left( \mathcal{F}_E \left( X_i, u_{ij}, \theta_F \right) \right) \mid X_i \right]
\]

Bivariate model (13)

- Model parameters \( \theta_{X} \), \( \theta_{E} \) in the bivariate model are estimated by fitting the model in the calibration sub-study by MLE using NL MIXED SAS procedure

- For any given set of covariate transformations, by using estimated model parameters \( \theta_{X}, \theta_{E} \), regression calibration predictors for transformed density and energy are calculated for each person in the main study

- The final set of covariate transformations is chosen to maximize the overall likelihood when fitting the risk model

NIH-AARP study (1)

- Prospective cohort of 567,169 men & women aged 50-71 in 1995-96 with FFQ administered at baseline

- Calibration substudy of ~1000 men and 1000 women with 2 non-consecutive 24HRs

- Analysis: relationships in men between
  - red meat density & lung cancer
  - orange vegetables density & lung cancer
  - adjusting for age, smoking, and energy intake

- Risk model: Cox regression on original and Box-Cox transformed scales with standard errors estimated by bootstrap to account for estimated RC predictors
NIH-AARP diet and health study (2)

- Due to zeros, for linear regression calibration (LRC), used Box-Cox transformation with a shift parameter:
  \[ g(v; \gamma, \delta) = \begin{cases} 
  \frac{(v + \delta)^{\gamma}}{\gamma} & \text{if } \gamma \neq 0 \\
  \log(v + \delta) & \text{if } \gamma = 0 
\end{cases} \]

- To assess covariate transformations, risk model fit was tested using cumulative martingales technique implemented in SAS (p < 0.05 indicates poor fit)

- Compared FFQ-based analysis (no correction for measurement error) with corrections using linear regression calibration and regression calibration based on the bivariate model

NIH-AARP diet and health study (3)

24HR reported consumption of orange vegetables and red meat

<table>
<thead>
<tr>
<th></th>
<th>Red Meat (g/day)</th>
<th>Orange Vegetables (cups/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean reported intake (s.e.)</td>
<td>62.8 (2.3)</td>
<td>0.14 (0.01)</td>
</tr>
<tr>
<td>Mean amount on consumption days (s.e.)</td>
<td>117.7 (2.6)</td>
<td>0.32 (0.01)</td>
</tr>
<tr>
<td>Mean probability to consume (s.e.)</td>
<td>0.70 (0.01)</td>
<td>0.44 (0.01)</td>
</tr>
<tr>
<td>% of subjects who consumed food:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 out of 2 days</td>
<td>14.7</td>
<td>33.5</td>
</tr>
<tr>
<td>1 out of 2 days</td>
<td>29.9</td>
<td>44.9</td>
</tr>
<tr>
<td>2 out of 2 days</td>
<td>55.4</td>
<td>21.6</td>
</tr>
</tbody>
</table>

NIH-AARP diet and health study (4)

NIH-AARP Diet and Health Study: Red meat intake and lung cancer risk in men; hazard ratios for red meat density between 10 & 60 g/1000 kcal

<table>
<thead>
<tr>
<th>Risk Model</th>
<th>Correction Method</th>
<th>Estimated Log Hazard Ratio (s.e.)</th>
<th>Estimated Hazard Ratio (95% CI)</th>
<th>Risk model fit test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransformed Intake:</td>
<td>No correction for ME</td>
<td>0.225(0.049)</td>
<td>1.255(1.158,1.354)</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>0.409(0.075)</td>
<td>1.502(1.300,1.744)</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC</td>
<td>0.441(0.097)</td>
<td>1.554(1.285,1.880)</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td>Transformed Intake:</td>
<td>No correction for ME</td>
<td>0.248(0.046)</td>
<td>1.281(1.171,1.402)</td>
<td>0.082</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>0.409(0.077)</td>
<td>1.501(1.394,1.751)</td>
<td>0.130</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC (chosen scale)</td>
<td>0.325(0.155)</td>
<td>1.373(1.017,1.886)</td>
<td>0.076</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC (fixed log scale)</td>
<td>0.004</td>
<td>1.004(1.000,1.008)</td>
<td>1.120(0.928,1.354)</td>
<td>0.364</td>
</tr>
</tbody>
</table>

NIH-AARP diet and health study (5)

NIH-AARP Diet and Health Study: Orange vegetable intake and lung cancer risk in men; hazard ratios for orange vegetable density from 0.2 to 0.10 cups/1000 kcal

<table>
<thead>
<tr>
<th>Risk Model</th>
<th>Correction Method</th>
<th>Estimated Log Hazard Ratio (s.e.)</th>
<th>Estimated Hazard Ratio (95% CI)</th>
<th>Risk model fit test</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Untransformed Intake:</td>
<td>No correction for ME</td>
<td>-0.076(0.021)</td>
<td>0.927(0.889,0.966)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>-0.269(0.070)</td>
<td>0.767(0.659,0.894)</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC</td>
<td>-0.233(0.066)</td>
<td>0.806(0.676,0.947)</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>Transformed Intake:</td>
<td>No correction for ME</td>
<td>-0.182(0.030)</td>
<td>0.834(0.786,0.884)</td>
<td>0.256</td>
<td></td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>-0.380(0.089)</td>
<td>0.684(0.574,0.814)</td>
<td>0.060</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC (chosen scale)</td>
<td>-0.593(0.146)</td>
<td>0.553(0.415,0.736)</td>
<td>0.202</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LRC (log scale)</td>
<td>0.005</td>
<td>0.995(0.887,1.118)</td>
<td>0.320</td>
<td></td>
</tr>
</tbody>
</table>

Simulation study (1)

- Main study: for 100 000 subjects generated FFQ and 1000 24HRs with distributions similar to those of orange vegetables and energy in NIH-AARP study
- Calibration substudy: for 1000 subjects used first 2 24HRs as reference measures
- True usual intakes: calculated as averages of 1000 24HRs; density intakes were calculated as ratios of true usual component to usual energy intakes
- Binary outcome: generated using logistic regression with Box-Cox transformed exposure

SIMULATION STUDY
Assessing diet–health relationships with FFQ: focus on episodically consumed dietary components

- **Goal:** estimating log RR for increasing main exposure between 0.02 & 0.10 cups/1000 kcal
- **Risk model:** logistic regression on original and Box-Cox transformed scales with standard errors estimated by bootstrap
- **Compared FFQ-based analysis (no correction for measurement error) with corrections using linear regression calibration and regression calibration based on the bivariate model.**

**Simulation study**

**Simulation study (2)**

- In theory, for risk models on original scale, LRC is approximately consistent, BUT leads to finite sample biases due to unaccounted excess zeros
- For risk models on a transformed scale, LRC may not perform well because
  - Trying to find a scale where both calibration and risk models have linear predictors and reference measurements are unbiased often leads to poor approximations
  - Applying LRC on the original scale by definition leads to a misspecified risk model

**Simulation study: results (1)**

Results of the simulation study: mean, standard deviation and root mean squared error of estimated log odds ratio in logistic regression of disease on orange vegetable intake and energy

<table>
<thead>
<tr>
<th>Sim</th>
<th>Measurement Error Correction Method</th>
<th>Mean log OR (Mean s.e.)</th>
<th>Standard Deviation</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>True parameters</td>
<td>-0.4</td>
<td>0.025</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>No correction for ME</td>
<td>0.42 (-0.242)</td>
<td>0.002</td>
<td>0.180</td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>0.85 (-0.417)</td>
<td>0.005</td>
<td>0.080</td>
</tr>
<tr>
<td></td>
<td>LRC (original scale)</td>
<td>1 (-0.436)</td>
<td>0.008</td>
<td>0.160</td>
</tr>
<tr>
<td></td>
<td>LRC (log scale)</td>
<td>0 (0.08)</td>
<td>-0.557 (0.007)</td>
<td>0.104</td>
</tr>
<tr>
<td></td>
<td>LRC (chosen scale)</td>
<td>-0.61 (0.18)</td>
<td>-0.543 (0.007)</td>
<td>0.191</td>
</tr>
<tr>
<td>2</td>
<td>True parameters</td>
<td>-0.4</td>
<td>0.024</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>No correction for ME</td>
<td>0.52 (-0.207)</td>
<td>0.002</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>RC (Bivariate model)</td>
<td>0.10 (-0.416)</td>
<td>0.007</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>LRC (original scale)</td>
<td>1 (-0.285)</td>
<td>0.006</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>LRC (log scale)</td>
<td>0 (0.006)</td>
<td>-0.342 (0.006)</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>LRC (chosen scale)</td>
<td>-2.18 (0.14)</td>
<td>-0.375 (0.009)</td>
<td>0.121</td>
</tr>
</tbody>
</table>

**Summary & Discussion**

**Summary (1)**

- Developed methodology addresses major challenges for bivariate modeling of short-term reference intakes of an episodic component & energy by allowing during any short-term period:
  - Energy intake to be correlated with the indicator of episodic component consumption
  - Energy intake to be correlated with consumption amount

**Summary (2)**

- Developed methodology allows for rigorous regression calibration to correct for nondifferential covariate measurement error in rather flexible risk models with multiple dietary exposures that:
  - Include energy-adjusted dietary components
  - Include covariates on transformed scales
- Simulations indicate that the developed method performs substantially better than conventional linear regression calibration
Discussion (1)

- Focus here: episodically-consumed dietary components that are eventually consumed in the long run
- What about never consumers?
  - Model could be extended to include never consumers
  - Depending on dietary component and a reference instrument, it may require more than 2 repeat reference measurements (e.g., 4-6 with 24HR-reported fish intake, Kipnis et al, Biometrics 2009)

Discussion (2)

- Developed methodology is based on the important assumption that a short-term reference instrument is unbiased for true usual dietary intake on individual level
- 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 other dietary components

Discussion (3)

- Recent publication (Freedman et al, JNCI 2011) based on OPEN biomarker study suggests that, in spite of biases, using 24HR as a reference to correct for FFQ measurement error on average leads to better results than FFQ-based analysis with no correction for measurement error
- Using more precise short-term reference instruments, such as 24HR, for correcting for FFQ measurement error is a step forward toward better understanding of diet-health outcome relationships

Discussion (4)

- Using reference measurements to calibrate FFQ (approximately) removes bias but does not fully restore the power to detect a relationship, which is lost due to measurement error
- Even bias correction may not be reliable if the attenuated effect is too small (weak signal problem)
- One can do better by using more precise short-term instruments (e.g., web-based ASA24) as the main dietary-assessment method and/or combine different instruments
- The corresponding methodologies will be presented in webinars 10-12

QUESTIONS & ANSWERS

Moderator: Kevin Dodd

Please submit questions using the Chat function

Next Session

Tuesday, November 15, 2011
10:00-11:30 EST

Estimating usual intake distributions for multivariate dietary variables

Raymond J. Carroll
Texas A&M University