Estimating usual intake distributions for dietary components consumed episodically

Janet A. Tooze, PhD, MPH
Wake Forest School of Medicine

Two main areas of interest

- Describing usual intake distributions: mean, percentiles, proportion above or below a threshold
- Estimating diet-health relationships: regression coefficients

Objectives

- There is interest in monitoring a population's usual intake of foods and nutrients
  - Informs research
  - Establishes population norms
  - Guides public policy

Presenters and collaborators

- Sharon Kirkpatrick
  Series Organizer
- Regan Bailey
- Launze Freedman
- Douglas Midthune
- Dennis Buckman
- Patricia Guenther
- Amy Subar
- Raymond Carroll
- Victor Kipnis
- Fran Thompson
- Regan Bailey
- Laurence Freedman
- Douglas Midthune
- Dennis Buckman
- Patricia Guenther
- Amy Subar
- Raymond Carroll
- Victor Kipnis
- Fran Thompson
- Janet Tooze

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.
Objectives

**Daily versus episodic consumption**

- Consumed nearly daily by nearly all persons
  - E.g., vitamin C, total grains, total vegetables, solid fats, added sugars
- Consumed episodically by most persons
  - E.g., vitamin A, whole grains, dark green vegetables, fish

Objectives

**Learning objectives**

- Define key concepts of food consumption related to usual intake estimation
- Identify challenges for estimating usual intake for episodically consumed dietary constituents
- Explain statistical modeling approaches
- Apply NCI macros

Objectives

**Key concepts**

- Consumption patterns vary across dietary constituents
- Usual intake is comprised of probability to consume and consumption-day amount
- Dietary intake data are often skewed
- Current dietary assessment measures are prone to error

Objectives

**Key concepts**

- Consumption patterns vary across dietary constituents
- Usual intake is comprised of probability to consume and consumption-day amount
- Dietary intake data are often skewed
- Current dietary assessment measures are prone to error
What makes up your diet
- Foods
- Food groups
- Components of foods
  - Macronutrients
  - Micronutrients
Many are "episodically consumed"
Most are consumed daily by most persons

Dietary constituents

Daily consumed: Total grains

Source: EATS, men, day 1

Total Grain Intake (ounce equivalents)

Skewed Distribution

Source: EATS, men, day 1
Whole Grain Intake (ounce equivalents)

Episodically consumed: Whole grains

Key concepts

- Consumption patterns vary across dietary constituents
- Usual intake is comprised of probability to consume and consumption-day amount
- Dietary intake data are often skewed
- Current dietary assessment measures are prone to error

Usual intake

- Usual intake = long-term or habitual intake
- Nutrients are stored in the body
- Nutrient and food intake recommendations should be met over time, not necessarily every day

Two parts of usual intake

Usual intake

Probability to consume × Amount consumed on consumption day
Key concepts

Usual intake for a given individual

Person A

Food intake (cups)

0 10
0

Food intake (cups)

0 10

Usual intake

True probability of consumption

Person A

Food intake (cups)

0 10
8/10 = 80%

Food intake (cups)

0 10

Usual intake

True consumption-day amount

Person A

Food intake (cups)

0 10
2/3 cup

Food intake (cups)

0 10

Usual amount

Probability x consumption-day amount

Person A

Food intake (cups)

0 10
Usual intake = 80% x 2/3 = 1/2 cup

Food intake (cups)

0 10

Usual amount

Correlation between probability and amount

Whole grain intake, men

24-hour recall

usual day

Usual intake is comprised of probability to consume and consumption-day amount

Dietary intake data are often skewed

Current dietary assessment measures are prone to error

Source: EATS

Key concepts
Key concepts

Skewness

- Consumption patterns vary across dietary constituents
- Usual intake is comprised of probability to consume and consumption-day amount
- Dietary intake data are often skewed
- Current dietary assessment measures are prone to error

Within-person error: Summary of Webinar 1

- Day-to-day variation
- Random error in reporting
- Additive error
- Intake-related bias
- Person-specific bias

OPEN findings: Structure of measurement error

- 24-hour recall (24HR)
  - Larger within-person random error
  - Smaller systematic error
- Food frequency questionnaire (FFQ)
  - Smaller within-person random error
  - Larger systematic error

Reported intake

- Unbiasedness of 24HR is a working assumption
- Required to proceed with development of methods
- May be more or less justified depending on dietary component of interest
Challenges

Challenges to assessing usual intake

- Account for measurement error
- Account for skewness
- Model probability and amount
- Allow for correlation between probability and amount
- Incorporate covariates

Effect of within-person variation

Methods of correcting for random measurement error

- Estimating usual intake distributions using short-term instruments – some existing methods:
  - Iowa State University Foods (ISUF) Method
  - National Cancer Institute (NCI) Method
  - EFCOVAL Consortium Multiple Source Method (MSM)
  - Statistical Program for Age-adjusted Dietary Assessment (SPADE)

General approach to estimating usual intake

- Separate between-person from within-person variation
  - Assume normality
- Estimate distribution with only between-person variation
Challenges of modeling episodically consumed constituents

- Account for measurement error
- Account for skewness
- Model probability and amount
- Allow for correlation between probability and amount
- Incorporate covariates

Estimating quantiles when transformations are used

Goal is to estimate a quantile of usual intake that corresponds to one in the normal distribution that exhibits only between-person variance.

Backtransformation

- Mean of transformed data ≠ transformation of mean on the original scale
- With nonlinear transformation is used, the estimated quantile is an integral that can be calculated/approximated in several ways
  - Taylor series approximation
  - Numerical integration for known distribution
    - Quadrature formulas, e.g., Gauss-Hermite
    - Monte Carlo integration

Challenges of modeling episodically consumed constituents

- Account for measurement error
- Account for skewness
- Model probability and amount
- Allow for correlation between probability and amount
- Incorporate covariates
For episodically-consumed dietary constituents, we fit two statistical models:
- **Probability**
  - Mixed model logistic regression
- **Amount**
  - Mixed model linear regression

**Two-part model: Person-specific effects**
- Also known as random effects
- Latent
- Constant for an individual
- Captures how an individual's value deviates from the average after adjusting for covariates, if appropriate
- Both the probability and amount models incorporate person-specific effects

**Challenges**
- Estimating usual intake distributions for dietary components consumed episodically

**Modeling correlation**
- Model probability and amount simultaneously
- Correlation between person-specific effects
  - Probability of consumption and consumption day amount
  - Covariates

**Challenges of modeling episodically consumed constituents**
- Account for measurement error
- Account for skewness
- Model probability and amount
- Allow for correlation between probability and amount
- Incorporate covariates

**Estimating distributions**
- Joint Distribution of Probability and Amount
- Account for measurement error
- Account for skewness
- Model probability and amount
- Allow for correlation between probability and amount
- Incorporate covariates
### Types of covariates

- **Individual-specific**
  - Affect true intake on all days
  - e.g., gender/age/race-ethnicity

- **Time-dependent**
  - Affect true intake on specific days
  - e.g., season/weekday

- **Nuisance**
  - Affect reporting error
  - e.g., interview sequence/mode effects

### NCI Method: Overview

**Two-part model:**

**Episodically-consumed constituents**

- **Part 1:** Probability
  - Mixed model logistic regression
  - Can incorporate covariates

- **Part 2:** Amount
  - Mixed model linear regression
  - Transformed scale – accounts for skewness
  - Can incorporate covariates
  - Separates between-/within-person random error

### Definitions

- Let $T_{ij}$ be true intake for a person $i$ on day $j$
  - Let $p_i$ be true probability to consume
  - $p_i = Pr(T_{ij} > 0 | i)$
  - Let $A_i$ be the true average consumption-day amount
  - $A_i = E[T_{ij} | i, T_{ij} > 0]$
  - True usual intake $T_i = E[T_{ij} | i] = p_i A_i$
Let $R_{ij}$ be intake reported on the 24HR for a person $i$ on day $j$.

i. A food is reported on 24HR if and only if consumed
   - Therefore, probability of consumption on recall is the same as the probability of true consumption
   \[ P(R_{ij} > 0|i) = P(T_{ij} > 0|i) = p_i \]

ii. 24HR is unbiased for usual amount consumed on a consumption day
   \[ E[R_{ij} | i; R_{ij} > 0] = A_i \]
   \[ \Rightarrow 24HR \text{ is unbiased for true usual intake} \]
   \[ E[R_{ij} | i] = p_i A_i = T_i \]

iii. On transformed scale the reported amount has additive and independent measurement error

**Part I: Probability to consume**

- Mixed model logistic regression
  \[ Pr(R_{ij} > 0 | X_{ij}, u_i) = h(p_{ij} + \beta_2 X_{ij} + u_i) \]
  - Where $h(\cdot)$ is the logistic function,
  - $X_{ij}$ is a vector of covariates, and
  - $u_i$ is a person-specific random effect
    - Allows a person's value to differ from that defined by covariates
    - $u_i \sim N(0, \sigma^2_{ui})$

**Part II: Amount on consumption days**

- Mixed model linear regression on $g(\cdot)$ Scale
  \[ g(R_{ij} | R_{ij} > 0; X_{2i}, u_{ij}) = \beta_{02} + \beta_2 X_{2i} + u_{ij} + \epsilon_{ij} \]
  - where $g(\cdot)$ is the Box-Cox transformation,
  - $X_{2i}$ is a vector of covariates,
  - $u_{ij} \sim N(0, \sigma^2_{uj})$ is a person-specific random effect,
  - $\epsilon_{ij} \sim N(0, \sigma^2_{\epsilon_{ij}})$ is within-person random error

**Link**

- Person-specific effects have bivariate normal distribution
  \[ (u_{ij}, u_{2ij}) \sim BVN(0, \Sigma) \]

\[ \Sigma = \begin{bmatrix}
\sigma_{u1}^2 & \rho \sigma_{u1} \sigma_{u2} \\
\rho \sigma_{u1} \sigma_{u2} & \sigma_{u2}^2
\end{bmatrix} \]
Estimating the distribution

- Use Monte Carlo approach to generate bivariate distribution of random effects using estimated model parameters
- Approximates integral using a numeric approach
- Combine with empirical distribution of fixed effects
- Backtransform estimate and multiply by estimated probability
- Taylor series
- 9-point quadrature method - recommended
- Estimate percentiles

Implemented in SAS macro DISTRIB
Currently uses Taylor series approximation
9-point approximation to be added

Eating at America’s Table Study (EATS)

- Men and women, 20-70 years
- Nationally representative sampling of 12,615 telephone numbers
  - Approximately 1600 recruited
- Four 24HRs, one in each season
- After one year: FFQ about past year
- 965 respondents completed four 24HRs and FFQ

EATS example: Whole grains

- 36% of men have no consumption on a given day

Correlation between probability and amount is around 0.3

Source: EATS

Conclusions
Example

**MIXTRAN macro: Call**

- %include "C:\NHANES\Macros\mixtran_macro_v1.1.sas";
- %MIXTRAN(data=men, response=r_g_whl_tot, foottype=gwhlm, subject=nid, repeat=notaken, covars_prob, covars_amt, outlib=webinar, modeltype=corr, titles=1, printlevel=2);
- Parameter estimates and predicted values are saved in datasets:
  - outlib_param_modeltype_foodtype_vcontrol
  - webinar_param_modeltype_gwhlm
  - outlib_pred_modeltype_vcontrol
  - webinar_pred_modeltype_gwhlm

Example

**MIXTRAN macro: Output**

- Correlated model with printlevel=2 produces:
  - 3 sets of NLMIXED output
  - Summary of the Un correlated model runs
    - Parameters
    - AIC and -2 log likelihood
  - Summary of the Correlated model runs
    - Parameters
    - AIC and -2 log likelihood with comparison to uncorrelated model

Example

**MIXTRAN macro: Uncorrelated**

Parameter estimates for usual intake:
- Intercept
- lambda

Example

**MIXTRAN macro: Output**

Parameter estimates for usual intake:
- Intercept
- lambda

Example

**DISTRIB macro**

- %include "C:\NHANES\Macros\distrib_macro_v1.1.sas";
- %DISTRIB(seed=0, nsim_mc=100, modeltype=corr, pred=webinar, pred=gwhlm, _param=webinar_param_gwhlm, outlib=webinar, cutpoints=1.25 3.35 5.65 7.75 11.5 2.5 3.25 4.5 5.35 7.5, ncutpoints=13, subject=nid, titles=1, foottype=gwhlm);
- Outputs one SAS data set that contains descriptive statistics for usual intake:
  - outlib.descript_food_freq
  - webinar.descript_gwhlm

Example

**DISTRIB macro: Output**

Selected percentiles and cutpoint probabilities from the distribution
- mean

10/04/2011
**Example**

**DISTRIB macro: descript_gwhlm_dataset**

<table>
<thead>
<tr>
<th>mean_1</th>
<th>tpercentile0</th>
<th>tpercentile10</th>
<th>tpercentile20</th>
<th>tpercentile30</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.55184</td>
<td>0.15964</td>
<td>0.18006</td>
<td>0.21505</td>
<td>0.24106</td>
</tr>
<tr>
<td>0.56062</td>
<td>0.34481</td>
<td>0.39905</td>
<td>0.39197</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tpercentile40</th>
<th>tpercentile50</th>
<th>tpercentile60</th>
<th>tpercentile70</th>
<th>tpercentile80</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15964</td>
<td>0.18006</td>
<td>0.21505</td>
<td>0.24106</td>
<td>0.26621</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>tpercentile90</th>
<th>tpercentile100</th>
<th>cutprob1</th>
<th>cutprob2</th>
<th>cutprob3</th>
<th>cutprob4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15964</td>
<td>0.18006</td>
<td>0.21505</td>
<td>0.24106</td>
<td>0.26621</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cutprob5</th>
<th>cutprob6</th>
<th>cutprob7</th>
<th>cutprob8</th>
<th>cutprob9</th>
<th>cutprob10</th>
<th>cutprob11</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.24432</td>
<td>0.28137</td>
<td>0.37865</td>
<td>0.56452</td>
<td>0.71131</td>
<td>0.81820</td>
<td>0.88930</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>cutprob12</th>
<th>cutprob13</th>
<th>numsubjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.93312</td>
<td>0.96186</td>
<td>446</td>
</tr>
</tbody>
</table>

Estimating usual intake distributions for dietary components consumed episodically

**Example**

**Whole grains (men): Distribution**

Source: EATS

**Whole grains (men): Cumulative distribution**

Source: EATS

**Whole grains: % above cutpoints**

<table>
<thead>
<tr>
<th>Ounce Equivalents</th>
<th>Gender</th>
<th>% Above (4-day mean)</th>
<th>% Above (NCI Method)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/3</td>
<td>Men</td>
<td>78.7%</td>
<td>89.7%</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>74.2%</td>
<td>84.5%</td>
</tr>
<tr>
<td>1</td>
<td>Men</td>
<td>56.7%</td>
<td>62.1%</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>39.5%</td>
<td>42.6%</td>
</tr>
<tr>
<td>3</td>
<td>Men</td>
<td>13.5%</td>
<td>11.1%</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>3.9%</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

Source: EATS; Tooze et al, 2006

**Simulations: Whole grains**

- Data were simulated based on EATS
  - Women
  - Probability and amount are correlated (r=0.23)
- 300 data sets of 2000 individuals
  - Simulate 365 days per person
  - Truth defined as the mean of 365 days
  - Fit model 300 times using only 2 days and take average
- Compare truth to the NCI method and 2-day mean
The NCI Method is less biased than the 2-day mean

For estimating the mean of the distribution both methods do well

In the tails of the distribution

- NCI Method is close to truth
- Simple 2-day mean overestimates the proportion of the population in the tails
The two-part model is appropriate for the estimation of the usual intake for episodically consumed foods.

The NCI Method meets the following challenges:
- Accounts for measurement error
- Accounts for skewness
- Models probability and amount
- Allows for correlation between probability and amount
- Incorporates covariates

**QUESTIONS & ANSWERS**
Moderator: Sharon Kirkpatrick

Please submit questions using the *Chat* function.