National Cancer Institute

measurementERRORwebinar series

Estimating usual intake distributions for dietary components consumed episodically

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

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

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

Wake Forest School of Medicine · The Gertner Institute · STATISTICS Texas A&M University · National Cancer Institute · Office of Dietary Supplements · USDA

---

measurementERRORwebinar series

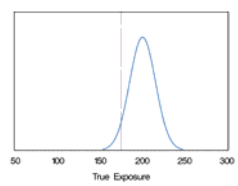
This series is dedicated to the memory of
Dr. Arthur Schatzkin

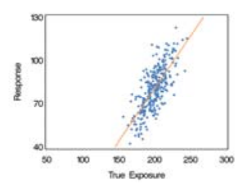
In recognition of his internationally renowned contributions to the field of nutrition epidemiology and his commitment to understanding measurement error associated with dietary assessment.

Objectives

Two main areas of interest

- Describing usual intake distributions: mean, percentiles, proportion above or below a threshold





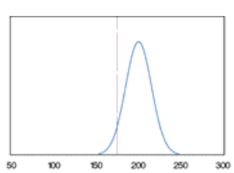
- Estimating diet-health relationships: regression coefficients

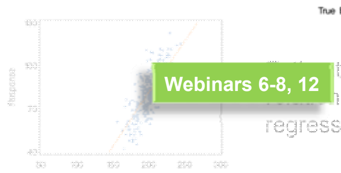
---

Objectives

Two main areas of interest

- Describing usual intake distributions: mean, percentiles, proportion above or below a threshold





...ting diet-health ...nships: regression coefficients

Objectives

Estimating distributions of usual intake

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

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

Daily versus episodic consumption

- **Consumed nearly daily by nearly all persons**
  - E.g., vitamin [Webinar 2] total vegetables, solid fats, added sugars

Webinar 2

- **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

Key concepts → Challenges → Statistical modeling → Example → Simulations → Conclusions

KEY CONCEPTS

Key concepts

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

Key concepts

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

Key concepts

Dietary constituents

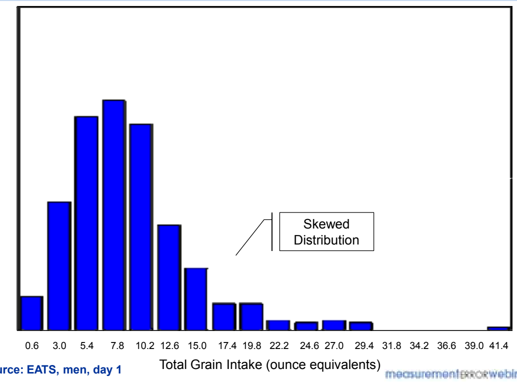
- What makes up your diet
  - Foods
  - Food groups
  - Components of foods
    - Macronutrients
    - Micronutrients

Many are "episodically consumed" (Foods, Food groups)

Most are consumed daily by most persons (Components of foods, Macronutrients, Micronutrients)

Key concepts

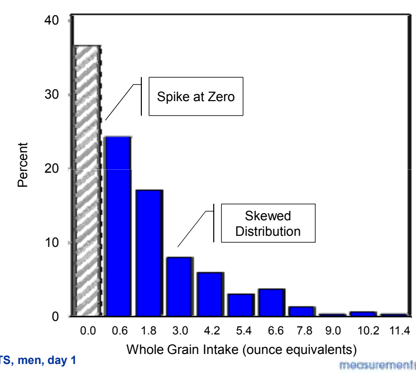
Daily consumed: Total grains



Source: EATS, men, day 1

Key concepts

Episodically consumed: Whole grains



Source: EATS, men, day 1

Key concepts

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

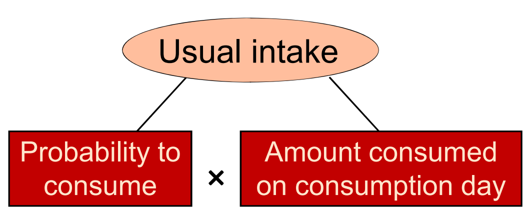
Key concepts

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

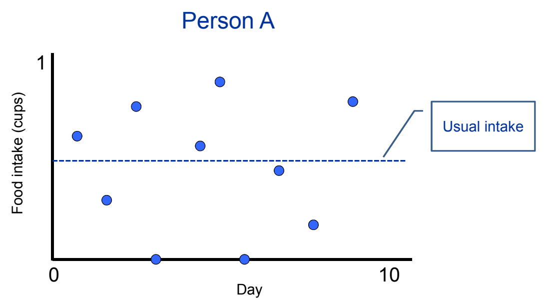
Key concepts

Two parts of usual intake



Key concepts

Usual intake for a given individual

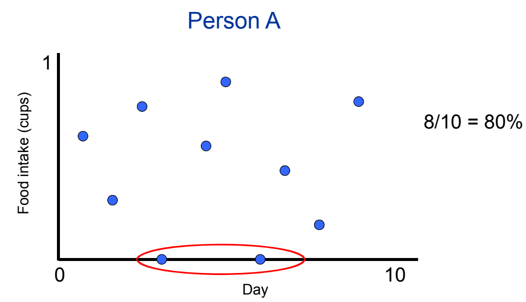


measurementERRORwebinar series

19 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

True probability of consumption

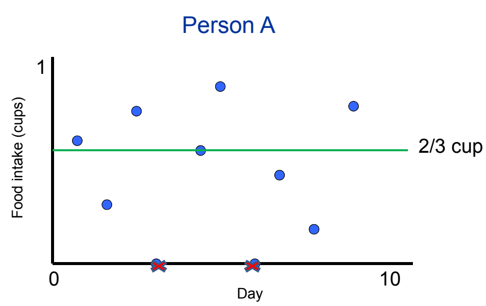


measurementERRORwebinar series

20 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

True consumption-day amount

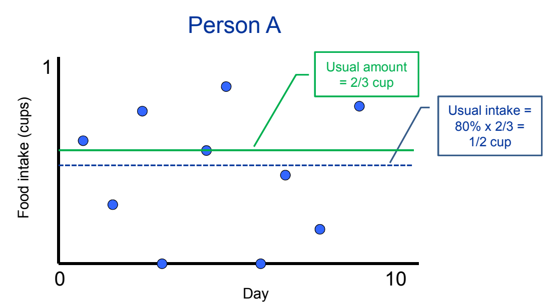


measurementERRORwebinar series

21 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Probability x consumption-day amount

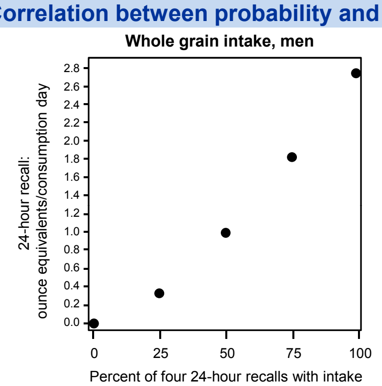


measurementERRORwebinar series

22 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Correlation between probability and amount



Source: EATS

measurementERRORwebinar series

23 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Key concepts

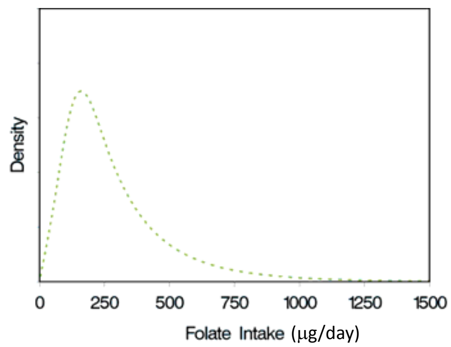
- Consumption patterns vary across dietary constituents
- Usual intake is comprised of probability to consume and consumption-day amount
- **Dietary intake data are often skewed**
- Current dietary assessment measures are prone to error

measurementERRORwebinar series

24 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Skewness



measurementERRORwebinar series

25 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

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**

measurementERRORwebinar series

26 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Within-person error: Summary of Webinar 1

- Day-to-day variation
- Random error in reporting

} Random

- Additive error
- Intake-related bias
- Person-specific bias

} Systematic

measurementERRORwebinar series

27 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

OPEN findings: Structure of measurement error

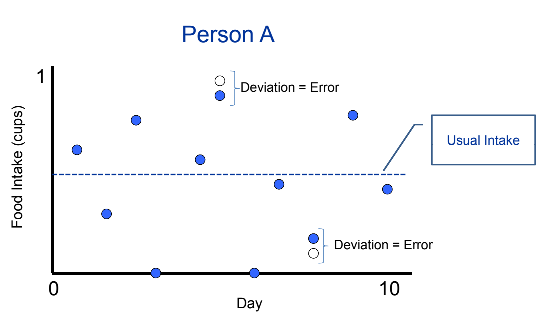
24-hour recall (24HR)	Food frequency questionnaire (FFQ)
• **Larger within-person random error** • Smaller systematic error	• Smaller within-person random error • **Larger systematic error**

measurementERRORwebinar series

28 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Reported intake



measurementERRORwebinar series

29 Estimating usual intake distributions for dietary components consumed episodically

Key concepts

Unbiasedness: Working assumption

- 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

measurementERRORwebinar series

30 Estimating usual intake distributions for dietary components consumed episodically

Key concepts → **Challenges** → Statistical modeling → Example → Simulations → Conclusions

CHALLENGES TO ASSESSING USUAL INTAKE

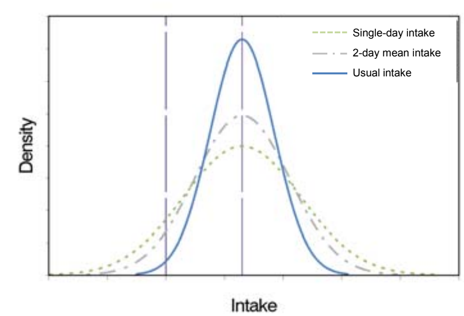
Challenges of modeling episodically consumed constituents

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

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

Effect of within-person variation

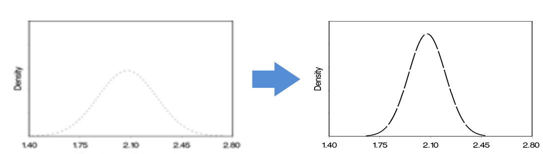


Methods of correcting for random measurement error

- 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



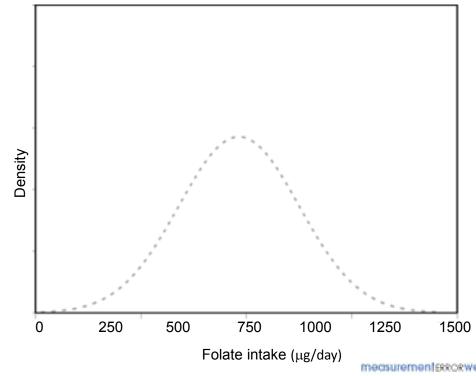
Challenges

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

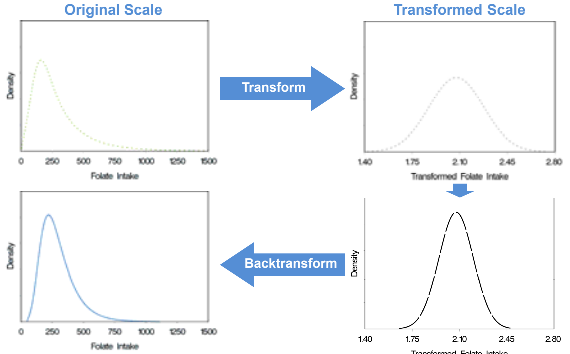
Challenges

Skewness



Challenges

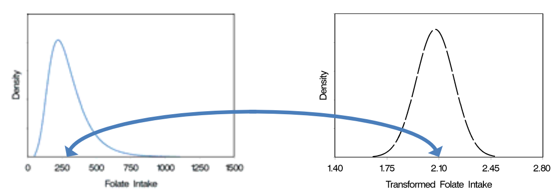
Accounting for nonlinear transformations



Challenges

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



Challenges

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

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

Challenges

Model probability and amount: Two-part model

- For episodically-consumed dietary constituents, we fit two statistical models:
  - Probability
    - Mixed model logistic regression
  - Amount
    - Mixed model linear regression

Challenges

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

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

Challenges

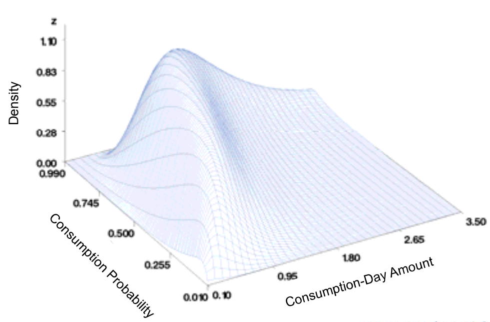
Modeling correlation

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

Challenges

Estimating distributions

Joint Distribution of Probability and Amount



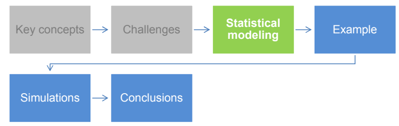
Challenges

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**

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



STATISTICAL MODELING: NCI METHOD

NCI Method: Overview

Two-part model:
Episodically-consumed constituents

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

NCI Method: Overview

Two-part model:
Episodically-consumed constituents

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

NCI Method: Overview

Two-part model:
Episodically-consumed constituents

- Link
  - Person-specific effects are correlated
  - May share covariates

Definitions

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

Statistical modeling

Assumptions

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

$$\Pr(R_{ij} > 0|i) = \Pr(T_{ij} > 0|i) = p_i$$

ii. 24HR is unbiased for usual amount consumed on a consumption day

$$E[R_{ij} \mid i; R_{ij} > 0] = A_i$$

$\Rightarrow$ 24HR is unbiased for true usual intake

$$E[R_{ij} \mid i] = p_i A_i = T_i$$

measurementERRORwebinar series

55 Estimating usual intake distributions for dietary components consumed episodically

Statistical modeling

Part I: Probability to consume

- Mixed model logistic regression

$$\Pr(R_{ij} > 0 \mid \mathbf{X}_{1i}, u_{1i}) = h(\beta_{10} + \boldsymbol{\beta}'_{X1}\mathbf{X}_{1i} + u_{1i})$$

- Where *h*( ) is the logistic function,
- $\mathbf{X}_{1i}$ is a vector of covariates, and
- u_{1i} is a person-specific random effect
  - Allows a person's value to differ from that defined by covariates
  - $u_{1i} \sim N(0, \sigma^2_{u1})$

measurementERRORwebinar series

56 Estimating usual intake distributions for dietary components consumed episodically

Statistical modeling

Assumptions revisited

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

$$P(R_{ij} > 0|i) = P(T_{ij} > 0|i)$$

ii. 24HR is unbiased for usual amount consumed on a consumption day

$$E[R_{ij} \mid i; R_{ij} > 0] = A_i$$

$\Rightarrow$ 24HR is unbiased for true usual intake

$$E[R_{ij} \mid i] = p_i A_i = T_i$$

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

measurementERRORwebinar series

57 Estimating usual intake distributions for dietary components consumed episodically

Statistical modeling

Part II: Amount on consumption days

- Mixed model linear regression on *g(·)* Scale

$$g(R_{ij}, \gamma \mid R_{ij} > 0; \mathbf{X}_{2i}, u_{2i}) = \beta_{20} + \boldsymbol{\beta}'_{X2}\mathbf{X}_{2i} + u_{2i} + e_{ij}$$

- where *g( )* is the Box-Cox transformation,
- $\mathbf{X}_{2i}$ is a vector of covariates,
- $u_{2i} \sim N(0, \sigma^2_{u2})$ is a person-specific random effect,
- $e_{ij} \sim N(0, \sigma^2_e)$ is within-person random error

measurementERRORwebinar series

58 Estimating usual intake distributions for dietary components consumed episodically

Statistical modeling

Link

- Person-specific effects have bivariate normal distribution

$$(u_{1i}, u_{2i})' \sim BVN(\mathbf{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2_{u1} & \rho\sigma_{u1}\sigma_{u2} \\ \rho\sigma_{u1}\sigma_{u2} & \sigma^2_{u2} \end{bmatrix}$$

measurementERRORwebinar series

59 Estimating usual intake distributions for dietary components consumed episodically

Statistical modeling

Fitting the model

- Implemented in SAS macro MIXTRAN that calls PROC NLMIXED
- Download at http://riskfactor.cancer.gov/diet/usualintakes/macros.html

measurementERRORwebinar series

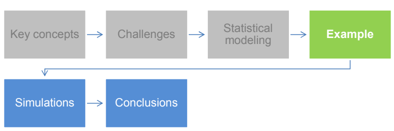
60 Estimating usual intake distributions for dietary components consumed episodically

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

Estimating the distribution

- Implemented in SAS macro DISTRIB
- Currently uses Taylor series approximation
- 9-point approximation to be added
- Download at http://riskfactor.cancer.gov/diet/usualintakes/macros.html



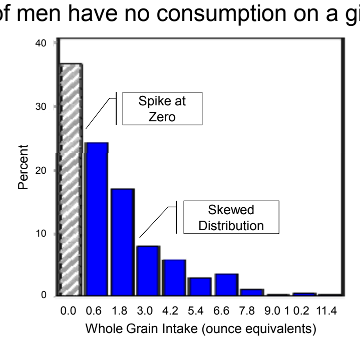
EXAMPLE

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

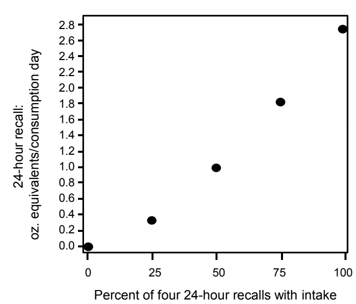
EATS example: Whole grains

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



EATS example: Whole grains

- Correlation between probability and amount is around 0.3



Source: EATS

Example

MIXTRAN macro: Call

- %include "C:\NHANES\Macros\mixtran_macro_v1.1.sas";
- %MIXTRAN(data=men, response=r_g_whl_tot, foodtype=gwhlm, subject=nid, repeat=intaken, 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_corr_gwhlm_
  - outlib._pred_modeltype_foodtype_vcontrol
    - webinar._pred_corr_gwhlm_

Example

MIXTRAN macro: Output

- Correlated model with printlevel=2 produces:
  - 3 sets of NLMIXED output
  - Summary of the Uncorrelated 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

Men
Results from Fitting Uncorrelated Model
Response Variable: r_g_whl_tot

Convergence Status:
Probability Model -- NOTE: GCONV convergence criterion satisfied.
Amount Model -- NOTE: GCONV convergence criterion satisfied.

Parameter	Name	Estimate	Std Err	Prob>|t|
P01_INTERCEPT	Intercept--bi	0.7311	0.0871	0.0000
P_LOGSDU1	Reparam Var(u1)--bi	0.2571	0.0857	0.0029
A01_INTERCEPT	Intercept--in	0.5538	0.0504	0.0000
A_LAMBDA	lambda--in	0.3134	0.0195	0.0000
A_LOGSDE	Resid, Reparam--in	0.1408	0.0269	0.0000
A_LOGSDU2	Reparam Var(u2)--in	-0.4863	0.0893	0.0000

Name	Value	Sum
AIC--bi	2228.41	.
AIC--amount	4252.40	6480.81
-2 Log Likelihood--bi	2224.41	.
-2 Log Likelihood--amount	4244.40	6468.81

Example

MIXTRAN macro: Correlated

Men
Results from Fitting Correlated Model
Response Variable: r_g_whl_tot

Convergence Status:
NOTE: GCONV convergence criterion satisfied.

Parameter	Name	Estimate	Std Err	Prob>|t|
P01_INTERCEPT	Intercept--bi	0.7249	0.0862	<.0001
P_LOGSDU1	Reparam Var(u1)--bi	0.2436	0.0853	0.0045
A01_INTERCEPT	Intercept--in	0.4168	0.0558	<.0001
A_LAMBDA	lambda--in	0.3108	0.0194	<.0001
A_LOGSDE	Resid, Reparam--in	0.1349	0.0267	<.0001
A_LOGSDU2	Reparam Var(u2)--in	-0.4058	0.0847	<.0001
Z_U	Z-trans of Correlation	0.9356	0.2187	<.0001

Name	Value	Diff in -2ll	p-value
AIC	6446.72	.	.
-2 Log Likelihood	6432.72	36.09	0.0000

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 .33 .5 .66 .75 1 1.5 2 2.5 3 3.5 4, ncutpnt=13, subject=nid, titles=1, food=gwhlm);
- Outputs one SAS data set that contains descriptive statistics for usual intake:
  - outlib.descript_food_freq_var
    - webinar.descript_gwhlm_

Example

DISTRIB macro: Output

Men
Selected percentiles and cutpoint probabilities from the distribution

numsubjects	mean_mc_t	tpercentile1	tpercentile5	tpercentile10
446	1.55184	0.062229	0.18606	0.31915

tpercentile15	tpercentile25	tpercentile40	tpercentile50	tpercentile75
0.43977	0.67415	1.04480	1.30851	2.16360

tpercentile85	tpercentile90	tpercentile95	tpercentile99	cutprob1
2.70120	3.09892	3.74623	5.15453	0.020169

cutprob2	cutprob3	cutprob4	cutprob5	cutprob6	cutprob7	cutprob8
0.073212	0.10439	0.17508	0.24432	0.28137	0.38185	0.56452

cutprob9	cutprob10	cutprob11	cutprob12	cutprob13
0.71131	0.81820	0.88930	0.93312	0.96186

Example

DISTRIB macro: descript_gwhlm_ dataset

mean_mc_t	tpercentile0	tpercentile1	tpercentile2	tpercentile3
1.55184	.002636961	0.062229	0.099405	0.13062

tpercentile4	tpercentile5	tpercentile6	tpercentile7	tpercentile8
0.15964	0.18606	0.21505	0.24106	0.26621

tpercentile9	tpercentile10	tpercentile11	tpercentile12	tpercentile13
0.29212	0.31915	0.34491	0.36905	0.39197

. . .

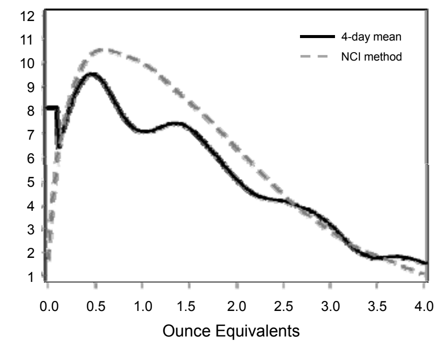
tpercentile99	tpercentile100	cutprob1	cutprob2	cutprob3	cutprob4
5.15453	9.68182	0.020169	0.073212	0.10439	0.17508

cutprob5	cutprob6	cutprob7	cutprob8	cutprob9	cutprob10	cutprob11
0.24432	0.28137	0.37865	0.56452	0.71131	0.81820	0.88930

cutprob12	cutprob13	numsubjects
0.93312	0.96186	446

Example

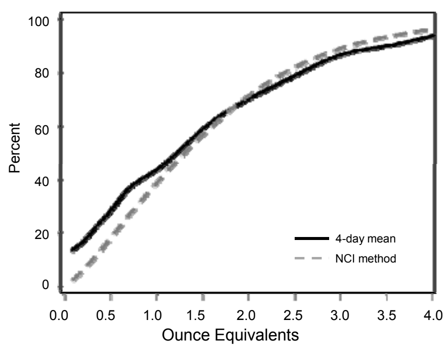
Whole grains (men): Distribution



Source: EATS

Example

Whole grains (men): Cumulative distribution



Source: EATS

Example

Whole grains: % above cutpoints

Ounce Equivalents	Gender	% Above (4-day mean)	% Above (NCI Method)
1/3	Men	78.7%	89.7%
	Women	74.2%	84.5%
1	Men	56.7%	62.1%
	Women	39.5%	42.6%
3	Men	13.5%	11.1%
	Women	3.9%	1.1%

Source: EATS; Tooze et al, 2006

SIMULATIONS

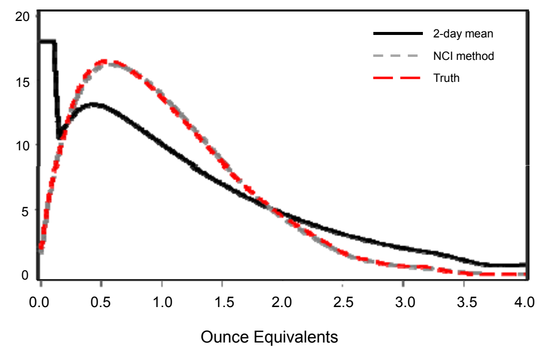
Simulations

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

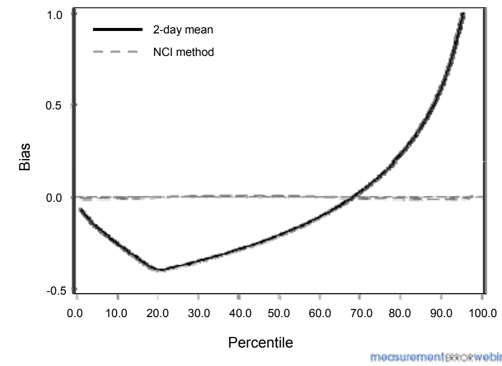
Simulations

Simulations: Distribution



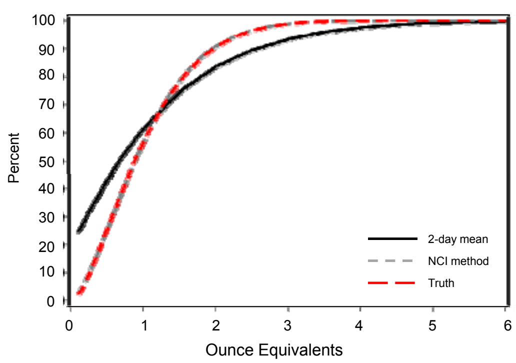
Simulations

Simulations: Bias



Simulations

Simulations: Cumulative distribution



Simulations

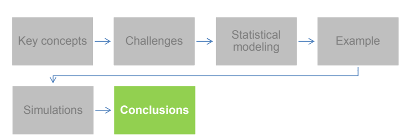
Simulations: Percent consuming below cutoffs

Ounce Equivalents	% Above (Truth)	% Above (2-day mean)	% Above (NCI Method)
1/3	86.1	65.9	87.3
1/2	75.8	57.9	76.2
1	44.5	38.4	43.3
2	9.2	16.1	8.3
3	1.2	6.1	1.1
4	0.0	2.7	0.1

Simulations

Summary of simulation studies

- 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



CONCLUSIONS

Conclusions

- 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

National Cancer Institute

measurementERRORwebinar series

Next Session

Tuesday, October 11, 2011
10:00-11:30 EDT

Estimating usual dietary intake distributions:

Accounting for complex survey design in modeling usual intake

Kevin W. Dodd, PhD
National Cancer Institute

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