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

Victor Kipnis, PhD
National Cancer Institute, USA

Slide 1

Hello and welcome to the eighth session in the Measurement Error Webinar Series. I'm Kevin Dodd, a statistician at the U.S. National Cancer Institute, and I'll be moderating today's webinar, in which we'll continue with our focus on examining diet and health relationships.

A few notes before we get started with today's presentation: The webinar is being recorded so that we can make it available on our Web site. All phone lines have been muted and will remain that way throughout the webinar. There will be a question and answer session following the presentation; please use the Chat feature to submit a question.

A reminder: You can find the slides for today's presentation on the Web site that has been set up for series participants. The URL is available in the Notes box at the top left of the screen. Other resources available include the glossary of key terms and notation, and the recordings of the preceding webinars.

Now I'd like to introduce Dr. Victor Kipnis, our presenter for today. Victor is a mathematical statistician in the Biometry Research Group, Division of Cancer Prevention, at the National Cancer Institute of the United States. Victor's research focus is on the design and analysis of nutritional studies, including the structure of dietary measurement error, its effects on study results, and methods of adjusting for it in nutritional epidemiology and surveillance. Dr. Kipnis played a leading role in the design and analysis of the Observing Protein and Energy Nutrition, or OPEN, biomarker study carried out at NCI in 2001-2002 and is a lead statistician on the NIH-AARP Diet and Health cohort study. In today's session, Victor will discuss methods of assessing diet and health relationships using a food frequency questionnaire as the main dietary instrument, with a focus on episodically consumed dietary components. Victor.

measurement ERROR webinar 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.

Slide 2

This lecture, as all webinars in the series, is dedicated to the memory of Arthur Schatzkin—a friend, a colleague who throughout his career had been very much interested in understanding measurement error and its role in nutritional epidemiology.

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

Slide 3

And this is a list of all the people involved in this project. As you may see they are from different institutions and even from different countries.

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

Learning objectives for today's talk: I will review statistical models for evaluating diet-health relationships in nutritional epidemiology. We'll learn applications of the regression calibration method to correct for measurement error when one uses short-term reference measurements in a substudy and the food frequency questionnaire, or FFQ, as the main instrument in the study. And my focus today will be on episodically consumed dietary components. And this will necessitate application of a new methodology to carry out regression calibration, and so we'll review this as well.

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

Slide 5

This is my outline. I would like to say from the beginning that this talk is going to be at a little bit more elevated level than the previous talks and I will review a little bit more flexible risk models in nutritional epidemiology that will involve some transformations of exposures and covariates. And because I will be focusing on episodically consumed dietary components, regression calibration needs to be explained in a slightly more general way, which will also involve some transformations. Some of them will be nonlinear and the models are going to be nonlinear. And, inevitably, I will have to introduce some formulas, but I will try my best, I promise, to explain what is behind those formulas in plain English.

I will show you some examples from the NIH-AARP Diet & Health Study and some simulation results, and we'll end up with a summary and discussion.
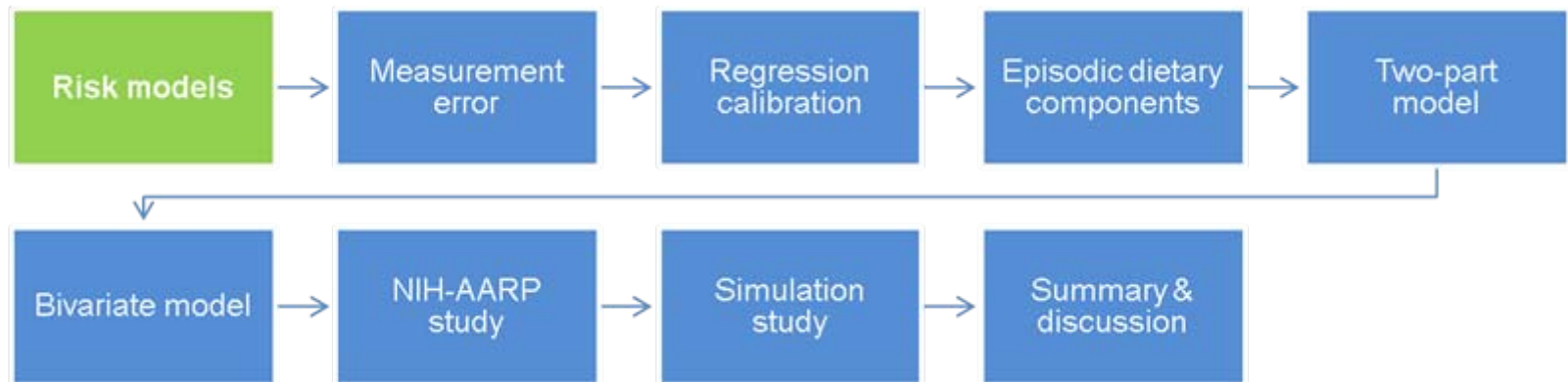
RISK MODELS IN NUTRITIONAL EPIDEMIOLOGY

Slide 6

And I will take you through this step by step, and I will start with the risk models.

Types of epidemiologic studies

- Animal experiments

- Ecological studies

- Cross-sectional studies

- Case-control studies

- **Cohort studies** (main focus here)

- Randomized prevention trials

There are many types of epidemiologic studies. I will concentrate today on cohort studies, and you will see in a moment why.

Risk models: exposure

- We consider studies that relate:

Dietary Exposure (adjusted for covariates) → Health Outcome
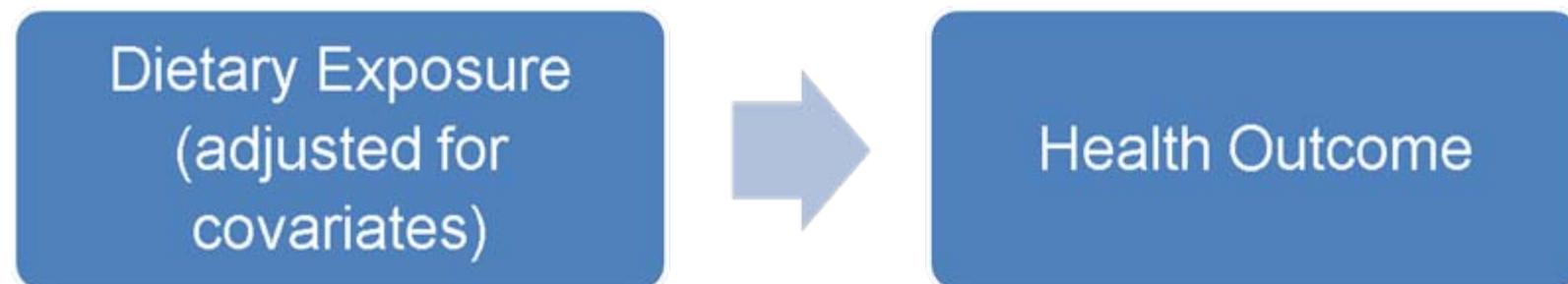
- Dietary exposure thought to be most relevant is usual (long-term average) daily dietary intake

- Health outcome examples: continuous (e.g., blood pressure), binary (event, no event), time to event (survival analysis)

Slide 8

What we would like to do is to relate dietary exposure to a health outcome, and this relationship, of course, will be adjusted for some confounders or covariates. The dietary exposure thought to be the most relevant is usual, which is long-term average daily dietary intake.

As to the health outcomes, there are several examples. They could be continuous; for example, blood pressure or cholesterol. They could be binary (event, no event), say, cancer, no cancer. Or they could be time to event in survival analysis.

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

    - $r(Y \mid \mathbf{T}, \mathbf{Z})$ - outcome risk function

    - $\eta(\mathbf{T}, \mathbf{Z}; \alpha)$ - covariate-based predictor (α is a vector of parameters)

- Risk model: $r(Y \mid \mathbf{T}, \mathbf{Z}) = \eta(\mathbf{T}, \mathbf{Z}; \alpha)$

Slide 9

In my notations, the health outcome will be denoted by Y. T is a vector of dietary components. Z is a vector of adjusting covariates or confounders. And by r I will consider the outcome risk function—I will explain what it means in a moment—and by η, a covariate-based predictor, because remember, what we want to do is to relate the risk function for the outcome to covariate-based predictors. So this model in its more general form is displayed here.

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)

Slide 10

What are the common risk models? Well, I'm sure you all are familiar with those. They could include the linear regression for a continuous outcome. You could do the logistic regression for a binary outcome or Cox regression in survival analysis.

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.,

$$r(Y \mid \mathbf{T}, \mathbf{Z}) = \mathrm{E}\left(Y \mid \mathbf{T}, \mathbf{Z}\right)$$

Slide 11

So in the linear regression, the outcome is a continuous variable; as I said, it could be blood pressure, cholesterol level, etc., etc. And the risk function is the conditional expectation or conditional mean given the covariates.

Let me explain what that means. Basically, for any given set of covariates, we have a bunch of possible outcomes—actually, the whole distribution. And what we will try to do is to consider as the risk of the health outcome some important characteristics of this distribution. And in this case, it's its conditional mean.

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.,

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

Slide 12

In the case of the logistic regression, outcome is a binary variable, usually denoted by 1 in case of event or 0 if there is no event. And the risk function is the logit of the probability of event or, in other words, is the log odds of event, which is the probability of having an event over the probability of not having an event, and then you take the log of this.

Risk models: risk function (3)

- ## **Cox regression**

  - Outcome: $Y = t$ (time to event)

  - Risk function: log of the hazard function $h(t \mid \mathbf{T}, \mathbf{Z})$ conditional on covariates, i.e.,

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

In the case of the Cox regression, the outcome is actually a time, time to event, and the risk function becomes the log of the hazard function, denoted here by *h*, which is again conditional on the covariates.

Risk models: risk predictor (1)

- Commonly used predictor is a linear function of covariates

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

  – Note: α_0 is a constant in linear and logistic regressions and $\alpha_0 = h_0\left(t\right)$ (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)

Slide 14

What we would like to do is to use a predictor of the covariates to relate the covariates to the risk function. And the commonly used predictor is a linear function of covariates given by this formula. And why a linear function? It's the simplest function. It can be well studied and so is the most convenient one.

I would like to emphasize that in this linear function, the intercept has a slightly different meaning in linear and logistic regressions versus the Cox regression. In the linear and logistic regressions, it's just a constant. In the Cox regression, it's a baseline hazard.

So, as I said, it's the most convenient risk predictor, but sometimes this most convenient predictor doesn't fit well. And we'll see an example of this when we study the relationship between orange vegetables versus lung cancer in the NIH-AARP Diet & Health Study.

Risk models: risk predictor (2)

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

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

where for any variable V, $V^* = g_v(V)$ denotes its transformed value

- Popular transformations include power functions $g_v(V) = V^{\gamma}$ and logarithm $g_v(V) = \log V$

There are many ways to consider more complicated risk predictors, and one which is still relatively simple but rather flexible is to transform covariates in such a way that over the transformed variables or on transformed scales, the predictor is still linear. The transformation is done using some function, g, and I will denote the transformed variables by putting this asterisk as a superscript. Popular transformations include the power functions such as square root, for example, or logarithms.

Risk models: risk predictor (3)

- Risk model: $r(Y \mid \mathbf{T}, \mathbf{Z}) = \alpha_o + \sum_{k=1}^{p} \alpha_{T_k} T_k^* + \sum_{l=1}^{q} \alpha_{Z_l} Z_l^*$

- Slope α_{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 $r(Y \mid \mathbf{T}, \mathbf{Z})$ is

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

And here is my risk model with a predictor linear over, generally speaking, transformed covariates. And α, or the slope, for exposure T_k represents the effect of this exposure on the outcome. Now, one has to be careful, because with the transformations the interpretation of the meaning of this effect should be considered with some care. If we use the linear predictor on the original scale, then the effect would depend only on change in the exposure. And so for any given change in the exposure, the product of the slope times the change will be your effect.

With the transformation, it's not only the change, but where you start. So the effect depends on the initial value as well. But it could be done for the initial value on the original scale and it's introduced here by this formula. So you take the initial value; you take the initial value plus the change, so this is the final value. You transform both to the scale where you fitted your risk model, and then you multiply it by the slope, and this is the effect.
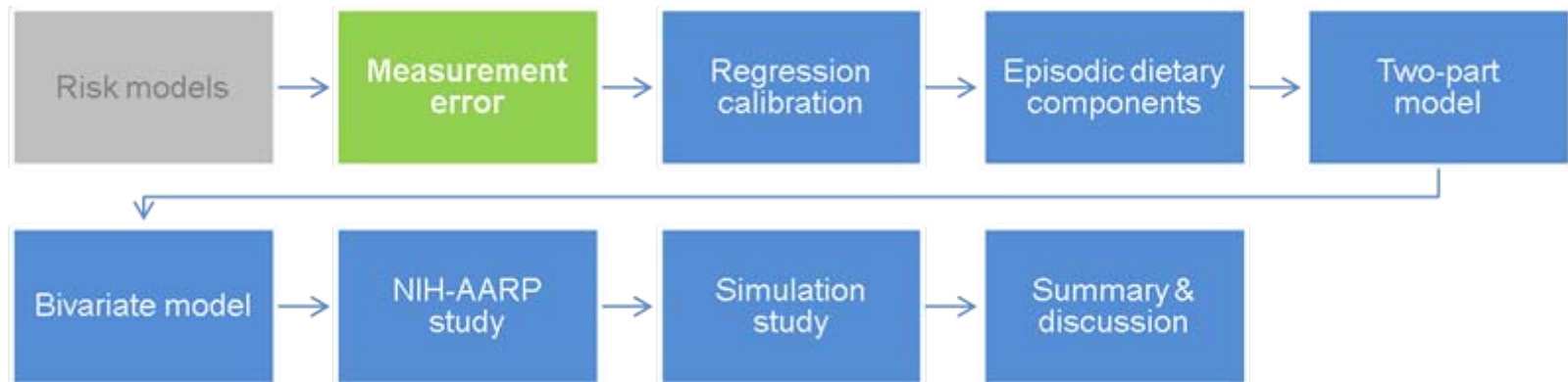
MEASUREMENT ERROR

Slide 17

All right, let's go to the second block, which is measurement error.

Dietary 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**

Slide 18

This is a big problem in nutritional epidemiology because the true usual intakes are unknown and could not be measured precisely in free-living populations. And all measured intake is measured with error.

In nutritional epidemiology, especially in large cohort studies, assessment of diet is commonly done by a food frequency questionnaire, or FFQ, which queries diet over a specified time period. It's usually one year. FFQs are 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**

Slide 19

And, generally, if we fit the risk model to error-prone dietary exposures, which I denote here by Q, it leads to three unpleasant things: first, bias (often attenuation) of estimated exposure effect—attenuation means that the relative risk is biased towards 1; reduced power to detect exposure effect, which means that to maintain the power, one has to increase substantially the sample size of the study; and, in theory, if the risk model contains multiple error-prone covariates, the significance test for the main exposure may become invalid.

The most popular method for correcting for dietary measurement error is regression calibration.
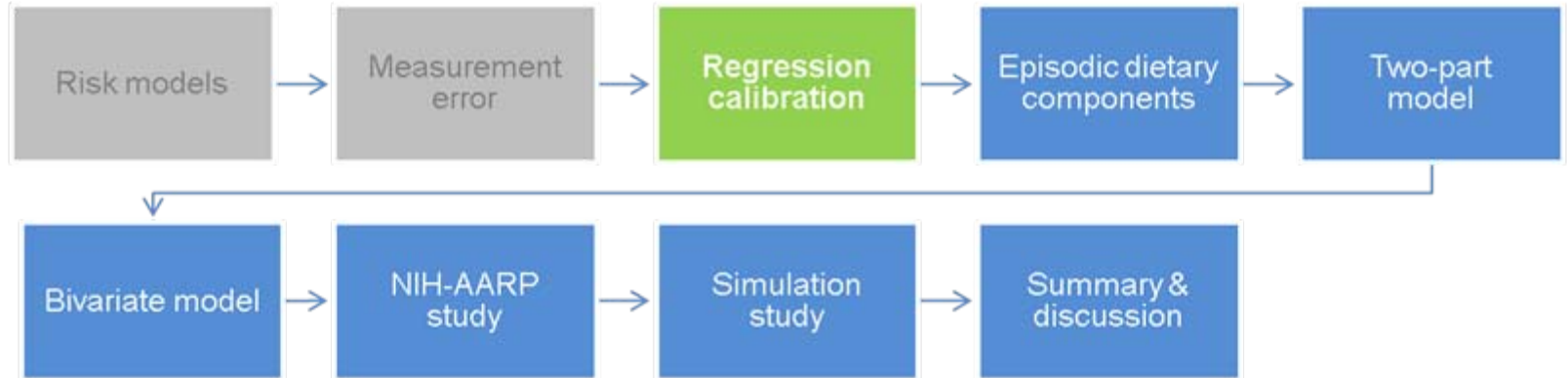
REGRESSION CALIBRATION

Slide 20

You've heard about regression calibration before. I will just review it and introduce it in a slightly more general way. And we will need this later on.

Regression calibration (1)

- 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

Slide 21

But before I do so, the regression calibration could be applied under the main assumption that the measurement error is nondifferential with respect to the health outcome. And what it means is that the measured exposure provides no additional information about the outcome beyond the information that is contained in true diet.

This assumption may be justified in cohort studies where diet is assessed at the baseline, before outcome is known, but not necessarily in case-control studies. And the reason for this is that in case-control studies, cases and controls could recall their past diet differently, which will lead to differential measurement error.

Regression calibration (2)

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

$$T_k^{*P}(\mathbf{X}) = \mathrm{E}(T_k^* \mid \mathbf{X}), \ k = 1, ..., p,$$

  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

$$r(Y \mid \mathbf{X}) = \tilde{\alpha}_0 + \sum_{k=1}^{p} \alpha_{T_k} T_k^{*P}(\mathbf{X}) + \sum_{l=1}^{q} \alpha_{Z_l} Z_l^*$$

Slide 22

So, assuming that measurement error is nondifferential, what is the regression calibration? What the regression calibration does is it replaces each mismeasured covariate, the true value of which is unknown, with its best predictor given everything that was observed, which in this case includes error-prone measurements Q and error-free covariates, which I denote by Z.

What it means, again, is that for each given set of components of vector X, there are a bunch of true usual intakes, here on the transformed scale, that represents the whole distribution. And we would like to take one characteristic of this distribution—in this case, it's conditional mean—and if we do so, this would be the best, in the mean squared error sense, predictor of the unknown true intake.

I would like to emphasize that when we calculate this conditional mean, it should be done on the transformed scale. The result of taking it on the original scale and then transforming might be very different, depending on the transformation. So the order here is very, very important.

And if we do so, we will retain the true regression slopes exactly for the linear regression or approximately for the nonlinear regressions and, therefore, will be able to estimate true covariate effects.

Regression calibration (3)

- In absence of true intakes, each predictor T_i^{*P} is estimated in a substudy (called calibration study) using (often repeat) reference measurements R_{ij}

- Requirement: reference measurements may contain error but should be unbiased for true individual usual intake, i.e., for person i, repeat j

$$\mathbb{E}(R_{ij} \mid i) = T_{ki}$$

- Regression calibration predictor is given by

$$T_i^{*P} = \mathbb{E}\left[g_T\left(T_i\right) \mid \mathbf{X}_i \right] = \mathbb{E}\left[g_T\left\{ \mathbb{E}\left(R_{ij} \mid i\right)\right\} \mid \mathbf{X}_i \right]$$

and its estimation requires a model for R_{ij}

Slide 23

Now, the main question becomes: How do you estimate this best predictor or this conditional mean? The best way to do it would be to have a substudy where the truth is observed. Unfortunately, it's not the case with dietary data. So in the absence of true intakes, each such predictor is estimated in a substudy, which is usually called a calibration substudy, using reference measurements, often repeat measurements. I will call them R.

And there is a requirement for a correct reference. The reference measurements may contain error of their own, but they should be unbiased for true individual usual intake. In other words, for each person, i, given all personal characteristics, the mean of the repeated reference measurements should be equal to true usual intake. And if this requirement is met, then the predicted value of true intake is given by this formula. Remember, this is the conditional mean of the transformed true intake given the values of observed covariates. This unknown true intake could be replaced by the conditional mean of the reference measurements.

Basically, this conditional mean means that we have to take many, many, many, many repeats for each person, i, and average them over. Of course, in real life we don't have those many, many, many repeats so we will need to use a statistical model for reference measurements, and then we can calculate this conditional mean by averaging, or doing weighted averaging—averaging with weights—the available repeats over the distribution of those repeats. And in mathematics this weighted average is an integral and it could be calculated numerically if it doesn't exist in closed forms.

Regression calibration (4)

- 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)

So the bottom line is that to apply this formula, one needs to model reference measurements. Before I go to modeling, I would like to say something about the reference measurements themselves.

The ideal reference measurements of dietary intake are short-term recovery biomarkers. Unfortunately, at the moment, only a few are known: doubly labeled water for energy intake, urinary nitrogen for protein intake, and urinary potassium for potassium intake.

What do we use in practice? Well, we use, usually, more detailed short-term self-reports, such as multiple-day food records or repeat 24 hour dietary recalls.

Regression calibration (5)

- Methodology below is developed for any correct reference measurement

- 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

Slide 25

The methodology that I'm going to present today is developed for any correct or valid reference measurement. I will demonstrate this methodology using the 24 hour recall, which is the reference instrument in many important dietary cohorts, including the NIH-AARP cohort or EPIC cohort in Europe. The short-term reference period for a 24 hour dietary recall is one day. And I will consider what I call a working assumption, and that is that the 24 hour recall is unbiased in reporting individuals' true usual intake and therefore is a valid reference instrument.

I will address the implications of possible biases in the 24 hour recall at the end of this lecture.

Regression calibration (6)

- Given a measurement error model for reference measurements, regression calibration predictor is a function of covariates $\mathbf{X}_i$ and model parameters $\boldsymbol{\theta}$

$$T_i^{*P} = \mathbb{E}\left[g_T \left\{ \mathbb{E}\left(R_{ij} \mid i \right) \right\} \mid \mathbf{X}_i \right] = \mathfrak{T}(\mathbf{X}_i; \boldsymbol{\theta})$$

- Parameters are estimated in a calibration substudy as $\hat{\boldsymbol{\theta}}$

- Regression calibration predictor is then estimated for all subjects in the main study as

$$\hat{T}_i^{*P} = \mathfrak{T}(\mathbf{X}_i; \hat{\boldsymbol{\theta}}), \ \ i = 1, ..., N$$

Slide 26

If we have a measurement error model for the reference measurements, then the regression calibration predictor given by this formula depends on observed covariates, which are components of vector X_i, and, of course, the parameters of the model.

I denote this function by a special form of letter *T*; it's a fancy T, which denotes a function that depends on X and the covariates of the model.

When I fit this model in a calibration substudy, I can estimate the model parameters by $\hat{\theta}$ [can't make symbol with "hat" on top]. And then I can calculate the regression calibration predictor for each person in the main study by using these estimated parameters and fit the risk model to those variables.

Regression calibration (7)

Regularly-consumed dietary components

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

$$R_{ij} = T_i + \varepsilon_{ij}, \ \varepsilon_{ij} \sim N\left(0; \ \sigma^2_\varepsilon\right)$$

where the regression of T_i on $\mathbf{X}_i$ is linear, i.e.,

$$T_i = \beta_0 + \boldsymbol{\beta}^t_X \mathbf{X}_i + u_i, \ u_i \sim N\left(0; \sigma^2_u\right)$$

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

$$R_{ij} = \beta_0 + \boldsymbol{\beta}^t_X \mathbf{X}_i + u_i + \varepsilon_{ij}$$

Slide 27

I would like to introduce measurement error models for short-term reference instruments, starting with the ideal situation when the references are true usual intake plus error. This error is normally distributed and has mean zero and a constant variance. Statisticians usually call this the classical measurement error model. Moreover, in this ideal world the true usual intake has a linear regression on the vector of covariates with a residual, u, which has a normal distribution with mean zero and a constant variance. Putting those two expressions together, the measurement error for the reference measurements in this ideal world is given by this formula.

Regression calibration (8)

- Measurement error model

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

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

  - **fixed** (in this case linear) **effect** of covariates defined by the population-level parameters (β_0, $\boldsymbol{\beta}_X$)

  - **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** ε_{ij} representing short-term variation

Slide 28

Here is this formula again. This model is a mixed effects linear model, which includes fixed (in this case, linear) effect of covariates, and this linear effect is defined by the population-level parameters, intercept β_0 and slope β_X. It also includes a random effect, u which represents a part of the within-person mean of R that is not explained by the covariates. It's person-specific, i.e., it is constant for each person but it randomly varies across people. And, of course, like in most statistical models, there is a within-person random error, ε, representing short-term variation in the reference measurement.

Regression calibration (9)

- Regression calibration predictor on a transformed scale is given by

$$T_i^{*P} = \mathbb{E}\left[g_T\left\{ \mathbb{E}\left(R_{ij}\mid i\right)\right\} \mid \mathbf{X}_i \right] = \mathbb{E}\left[g_T\left\{ \beta_0 + \boldsymbol{\beta}_X^t \mathbf{X}_i + u_i \right\} \mid \mathbf{X}_i \right]$$

- 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

$$T_i^P = \beta_0 + \boldsymbol{\beta}_X^t \mathbf{X}$$

Slide 29

The regression calibration predictor, generally on the transformed scale, is given by this formula. Remember I said that we could replace unknown true values by the conditional means of the reference measurements. And under this ideal model, there is a finite form for expressing this conditional mean given over here by this expression, and then one has to calculate the transformation of this and after that to calculate the conditional mean of the transformed value for given values of covariates, X. In a general case, this conditional mean doesn't exist in closed forms but could be calculated numerically, as I mentioned before.

There is one case when it does exist in closed forms, and this case is the one when the transformation is the identity function; in other words, when we consider the risk model on the original scale and then the regression calibration predictor is given by a linear function of covariates, the simplest predictor possible.

Regression calibration (10)

- **Real world:** often within-person random error in R_{ij} depends on individual mean and has a skewed distribution, violating classical model assumptions

- Usual remedy: transformation to a scale where classical model is a good approximation, i.e.,

$$g_R\left(R_{ij}\right) = \beta_0 + \boldsymbol{\beta}_X^t \mathbf{X}_i + u_i + \varepsilon_{ij}, \ u_i \sim N\left(0; \sigma_u^2\right), \ \varepsilon_{ij} \quad N\left(0; \quad _\varepsilon\right)$$

- Regression calibration predictor is then given by

$$T_i^{*P} = \mathbb{E}\left[\, g_T\left\{\mathbb{E}\left(R_{ij}\mid i\right)\right\}\mid \mathbf{X}_i\,\right]$$

$$= \mathbb{E}\left[\, g_T\left\{\mathbb{E}\left(g_R^{-1}\left\{\beta_0 + \boldsymbol{\beta}_X^t \mathbf{X}_i + u_i + \varepsilon_{ij}\right\}\mid i\right)\right\}\mid \mathbf{X}_i\,\right]$$

Slide 30

Well, that was an ideal world. Unfortunately, the real world sometimes is very different. Often, within-person random error in reference measurements depends on the individual mean and has a very skewed distribution, therefore violating classical model assumptions. The common remedy in this case is to transform the reference values to a scale where the classical model is a good approximation. So what we do is perform the transformation, g_R, such that the mixed effects linear model still holds on the transformed scale.

And then the regression calibration predictor is given by this formula, so we have first to transform back to the original scale, then calculate the conditional mean—again, it could be done numerically—then transform it to a scale where we fit the risk model, and again calculate the conditional mean of this expression given the covariates that are observed.

Linear regression calibration

- Linear regression calibration approximation (LRC): commonly used in nutritional epidemiology

- **Working assumptions**:

  - There is a scale where R_{ij}^* 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 $\mathbb{E}\left(R_{ij}^* \mid i\right) \approx T_i^*$

- Then $$T_i^{*P} = \mathbb{E}\left[\mathbb{E}\left(R_{ij}^* \mid i\right) \mid \mathbf{X}_i\right] = \beta_0 + \boldsymbol{\beta}_X^t \mathbf{X}_i$$

- We will see later that LRC may fail to provide a good approximation for nonlinear models

Slide 31

You have heard before and probably know from your experience that the most commonly used form of the regression calibration is the linear regression calibration. And I just showed you all those complicated formulas and said that it should be done by numerical integration. Why? Why bother?

Well, actually, the linear regression calibration in many cases works just fine, but not always. The reason is that it's based on two important working assumptions. The first assumption is that there is a scale where the transformed reference measurements are well approximated by the linear mixed effects model and where the risk model predictor is linear. In other words, the g_T and g_R are the same function; I call it g. Moreover, on this transformed scale, approximately, the transformed reference is unbiased for transformed usual intake. And, remember, we assumed that the reference instrument is unbiased on the original scale. It cannot be unbiased on both original and transformed scales. The quality of this approximate assumption depends on the transformation. If these two assumptions are met, more or less, even approximately, then the calibrated predictor of the transformed scale, again, is given by this simple linear function of the covariates.

As I mentioned, it may work pretty well in many cases, but not always, and we will see later that linear regression calibration may fail to provide a good approximation for nonlinear models.
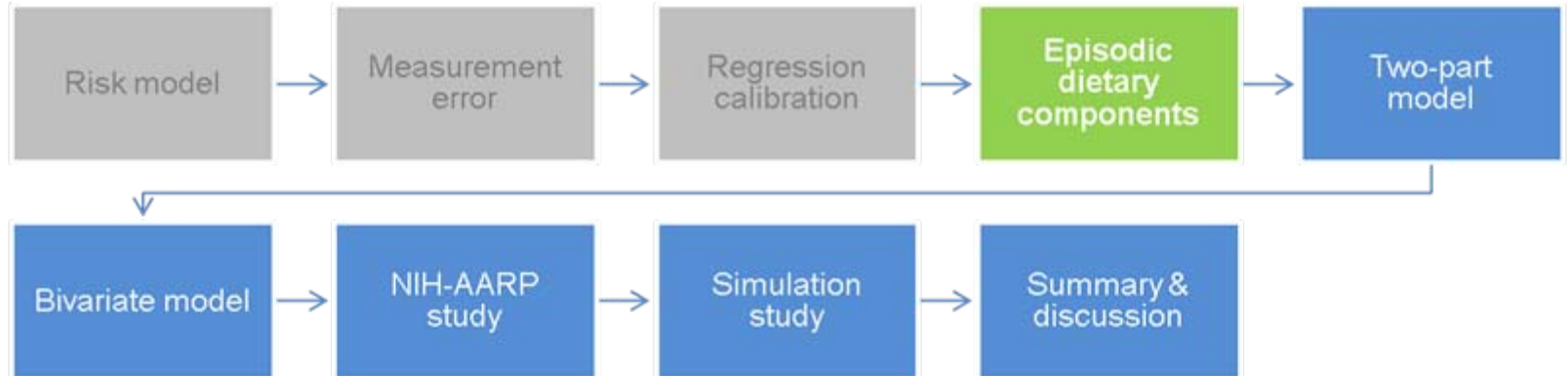
EPISODIC DIETARY COMPONENTS

Slide 32

Now, let me consider episodic dietary components, the main focus of today's webinar.

Episodic dietary components (1)

- 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)

Slide 33

The episodically consumed dietary components are the components that are not consumed daily by nearly everyone, but we assume that they are eventually consumed in the long run. The examples are many foods, such as fish, red meat, whole grains, dark green or orange vegetables, etc., and some nutrients, such as vitamin A or B12.

Episodic dietary components (2)

- ## Example: typical short-term report

measurementERRORwebinar series
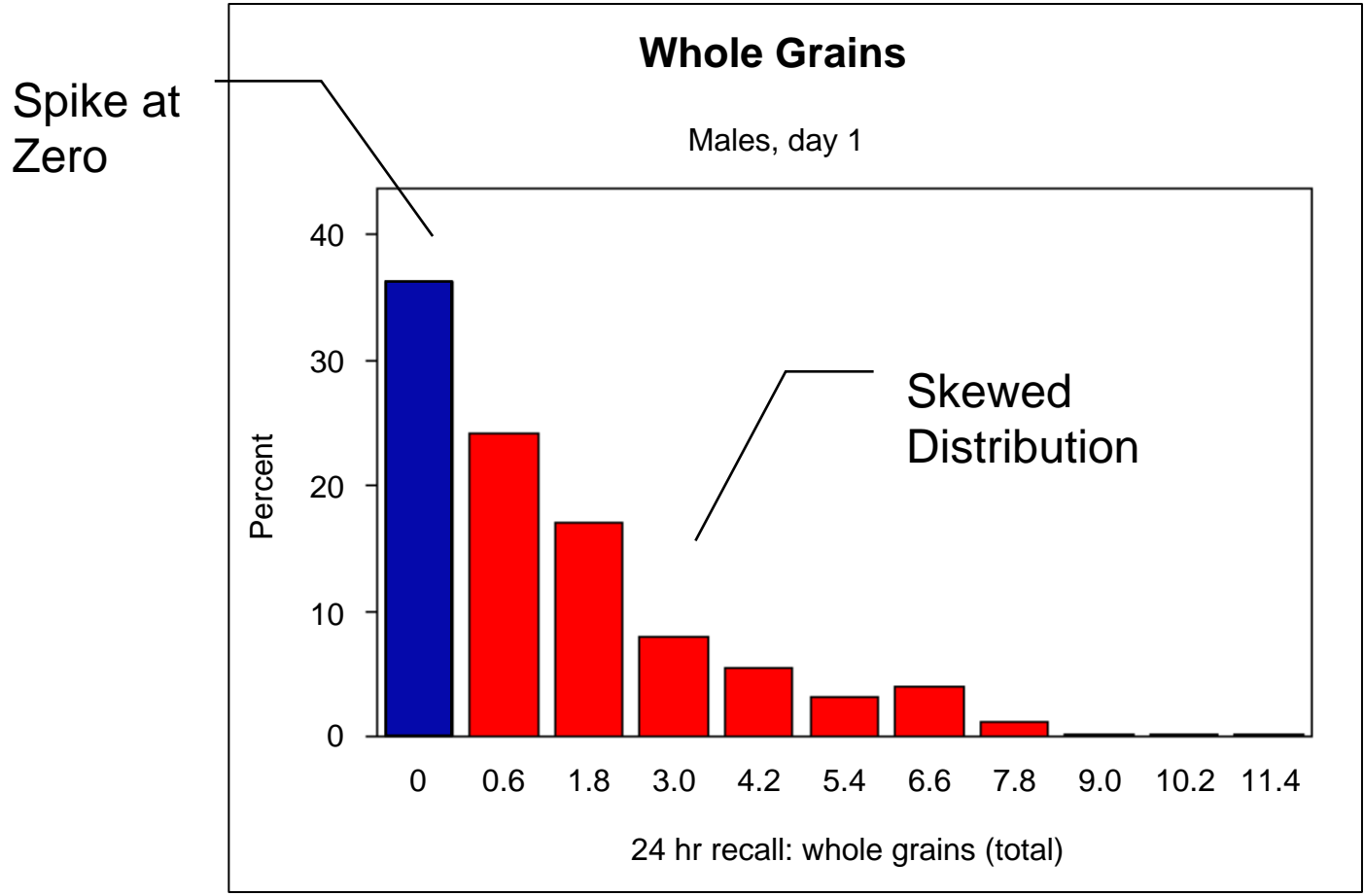
This is a typical example. This is a histogram from the Eating at America's Table Study of a short-term report on the 24 hour recall, one day of the 24 hour recall, of whole grains by men. And what you can see is that, first of all, you have about 36 or 37% of individuals reporting zero intake of whole grains on this particular day. And then the distribution of positive values is skewed to the right with a very ugly, long right-hand tail.

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

Slide 35

So it's a typical situation when short-term reference measurements for episodically consumed dietary components are what statisticians call semicontinuous variables with excess numbers of zeros and often skewed to the right positive values. And even if otherwise precise, they contain substantial within-person measurement error due to short-term variation in reported intake.
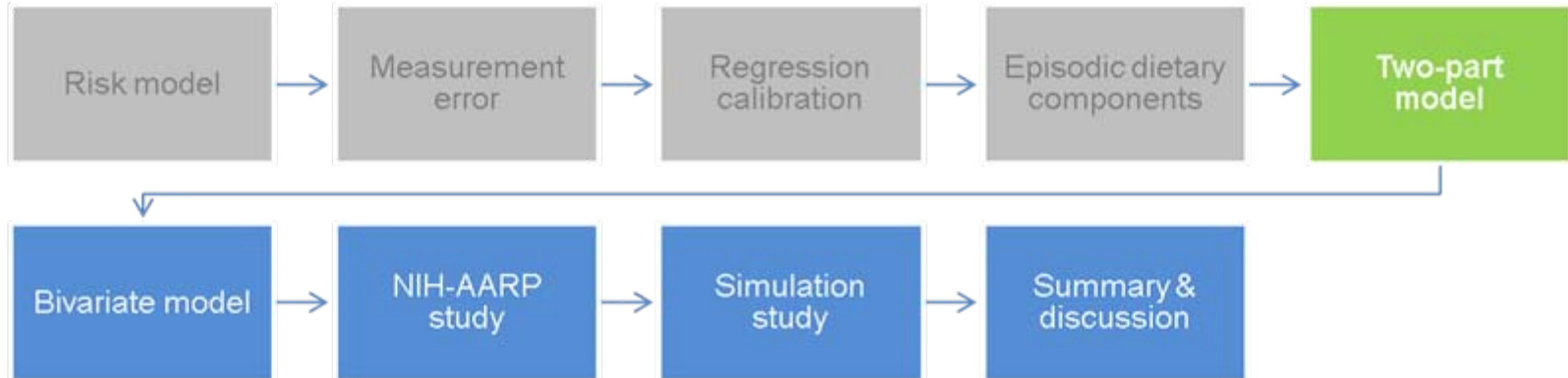
TWO-PART MODEL

Slide 36

So how would one model such data?

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

Slide 37

The goal is to specify a measurement error model for semicontinuous reference measurements. The main idea is to model a semicontinuous variable as the result of two distinct, although generally correlated, processes. One determines whether the variable takes positive or zero value, and the other one 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

This leads to a two-part model, which was first proposed in econometrics literature by Cragg in 1971, and it was intensively studied first in econometrics literature and then later on in biostatistical literature. The part I of this two-part model is usually the logit or probit regression specifying the probability of positive values. And the part II models the positive values through the linear regression of their log transformation.

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

Slide 39

This model was relatively recently extended to longitudinal data in 2001 by Olsen and Schafer and in 2002 independently by Tooze et al. by introducing a mixed effects two-part model. We are already familiar with the mixed effects linear model. In this case, the model may be nonlinear but still it has fixed effects that are defined by a function of covariates with population-level parameters; random effects that represent part of the within-subject mean which is not explained by the covariates, which is subject-specific but randomly varies across subjects; and 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

Slide 40

In the case of the longitudinal two-part model considered in those two papers, the part I was the mixed effects logistic regression specifying the probability of positive values. Part II was the mixed effects linear regression for log-transformed positive values. Both parts were linked by allowing correlated person-specific random effects in the two parts and overlapping covariates. And for model identifiability—in other words, to be able to uniquely estimate all model parameters—one would need at least two repeat observations on at least a subsample of the 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 skewedness)

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

  – Allowing for within-subject random measurement error

And then this longitudinal two-part model was further extended by the new methodology known as the NCI method to the case of modeling short-term reference measurements of episodically consumed dietary components. The extension was twofold. First of all, the log transformation was replaced by a much more flexible Box-Cox family of transformations given by this formula. Basically, it's a power transformation for all parameter values which are not zero, and for the zero parameter value the transformation is a log. And the second extension was allowing for within-subject random measurement error.

Two-part model (6)

- Two-part NCI model:

$$\mathbb{P}(R_{ij} > 0 \mid i) = H\left(\beta_{10} + \boldsymbol{\beta}_1^t \mathbf{X}_{1i} + u_{1i}\right), \; H(v) = \left(1 + e^{-v}\right)^{-1}$$

$$g(R_{ij}; \gamma_R \mid R_{ij} > 0) = \beta_{20} + \boldsymbol{\beta}_2^t \mathbf{X}_{2i} + u_{2i} + \varepsilon_{2ij}, \; \varepsilon_{2ij} \sim N\left(0, \sigma_{\varepsilon_2}^2\right)$$

where:

$$\mathbf{u}_i = (u_{1i}, u_{2i})^t \sim N\left(\mathbf{0}; \boldsymbol{\Sigma}_\mathbf{u}\right), \; \boldsymbol{\Sigma}_\mathbf{u} = \begin{pmatrix} \sigma_{u1}^2 & \sigma_{u1,u2} \\ \sigma_{u1,u2} & \sigma_{u2}^2 \end{pmatrix}$$

- 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

Here is the formula for the two-part NCI model. It was already introduced in the third webinar by Dr. Tooze. And what's important is that the first part specifies the probability of consumption; in this case, using the logistic regression. And the second part specifies the consumption amount. And on the transformed scale, the amount is given by the mixed effects linear model. The random effects have a normal distribution and are correlated, as is depicted here. And also, covariates in the two parts of the model could overlap.

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)
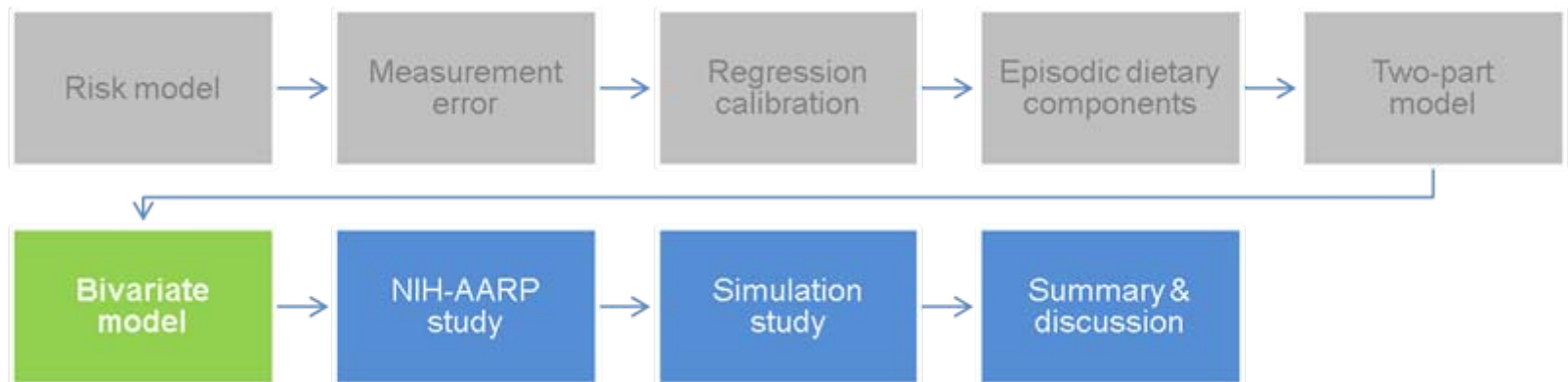
Slide 43

For a single episodically consumed dietary component, using two repeat 24 hour recalls in the U.S. survey, which is called NHANES for National Health and Nutrition Examination Survey, as the main dietary assessment instrument, the NCI method was applied to estimating the distribution of usual intake and its characteristics in a paper in *JADA* in 2006, and estimating the relationships of usual intake with health outcomes in a *Biometrics* paper in 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

Slide 44

Our goal today is to extend this NCI methodology for adjusting diet-health relationships for FFQ measurement error when the risk model includes several dietary components. Now, in many cases, regression calibration can be applied to error-prone covariates in a risk model one at a time, i.e., one by one. But in the case of diet-health relationships, there is a complication, and this complication is related to energy adjustment, which is often used with dietary exposures.

BIVARIATE MODEL

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

Slide 45

[No notes.]

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

To understand the effect of dietary composition, epidemiologists usually consider risk models with energy-adjusted dietary covariates such as the density, or the ratio of usual intake of interest to usual energy intake, or the residual from regressing usual intake of interest on usual energy intake. And energy itself is usually a covariate in the risk model as well.

What is achieved by this is we are looking at the effect of main exposure given that total energy intake is the same; in other words, when the change in the exposure is achieved by substituting the exposure value for some other dietary components. And since many dietary intakes are correlated with energy intake, estimation of the regression calibration predictor, in this case of density or residual, requires modeling episodic components and energy simultaneously.

Bivariate model (2)

- **Observed data** in calibration sub-study: for person i, time period j

  - R_{Fij}, R_{Eij} - short-term reference measurements of episodic dietary component F and energy E

  - $\mathbf{X}_i$ - vector of observed covariates, including FFQ-reported intakes $\mathbf{Q}_i$ and error-free covariates $\mathbf{Z}_i$

  - Indicator variable of reference consumption in period j for episodic component

$$I_{Fij} = I(R_{Fij} > 0) = \begin{cases} 1 & \text{if } R_{Fij} > 0 \\ 0 & \text{if } R_{Fij} = 0 \end{cases}$$

Before I go to this bivariate model, I would like to go over what is observed and what is not. The observed data consist in the calibration substudy of short-term reference measurements for dietary components of interest—in our case, episodic dietary components—and for energy. They also consist of vector X of observed covariates, which include FFQ reported intakes, denoted by Q, and error-free confounders, denoted by Z. And, also, one observes the indicator variable of reference consumption in any short-term period, j. In other words, we do observe whether the reported episodic component is consumed or not consumed during a specified time period.

Bivariate model (3)

- **Latent (unobserved) variables**:

  - T_{Fij}, T_{Eij} - true intakes of interest in period j

  - T_{Fi} - true usual intake of component of interest

  - T_{Ei} - true usual energy intake

  - $T_{Di} = T_{Fi} / T_{Ei}$ - true density intake of interest

- Additional latent variables: person-specific random effects and within-person random errors

What are unobserved are the variables which, first of all, are true intakes of components of interest or energy, and I will consider density in this presentation, so the density is given by the ratio of true usual intake of the component of interest to true usual intake of energy. And there are also additional latent variables in the model which consist of person-specific random effects and within-person random errors, as we saw previously.

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

As to energy intake, it's natural to specify it as part II of the NCI model since energy is always consumed. Part I, which is the probability of consumption, is always 1, so it doesn't have to be specified.

Now, allowing correlations between person-specific random effects in the energy model and in the model of the episodic components 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 components during short-term consumption periods.

Is this enough to come up with a good model?

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

Slide 50

Well, there is an additional model requirement. During any short-term period, energy intake, which is a continuous variable, should be allowed to be correlated with the indicator of consumption of the dietary component of interest, which is a binary variable for episodically consumed components.

This has an intuitive sense because, for example, on a day when one consumes, say, red meat, his or her energy intake could go up. And if one doesn't and consumes only, say, vegetables, the energy intake on that particular day or during this particular short-term period could go down. So one feels that this correlation might be important.

Now, the original part I of the NCI model specifies a model for the probability of consumption but not for the indicator or the fact of short-term consumption. And so to satisfy the mentioned requirement, one needs to modify part I of the two-part NCI model.

Bivariate model (6)

- **Modified part I**: to allow I_{Fij} & R_{Eij} to be correlated

  - Consider continuous **latent variable** in period j

  $$\tilde{R}_{Fij} = \beta_{F10} + \boldsymbol{\beta}_{F1}^t \mathbf{X}_i + u_{F1i} + \varepsilon_{F1ij},$$

  where $u_{F1i} \sim N\left(0, \sigma_{u_{F1}}^2\right), \ \varepsilon_{F1ij} \sim N(0,1)$

  - Let $\tilde{R}_{Fij}$ underlie binary indicator of episodic component's reference consumption

  $$I_{Fij} = 1 \Leftrightarrow \tilde{R}_{Fij} > 0$$

  - Allow ε_{F1ij} and within-person error in the model for reference energy intake R_{Eij} to be correlated

And to do this modification to allow the indicator of consumption, I, and the reference energy amount during a particular short-term period, j, to be correlated, let's consider a latent variable. It's an additional latent variable. I call it R_F tilde, which is represented by the linear mixed effects model, which we're familiar with already. The only difference here is that the variance of epsilon, or within-person variation, is supposed to be 1. One needs to make this assumption to be able to identify or uniquely estimate the model parameters.

And let us relate this latent variable to the indicator of episodic components reference consumption. In other words, we will say that consumption takes place if and only if this latent variable is positive.

And now, because we have a model for the fact of consumption, not just the probability to consume, and we have this within-person variation, ε_1, we could allow it to be correlated with its counterpart, within-person variation, in the model for energy.

Bivariate model (7)

- The use of additional latent variable $\tilde{R}_{Fij}$ leads to the probability of consumption specified as the mixed effects **probit** model

$$\mathbb{P}(R_{Fij} > 0 \mid i) = \Phi\left(\beta_{F10} + \boldsymbol{\beta}_{F1}^{t}\mathbf{X}_i + u_{F1i}\right)$$

  where Φ 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

Slide 52

By the way, if one introduces this latent variable in the form I did, it means that the probability of consumption now is given not by the logit but by the probit mixed effects model. Usually, the two models are in close agreement so it's not a big deal. I just want to emphasize that it's a little change from the original NCI model, where it was logit.

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:

$$g(R_{Fij}; \gamma_{R_F} \mid R_{Fij} > 0) = \beta_{20} + \boldsymbol{\beta}_{F2}^{t}\mathbf{X}_i + u_{F2i} + \phantom{_{F2ij}}{}_{F2ij}$$

where:

$$u_{F2i} \sim N\left(0, \sigma_{u_{F2}}^2\right), \ \varepsilon_{F2ij} \quad N\left(0, \sigma_{\varepsilon}\right)$$

- Part I and II are **linked** by using the same covariates and allowing person-specific random effects to be correlated

Okay, part II is unchanged. It's the same as in the original NCI model. It says that on the transformed scale, the positive reported values of episodic components are represented by the linear mixed effects model. And part I and part II are linked by using the same covariates, X, and by allowing person-specific random effects, u, to be correlated.

Bivariate model (9)

- **Model for Energy**: transformed energy amount for period j is specified as linear mixed effects model:

$$g(R_{Eij}; \gamma_{R_E}) = \beta_{E0} + \boldsymbol{\beta}_E^t \mathbf{X}_i + u_{Ei} + \varepsilon_{Eij}$$

where:

$$u_{Ei} \sim N\left(0, \sigma_{u_E}^2\right), \ \varepsilon_{Eij} \sim N\left(0, \sigma_{\varepsilon_E}^2\right)$$

- 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 ε_{Eij} to be correlated with ε_{F1ij} and ε_{F2ij}

measurementERRORwebinar series

Slide 54

The model for energy is, as I said, represented the same way as in part II of the NCI model; on the transformed scale, it's the linear mixed effects model. And this model is linked to both part I and part II of the episodic component model by using the same covariates, X's, and by allowing person-specific random effects to be pair-wise correlated and within-person error in the model for energy, ε_E, to be correlated with both ε_1 and ε_2 in the two-part model for the episodic components.

Bivariate model (10)

- Bivariate model is formally specified as:

$$R_{Fij} = I\left[\beta_{F10} + \boldsymbol{\beta}_{F1}^{t}\mathbf{X}_{i} + u_{F1i} + \varepsilon_{F1ij} > 0 \right]$$

$$\times \; g^{-1}\left(\beta_{F20} + \boldsymbol{\beta}_{F2}^{t}\mathbf{X}_{i} + u_{F2i} + \varepsilon_{F2ij} ; \gamma_{R_F} \right)$$

$$R_{Eij} = g^{-1}\left(\beta_{E0} + \boldsymbol{\beta}_{E}^{t}\mathbf{X}_{i} + u_{Ei} + \varepsilon_{Eij} ; \gamma_{R_E} \right)$$

where: $\mathbf{u}_{i} = (u_{F1i}, u_{F2i}, u_{Ei})^{t} \sim MVN\left(\mathbf{0}; \boldsymbol{\Sigma}_{\mathbf{u}} \right)$

$$\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{F1ij}, \varepsilon_{F2ij}, \varepsilon_{Eij})^{t} \sim MVN\left\{ \mathbf{0}; \begin{pmatrix} 1 & 0 & \sigma_{\varepsilon_{F1E}} \\ 0 & \sigma^{2}_{\varepsilon_{F2}} & \sigma_{\varepsilon_{F2E}} \\ \sigma_{\varepsilon_{F1E}} & \sigma_{\varepsilon_{F2E}} & \sigma^{2}_{\varepsilon_{E}} \end{pmatrix} \right\}$$

Slide 55

This bivariate or three-part model is formally specified on the slide. This is provided for those who would like to see this formal model specification, but I just want to emphasize one thing. The variance covariates matrix for within-person random variation has a structured form; namely, it has 1 as the first element of the main diagonal, and it has two zeros, here and here, and so when fitting the model, one should take care and not forget that this is not just a three-by three variance covariate matrix with parameters free to vary; it has a very specific and structured form.

Bivariate model (11)

- Denoting by θ_F, θ_E model parameters for episodic component and energy, respectively, we have:

$$R_{Fij} = \Re_F\left(\mathbf{X}_i, \mathbf{u}_{Fi}, \varepsilon_{Fij}; \theta_F\right), \; R_{Eij} = \Re_E\left(\mathbf{X}_i, u_{Ei}, \varepsilon_{Eij}; \theta_E\right)$$

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

$$T_{Fi} = \mathbb{E}\left\{\Re_F\left(\mathbf{X}_i, \mathbf{u}_{Fi}, \varepsilon_{Fij}; \theta_F\right) \mid i\right\} = \mathfrak{T}_F\left(\mathbf{X}_i, \mathbf{u}_{Fi}; \theta_F\right),$$

$$T_{Ei} = \mathbb{E}\left\{\Re_E\left(\mathbf{X}_i, u_{Ei}, \varepsilon_{Eij}; \theta_E\right) \mid i\right\} = \mathfrak{T}_E\left(\mathbf{X}_i, u_{Ei}; \theta_E\right)$$

Slide 56

Now, once we have specified the model, then the reference measurement for the episodic component is a function, denoted by this fancy R, of observed covariates; of a latent variable, which is the person-specific random effect; of another latent variable, which represents the within-person variation; and of model parameters. The same is true for the model for energy intake.

Bivariate model (12)

- True episodic component density is given by:

$$T_{Di} \equiv \frac{T_{Fi}}{T_{Ei}} = \mathfrak{T}_D(\mathbf{X}_i, \mathbf{u}_i; \boldsymbol{\theta}_F, \boldsymbol{\theta}_E)$$

- Regression calibration predictor for T_{Di}^* is given by:

$$T_{Di}^{*P} \equiv \mathbb{E}\left[g\left\{T_{Di}; \gamma_{T_D}\right\} | \mathbf{X}_i\right] = \mathbb{E}\left[g\left\{\mathfrak{T}_D(\mathbf{X}_i, \mathbf{u}_i; \boldsymbol{\theta}_F, \boldsymbol{\theta}_E, \gamma_{T_D})\right\} | \mathbf{X}_i\right]$$

- Regression calibration predictor for T_{Ei}^* is given by:

$$T_{Ei}^{*P} \equiv \mathbb{E}\left[g\left\{T_{Ei}; \gamma_{T_E}\right\} | \mathbf{X}_i\right] = \mathbb{E}\left[g\left\{\mathfrak{T}_E(\mathbf{X}, u_E; \boldsymbol{\theta}_E, \gamma_{T_E})\right\} | \mathbf{X}_i\right]$$

Now, to calculate the true usual intake of the episodic component, or energy, one needs to take the conditional expectation of this function, which is, as I said, the weighted average over the distribution of epsilons. And so it could be done numerically because this weighted average is just an integral.

And then, once we've calculated this weighted average, this becomes a function of observed covariates, person-specific random effects, and the model parameters. The same is true for energy intake.

And once we have those two, we can calculate their ratio and, again, it's a function of covariates, person-specific random effects, and model parameters. And, remember, what one needs to do is one needs to predict this on the transformed scale, generally speaking, given the observed covariates.

So to calculate this conditional mean, we need, again, to calculate the weighted average of this function over the distributions of u, because X is fixed. Again, it could be done by numerical integration. The same is true for the model for energy.

Bivariate model (13)

- Model parameters $\boldsymbol{\theta}_F$, $\boldsymbol{\theta}_E$ in the bivariate model are estimated by fitting the model in the calibration sub-study by MLE using NLMIXED SAS procedure

- For any given set of covariate transformations, by using estimated model parameters $\hat{\boldsymbol{\theta}}_F$, $\hat{\boldsymbol{\theta}}_E$, 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

Slide 58

In terms of model parameters, θ_F and θ_E in this bivariate model are estimated by fitting this model in the calibration substudy using the maximum likelihood estimation procedure, which is part of the NLMIXED procedure in SAS. For any given set of covariate transformations, by using these estimated model parameters from the calibration substudy, the regression calibration predictor for transformed density and energy could be calculated for each person in the main study. Then, the risk model could be fit for the given transformations of predicted covariates, and the final set of covariate transformations could be chosen by maximizing the overall likelihood when fitting the risk model.
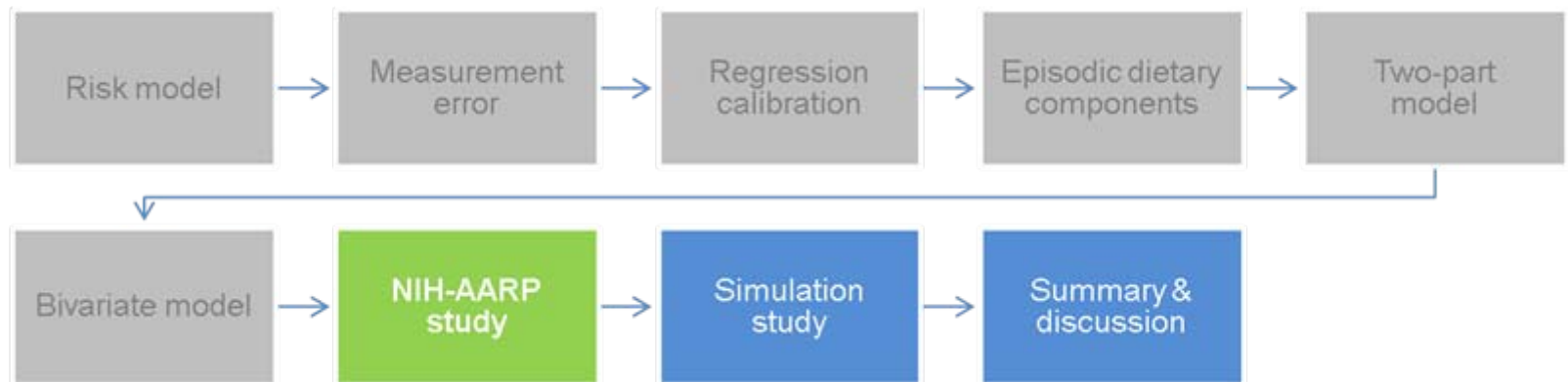
NIH-AARP STUDY

Slide 59

Now, let us consider how this methodology could be applied in the real situation in the NIH-AARP Diet & Health Study.

NIH-AARP diet and health 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

Slide 60

This is a prospective cohort of more than half a million men and women who were aged 50-71 in 1995-96 when the study began, with an FFQ administered at the baseline. This study has a calibration substudy of approximately 1,000 men and 1,000 women with an additional FFQ and two nonconsecutive 24 hour dietary recalls.

The analysis that I want to consider is the relationships in men between red meat density and lung cancer, and also orange vegetable density and lung cancer, adjusting for age, smoking, and energy intake.

The risk model was the Cox regression. It's a survival analysis—Cox regression on either original or Box-Cox transformed scales The standard errors were estimated by the bootstrap method. One needs to do so because one cannot use SAS-produced standard errors from fitting the Cox regression. The assumption behind SAS-estimated standard errors is that all the covariates in the model are known. In our case, we predict some of the covariates using the regression calibration prediction and because it's done in the calibration substudy, there are certain uncertainties in the estimated parameters. And to take them into account, one has to use the bootstrap or any such method.

NIH-AARP diet and health study (2)

- Due to zeros, for linear regression calibration (LRC), used Box-Cox transformation with a shift parameter:

$$g(v; \gamma_v) = \begin{cases} \dfrac{(v + \delta)^{\gamma_v}}{\gamma_v} & \text{if } \gamma_v \neq 0 \\ \log(v + \delta) & \text{if } \gamma_v = 0 \end{cases}$$

- 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

Slide 61

I would like to compare the methodology that I just presented with the commonly used linear regression calibration. And in the case of the linear regression calibration, because it basically consists of linearly regressing reference measurements on the FFQ and other covariates in the model, and the reference measurements contain lots of zeros, we have to use a different form of the Box-Cox transformation, a more general form which involves a shift parameter. I call it delta here. That allows you to, for example, take a log when the true values are zeros.

To assess whether after the transformation the risk model, Cox regression in our case, fits the data well, we apply the test using the cumulative martingale technique, which is implemented in SAS. And what it produces is the p value. If this p value is less than the conventional 5 percent, the fit may not be so good. If it's greater than 5 percent, it's an indication that the fit is okay.

In our examples we compare the FFQ-based analysis with no correction for measurement error with two types of corrections, first using the linear regression calibration and, second, using regression calibration based on the bivariate model that I just presented.

NIH-AARP diet and health study (3)

24HR reported consumption of orange vegetables and red meat

	Red Meat (g/day)	Orange Vegetables (cups/day)
Mean reported intake (s.e.)	82.8 (2.3)	0.14 (0.01)
Mean amount on consumption days (s.e.)	117.7 (2.6)	0.32 (0.01)
Mean probability to consume (s.e.)	0.70 (0.01)	0.44 (0.01)
% of subjects who consumed food:		
0 out of 2 days	14.7	33.5
1 out of 2 days	29.9	44.9
2 out of 2 days	55.4	21.6

measurementERRORwebinar series

Slide 62

Before I go to the results, I would like to show you this table, which contains mean reported intake, the standard errors, mean amount on the consumption days, mean probability to consume, and also the percentage of people in the study who consumed either red meat or orange vegetables on two days out of two, on one day out of two, or on zero days out of two.

You can see that about 33.5 percent of people in the calibration substudy reported no consumption of orange vegetables on any of the two days when the 24 hour recall was administered. Thus it's pretty episodically consumed, the mean probability of consumption was only about 40 percent.

With red meat the picture is slightly different. About 15 percent of the subjects didn't consume red meat, and we're talking about males, on any of the two days, which is interesting. But remember, this is the U.S. population and although they are middle-aged men, they still do eat red meat. So it's somewhat episodically consumed but not too episodically.

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

Risk Model: Measurement Error Correction Method	γ	δ	Estimated Log Hazard Ratio (s.e.)	Estimated Hazard Ratio (95% CI)	Risk model fit test p-value
Untransformed Intake:					
No correction for ME	1		0.225(0.040)	1.252(1.158,1.354)	0.041
RC (Bivariate model)	**1**		**0.409(0.075)**	**1.505(1.300,1.744)**	**0.130**
LRC	1		0.441(0.097)	1.554(1.285,1.880)	0.041
Transformed Intake:					
No correction for ME	0.4		0.248(0.046)	1.281(1.171,1.402)	0.082
RC (Bivariate model)	**1**		**0.409(0.077)**	**1.505(1.294,1.751)**	**0.130**
LRC (chosen scale)	0.4	0.001	0.321(0.155)	1.379(1.017,1.868)	0.076
LRC (fixed log scale)	0	0.001	0.113(0.096)	1.120(0.928,1.351)	0.564

Okay, here the table represents the results of the analysis for the red meat intake versus lung cancer in men. We have the estimated log hazard ratios in this column, and hazard ratios and 95 percent confidence intervals in this column. This column is, then, the value of the Box-Cox transformation parameter. And remember, in the case of the linear regression calibration, we also use the shift, which is represented in this column.

Now, we'll first consider the case when we fit the Cox regression on the original scale. If one uses the FFQ on the original scale, and the value of the Box-Cox transformation parameter for the original scale is always 1, the estimated log hazard ratio is .225; the estimated hazard ratio is 1.2. It's statistically significantly different from 1, so it's a risk factor. But when one calculates the model fit, or whether the scale is an appropriate scale, the p value is slightly less than .05.

Strictly speaking, it means that the model may not fit very well, although keep in mind it's a very large study with a quarter million men, and so when you are dealing with a large sample, any statistical test sooner or later will show deviations from the null hypothesis. And in this case, the null hypothesis is that the model fits on this scale. So when the value is close to 5 percent, the fit, although strictly speaking is not great, could be rather appropriate.

Now, if one uses the regression calibration based on the bivariate model, the results change. The hazard ratio becomes 1.5, so the effect is much larger, actually, as compared to the one estimated by using FFQ without correction. And, interestingly, the model fits rather well.

What about linear regression calibration? In this particular case, the results are very, very similar. Again, we estimate the de-attenuated hazard ratio and, again, the fit, strictly speaking, is not that great but may be rather appropriate.

Yet, because of those two p-values that indicate a not so great fit, let's see what happens if one does transform the main exposure variable. If one uses the FFQ, the transformation parameter is .4, so strictly it's close to—well, not strictly, but approximately—it's close to the square root. The estimated hazard ratio is very similar to the one estimated on the original scale: 1.25-1.28. And the model fits okay.

If one uses the regression calibration based on the bivariate model, the chosen transformed scale happens to be the original scale; in other words, it's an identity transformation. And of course, we already saw this result up here.

Now, with the linear regression calibration, one could apply two different transformations. In the literature, epidemiologists often use linear regression calibration on the log scale, always on the log scale. In other words, they don't choose this parameter in the Box-Cox family; it's fixed at zero, which represents log. But because

we're using shift, shift is still being chosen by maximizing the fit of the risk model, and in this case we add just a very, very little amount to the red meat.

Now, the results of those two different transformations are also very different. If one uses a log scale instead of deattenuating the hazard ratio estimated with FFQ, one attenuates it a little bit further and it becomes null. If one chooses the scale with the linear regression calibration, again one chooses something close to the square root, and again the shift has a very, very small value. And the result of this is still deattenuation. The effect is not as large as using the bivariate model, somewhat different, and the fit of the model is okay.

So basically, from this slide one can conclude that with or without transformation, the results for the bivariate model and linear regression calibration are not that different unless one uses a log scale with the linear regression calibration. Then, of course, the results differ very much.

[Page intentionally blank.]

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 0.02 to 0.10 cups/1000 kcal

| Risk Model: Measurement Error Correction Method	γ	δ	Estimated Log Hazard Ratio (s.e.)	Estimated Hazard Ratio (95% CI)	Risk model fit test p-value
Untransformed Intake:					
No correction for ME	1		-0.076(0.021)	0.927(0.889,0.966)	<0.0001
RC (Bivariate model)	**1**		**-0.265(0.078)**	**0.767(0.658,0.894)**	**0.002**
LRC	1		-0.223(0.086)	0.800(0.676,0.947)	<0.0001
Transformed Intake					
No correction for ME	-0.3		-0.182(0.030)	0.834(0.786,0.884)	0.256
RC (Bivariate model)	**0.1**		**-0.380(0.089)**	**0.684(0.574,0.814)**	**0.060**
LRC (chosen scale)	-20	1	-0.593(0.146)	0.553(0.415,0.736)	0.202
LRC (log scale)	0	0.005	-0.387(0.107)	0.679(0.551,0.838)	0.022

What about orange vegetables? Well, here we estimated the hazard ratio for increasing the orange vegetable density from .02 to .10 cups/1,000 kilocalories. Orange vegetables are presumably a factor which helps to lower the hazard ratio for the lung cancer, which means that the estimated log hazard ratio would have a negative sign, as we see here in this table. So on the untransformed scale, the situation now is quite different. In what sense? Well, the model doesn't fit on this scale. And we can still look at the results and we can still say that with the FFQ the results are very close to—the hazard ratio is very close to 1. This is being deattenuated using the bivariate model, very similar to deattenuating using the linear regression calibration. But if the model doesn't fit, do those numbers represent true hazard ratios? Probably not.

So let's consider some transformations. With the FFQ without any correction, the estimated hazard ratio is .83, so it's different from .92, the effect on the original scale, and the model fits okay. And the chosen scale is very close to the log scale; it's tenth root. With LRC, if one chooses the log scale, one gets the results very similar to those with the bivariate model. Unfortunately, the model doesn't fit. If one chooses both the scale and the shift parameter, the results of the FFQ are deattenuated even further than those based on the bivariate model, and the model fits. The results are rather different.

And if you remember with the previous analyses, the log scale produced results very similar to those with the bivariate model. In this particular case, the log scale produces the results similar—sorry—in the previous case it was not similar; it was very different. Let's go back and see. Yes, it was very different. But on the chosen scale it was somewhat closer. Here, it's closer on the log scale. On the chosen scale, it's not.

Of course, in the real example, we don't know what truth is. We don't know which of the two approaches—one which is based on the bivariate model, or linear regression calibration—works better.
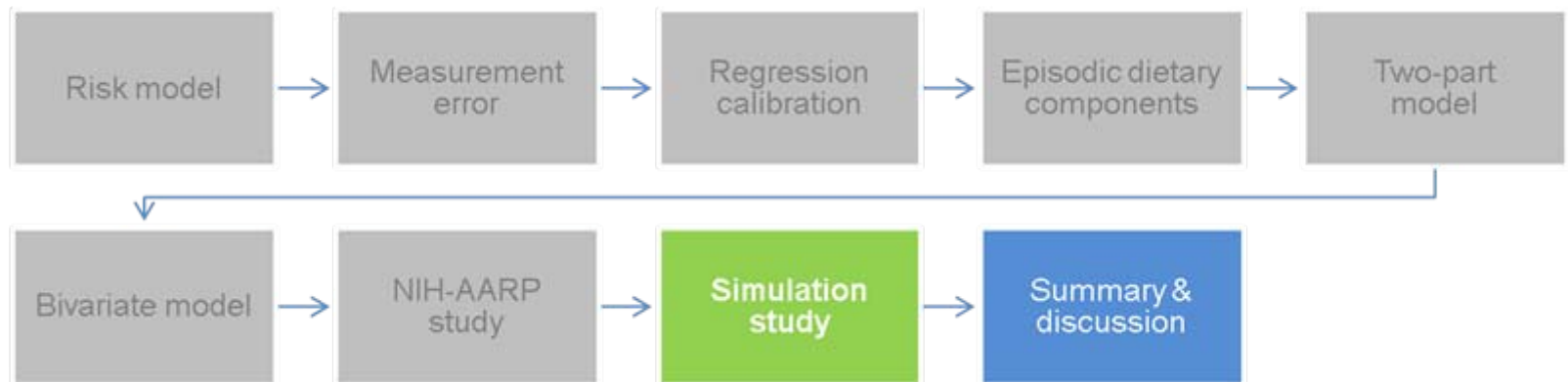
SIMULATION STUDY

So to examine it, we did a simulation study where the truth is known.

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

Slide 66

And so we conducted such a study. We generated the FFQ for the 100,000 individuals. We also generated 1,000 24 hour recalls, and in both cases the distributions of the simulated data were similar to those of orange vegetables and energy in the NIH-AARP Diet & Health Study.

The calibration substudy consisted of 1,000 subjects using the first two 24 hour recalls as reference measurements. Now, why did we simulate 1,000? Because we wanted to estimate true usual intake and, as I said, we assume that true usual intake is the mean of many, many repeats. So with 1,000 repeats, we just calculated this mean, and then we calculated the density by dividing the true usual intake for orange vegetables, pseudo-orange vegetables, over the true usual intake of energy.

We also generated the binary outcome using the logistic regression on Box-Cox transformed exposure.

Simulation study (2)

- 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

Our goal was to estimate the log relative risk for increasing main exposure between .02 and .10 cups/1,000 kilocalories, like in the earlier example. The risk model here was a logistic regression on the original and Box-Cox transformed scales. Standard errors were estimated by bootstrap. And we compared the FFQ-based analysis with no correction for measurement error with corrections using linear regression calibration and regression calibration based on the presented methodology.

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

Sim	Measurement Error Correction Method	Mean γ (Mean δ)	Mean Log OR (s.e.)	Standard Deviation	RMSE
1	True parameters	1	-0.4		
	No correction for ME	0.42	-0.242 (0.002)	0.025	0.160
	RC (Bivariate model)	**0.85**	**-0.417(0.005)**	**0.078**	**0.080**
	LRC (original scale)	1	-0.436(0.008)	0.118	0.123
	LRC (log scale)	0 (0.08)	-0.557(0.008)	0.108	0.191
	LRC (chosen scale)	-0.61 (0.18)	-0.543(0.007)	0.104	0.177
2	True parameters	0.1	-0.4		
	No correction for ME	0.02	-0.207(0.002)	0.024	0.194
	RC (Bivariate model)	**0.10**	**-0.416(0.007)**	**0.093**	**0.094**
	LRC (original scale)	1	-0.285(0.006)	0.088	0.145
	LRC (log scale)	0 (0.006)	-0.342(0.006)	0.083	0.101
	LRC (chosen scale)	-2.18 (0.14)	-0.375(0.009)	0.121	0.124

Slide 68

Here are the results. I present here two scenarios: one when the true logistic regression was fit using the original scale, and one when it was fit using the transformed (almost log) scale. Again, negative log odds ratio means that the exposure has a protective effect.

Let's look at this panel first. So when there is no correction for measurement error and one chooses the scale, the chosen scale is close to the square root, as we saw in the real example. The mean estimated log odds ratio is minus .24. This is the standard deviation. It's quite small, which is usually the case with using the FFQ without correction.

And in this column, I present what is called the root mean squared error. The root mean squared error somehow balances the bias in the estimated parameter and its standard deviation on its variance. For example, in using FFQ without correction, the estimated log odds ratio is very biased. It's almost half of the true value. But its variation, as depicted by the standard deviation, is rather low. Yet, because of the high bias, the root mean squared error is relatively large.

What happens if one uses the presented methodology, or bivariate model? The mean estimated transformation parameter is close to 1—not exactly, but close. The estimated mean log odds ratio is slightly biased, but very slightly. It's very close to minus .4. The standard deviation is larger than in the case of FFQ but not bad. As a result, the root mean squared error is only half of the root mean squared error of no correction when using FFQ, which is a very good reduction.

What about the linear regression calibration? Now, we have three choices. One is to consider always the original scale. The other one is to consider always the log scale but chose the shift parameter. And the third one is when both the scale and shift parameters are being chosen.

When one does the analysis on the original scale, the results are slightly more biased than with the bivariate model, but not by much. The standard deviation increases and, as a result, the root mean squared error is 50% higher compared to the bivariate model. If one uses the log scale to fit the risk model, one over adjusts the log odds ratio. So instead of minus .4, one gets minus .56. The variation is slightly less but not by much. Because of the bias, the root mean squared error is larger in this case. It's even larger than for the FFQ without any correction.

If one applies a linear regression calibration with both choice of the scale and the shift parameter, the estimated log odds ratio is still biased, a little bit less so than in the previous case. The root mean squared error is still large.

Okay, so based on this panel, it seems that the best strategy is to use linear regression calibration on the original scale; in other words, to do the right thing, although we don't

know it in advance, but we could always use the original scale. Why not? Well, the second panel shows why not. Because if the true scale is actually close to the log, with no correction we basically choose, on average, the right scale, yet with no correction for measurement error while using the FFQ, the results are very biased, half of the true value, and the root mean squared error, as a result of this, is pretty large.

With the bivariate model, on average, we choose the right scale. Look at this. The mean log odds ratio is very close to the true value. This variation of the estimate increases somewhat compared to the previous case but not by much. As a result, the root mean squared error again is half of the previous case.

Now, what about the linear regression calibration? Remember on the previous panel, if you fit with the original scale, you get results close to true ones. Here, if you fit in the original scale, you have huge bias. Basically, you don't deattenuate the FFQ-based result by much. As a result, the root mean squared error is pretty large.

If you do it on the log scale, there is still bias but less so. The root mean squared error improves. If you choose both the scale and the shift parameter, you get the results with a slight bias but rather close to the true value. Unfortunately, the variation of estimated log odds ratio is quite large. As a result, the root mean squared error is pretty large.

The main point here is that with the linear regression calibration, it seems that there is no consistent strategy of how it should be applied. On this panel, the original scale produces good results; the other two scales, not so much so. In the other case, the original scale failed very miserably. Using the log scale improved the result. Choosing the scales improved it even further, although the variation may not be so good in this particular case.

[Page intentionally blank.]

Simulation study: results (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

Well, in theory, for the risk model on the original scale, linear regression calibration is approximately consistent. Unfortunately, this consistency doesn't kick in for the finite samples, even as large as in the simulated study. And so for the finite samples, it still involves biases due to unaccounted excess zeros.

For the risk models on the transformed scale, when one applies the linear regression calibration, we also may have a problem. Why? Because trying to find a scale where both calibration and risk models have linear predictors and the reference measurements are unbiased—those two conditions that are important—often leads to poor approximations. If one does it without choosing the scale but on the original scale, by definition, one misspecifies the true scale of the risk model.

SUMMARY & DISCUSSION

Slide 70

[No notes.]

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

Slide 71

The developed methodology addresses major challenges for bivariate modeling of short-term reference intake of an episodic component and energy by allowing during any short-term period energy intake to be correlated with the indicator of episodic component consumption and by allowing energy intake to be correlated with non-zero episodic component consumed.

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

The developed methodology also allows for the rigorous regression calibration to correct for nondifferential covariate measurement error in rather flexible risk models which a) include multiple dietary exposures that could be energy adjusted, and b) the risk models could include covariates on the transformed scales.

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)

Slide 73

Our focus was on episodically consumed dietary components and we assumed that although they are episodically consumed during any short-term period, eventually they are consumed by nearly everyone in the long run.

What about never consumers? Those probably exist in some cases; for example, alcohol intake, [and] maybe even intake of dark green vegetables or orange vegetables or red meat. The model could be relatively easily extended to include never consumers, but depending on the dietary component and the reference instrument, it may require more than two repeats of the reference. Never consumers required four to six 24 hour recalls. With two 24 hour recalls in some cases in our simulation study, the model did not converge.

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

Now, the methodology is based on the important assumption that the short-term reference instrument is unbiased for true usual dietary intake at the individual level, which means that it's a valid reference instrument. In the application that I considered, such an instrument was a 24 hour recall.

Now, studies with recovery biomarkers such as doubly labeled water for energy and urinary nitrogen and urinary potassium for protein and potassium intakes, respectively, have demonstrated that 24 hour recall involves some biases.

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

[A paper] in the *JNCI* this year by Freedman and colleagues, which was based on the OPEN biomarker study, suggests that in spite of those biases, if one uses 24 hour recall as a reference to correct for the FFQ measurement error, this on average leads to better results than relying on FFQ without correction. And this study was conducted considering different scenarios that involved risk models with multiple dietary exposures. So it seems, based on this study, that using 24HR to correct for FFQ measurement error leads to better understanding of diet-health outcome relationships. It's still, at least on average, better than not adjusting at all.

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

Now, using reference instruments to calibrate the FFQ approximately removes bias due to measurement error, but it never fully restores the power to detect a relationship, the power which is lost due to measurement error in the FFQ. Even bias correction may not be very reliable if the attenuated effect is too small because of some unmeasured confounders that could be involved in the risk modeling. This is a well-known weak signal problem.

So those two considerations mean that one could do better—why not using the short-term instruments in the main study, especially now when there are several Web-based 24 hour recalls, such as the, for example, ASA24 developed at the National Cancer Institute? Another possibility is to combine different instruments—for example, Web-based 24 hour recall and FFQ—to achieve the same purpose.

And the methodologies that are involved will be presented in webinars 10 to 12.

QUESTIONS & ANSWERS

Moderator: Kevin Dodd

Please submit questions using the *Chat* function

Slide 77

Thank you Victor. We'll now move on to the question and answer period of the webinar.

Measurement Error Webinar 8 Q&A

Question: **What is the best way to determine if a dietary variable is episodically consumed? When is it the case that you need to use these multipart models? Is there a rule of thumb that you use for that?**

Basically, assuming that one has a reference instrument which is 24 HR, let's say—and if you have repeats, you look at those numbers, and if the percentage of people who report zero intakes on, say, two days, is maybe 1 or 2 or 3 percent, the nutrient or food may not be too episodically consumed and the usual methodology could be applied. On the other hand, we saw in our examples with red meat it was about 15 percent. With orange vegetables, it was more than 30 percent. Those foods are really episodically consumed. So is it a rule of thumb? To some degree, but looking at the reference measurements, one can make a judgment which usually is pretty good. *(V. Kipnis)*

You showed the linear regression calibration doesn't have necessarily a consistent message. How do you think that's going to perform if the main exposure was one of these regularly consumed dietary components, rather than an episodic one?

That's a good question. Actually, cases of nonlinearity—the first one was because the risk model could be fitted on the transformed scale. The second one is that, in our case, the model for the predictor or regression calibration model could be very nonlinear because of the nature of episodic consumption. Of course, for the regularly consumed dietary components, this second consideration may go away, although one could still use transformations of the reported values. The first one, though, remains. So, all in all, if one fits a model on the original scale, linear regression calibration, in theory, provides consistent estimates. In practice, it depends on the size of the study. For very large studies with very large calibration substudies, I think the results are going to be pretty good. In other words, the approximation would be very good. If one fits a risk model on the transformed scale, then it depends on the transformation, although I think that in many cases the results could be rather good. But in combination, if you use both nonlinearities as in the

case of episodically consumed dietary components, as we saw, linear regression calibration may fail. *(V. Kipnis)*

How do you set up a calibration study for these big cohorts with FFQ as the main instrument? In the example you gave with the AARP study, you had two 24 hour recalls in the calibration substudy in that situation.

That's a very important question because of EPIC. EPIC has a very large calibration substudy, [with] more than 35,000 people, I think, in it. But as you said, it was one 24 hour recall. Now, with one 24 hour recall used as a reference, one doesn't have a longitudinal study. One is one. So one could go and fit the original two-part model that was suggested for non-longitudinal data; in other words, that doesn't have person-specific random effects. Those random effects are very important. Not only they model unexplained by covariate variation in true usual intake, but also, because they are allowed to be correlated, they therefore link together the probability to consume and the consumption amount. And as you, Kevin, explained in your webinar, those two are often correlated. So when one uses the two-part model without person-specific random effects, the only way to do this correlation would be through the same covariates in the model. And I think it would be interesting to try it and apply it and see how much you would lose by applying such modeling. But as I said, in principle, it could be done. In EPIC it is the only way it could be done. *(V. Kipnis)*

I think that may have been Janet Tooze's webinar, not mine, that talked about the correlation. That's all right; I don't want to have you give me credit for something I haven't done. *(K. Dodd)*

Just to follow up on that, then, so, with one 24HR, you could do an EPIC if you change things around a little bit, and obviously you can do it with two 24HRs, because you just did. But is there some sort of a rule as to how many 24HRs would be optimal to use?

What would be better—to increase the number of repeats of the reference instrument, or to increase the size of the calibration substudy? Because, remember, the parameters of the measurements in our model

are being estimated, and so the larger the calibration substudy, the better you estimate them. And the results seem to be indicating that it's better to use a larger study rather than to increase the number of repeats. Actually, the EPIC study was designed based on this consideration. Now, with episodically consumed dietary components, one needs at least two repeats. To model never-consumers, one might have to increase the number of 24 hour recalls to, say, four. But what I would like to emphasize is that as far as I know there have not been any studies that have looked at this theoretically, and it probably needs to be done. *(V. Kipnis)*

There have been some questions asked about sort of the complication effect of this model. And there's one question based on why couldn't one use a univariate model that doesn't actually have all of these additional complications to it, instead of the bivariate model?

In the case of energy adjustment, let's consider density, for example. We define this density as the ratio of usual intake of a dietary component of interest to the usual energy intake, instead of the usual ratio. Depending on the dietary component and depending on the transformation involved, those two values could be very close together or could be rather far apart. So the answer to your question is, in some cases, if you do it univariately you may not see too much of a difference, but there will inevitably be cases where the difference would be sizeable, so I would suggest applying this more advanced methodology. *(V. Kipnis)*

And is it possible that after doing all of this best selection of transformation for the disease model, you do the best you can, but can you still get a situation where that model fit isn't very good? And if so, what do you do then?

Yes, you could. With all those transformations, I presume that it's a curvilinear relationship. It could be made linear if one does the transformation. It's conceivable that there are some other ways, of course, of considering more complex models. One can use nonparametric modeling with, e.g.,splines. One can use a Bayesian approach, which is more flexible. But in my experience, in many cases, finding a

transformation can lead to not too much complication and yet to a reasonable fit. *(V. Kipnis)*

As a corollary to that, what about doing some categorization?

Well, that's another very good question because in most epidemiologic studies, if you look at the analyses, people do use categorized exposure. And of course, the main reason, as I understand it, is to avoid the problem of specifying a form of a relationship or a form of the risk model predictor. Nothing comes free, though. There is no free lunch. When you categorize, you lose a lot of power unless your study is very, very, very large. And that is not such a good thing. Another thing is, then, that when you categorize, the induced misclassification is going to be differential. And so regression calibration, which we consider here, could not be applied. There are other methods for measurement error correction which could be applied in the differential case, in the case of differential misclassification; for example, multiple imputations. They are more involved, so you win by considering a simple risk model; you lose by the necessity to consider a much more involved method for measurement error correction. So, all in all, it may not be such a good idea, if not always. *(V. Kipnis)*

What happens if, after measurement error correction, your relationship is no longer statistically significant? What does that mean?

It could mean two things. It could mean that your calibration substudy was too small. As a result, when you take into account uncertainty in estimating measurement error model parameters, this uncertainty is so large that the resulting confidence interval in the risk model becomes too large and so you don't have enough power to see the relationship. It could also mean a wrong procedure. When I applied the linear regression calibration on the log scale, the relationship between—I believe it was orange vegetables versus lung cancer—became statistically nonsignificant. And it was simply because the linear regression calibration was not the right approach. *(V. Kipnis)*

Slide 78

Thank you, Victor, and thanks to our audience for joining today's webinar. Please join us next week for webinar 9, the first session in the Advanced Methods section of our series, in which Dr. Raymond Carroll will discuss the estimation of usual intake distributions for multivariate dietary variables. Thank you.