National Cancer Institute

measurement ERROR webinar series

Assessing diet-health relationships: Focus on dietary components consumed daily by nearly all persons

Douglas Midthune, MS
National Cancer Institute

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

Slide 1

Hello and welcome to the seventh session in the Measurement Error Webinar Series. I'm Amy Subar, a nutritionist with the Risk Factor Monitoring and Methods Branch 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.

Before we get started with today's presentation, please note that 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; you can 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 the presenter for today's webinar. Doug Midthune is a mathematical statistician in the Biometry Research Group, Division of Cancer Prevention, at the National Cancer Institute. He is an integral member of the Surveillance Measurement Error Group at the National Cancer Institute, helping to develop the NCI method for modeling episodically consumed foods. Recently, he has played an important role in the extension of the method to accommodate simultaneous modeling of multiple nutrients and foods, with applications to both estimation of usual intake distributions and examination of diet and health relationships. Today Doug will discuss methods of assessing diet and health relationships using a food frequency questionnaire as the main dietary instrument and with a focus on non-episodically consumed dietary components. Doug.

Today's webinar is about "assessing diet-health relationships." Last week, Larry Freedman discussed the problems caused by measurement error when trying to estimate diet-health relationships. Today I'm going to talk about methods that can be used to address some of these problems.

Since this is the first webinar on this topic, I'll be focusing on relatively simple methods that can be used when the dietary variables are consumed nearly every day by nearly everyone. I'm also going to focus on the case when the main dietary instrument is a food frequency questionnaire, which is the case for most large cohort studies.

Future webinars will address more complex methods that may be required when some of the dietary components are episodically consumed, or when the main instrument is a 24-hour dietary recall. *(D. Midthune)*

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 webinar series is dedicated to the memory of Arthur Schatzkin, a colleague who collaborated with us for many years on the problem of measurement error in dietary assessment.

Presenters and Collaborators

Sharon Kirkpatrick
Series Organizer

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

Slide 3

And this is a list of the many people involved in this project.

Learning objectives

- Understanding:
  - That measurement error leads to bias in estimated diet-health associations
  - Concepts involved in regression calibration, a method to correct for this bias
  - The role of calibration studies in regression calibration
- Learning how to apply regression calibration in diet and health studies

Slide 4

The learning objectives for this webinar are: first, to understand that measurement error leads to bias in estimated diet-health associations; second, to understand the concepts involved in regression calibration, which is a method to correct for this bias; third, to understand the role of calibration studies in regression calibration; and, finally, to learn how to apply regression calibration in diet and health studies.

Main results on impact of measurement error

From webinar 6,

- When there is a **single** dietary exposure measured with error in a diet-health model:
  1) Estimated diet-health relationship (risk) is **attenuated** (underestimated)
  2) Power to detect relationship is **decreased**
  3) Statistical tests are still **valid**
- Same conclusions seem to hold approximately when **several** dietary exposures are included in a model
- In this webinar we will (mostly) address problem 1)

Slide 5

I'll begin with a summary of the main results from last week. We learned that when there is a **single** dietary exposure measured with error in a diet-health model, three things happen. First, the estimated diet-health relationship (or risk due to exposure), is attenuated, or underestimated. Second, the power to detect the relationship is decreased. And, third, although the power is decreased, the statistical tests used to test for the relationship are still valid.

We also learned that these same conclusions seem to hold approximately when there are **several** dietary exposures included in the diet-health model.

In this webinar we'll focus mostly on the **first** problem and talk about ways to adjust attenuated risk estimates to make them approximately unbiased.

Why adjust the risk estimate?

- Unadjusted estimates will underestimate
  - True health risk due to unhealthy eating
  - True health benefit due to healthy eating
- As a result:
  - Public health impact of dietary change would be underestimated
  - Health officials could mistakenly ignore the potential impact

Slide 6

Why do we need to adjust the risk estimate if the statistical tests are still valid? The reason is that the unadjusted estimates will underestimate the true health risk due to unhealthy eating or true benefit due to healthy eating.

As a result, even if a risk or benefit were found to be statistically significant, the public health impact of dietary change would be understated and health officials could mistakenly ignore the potential impact.

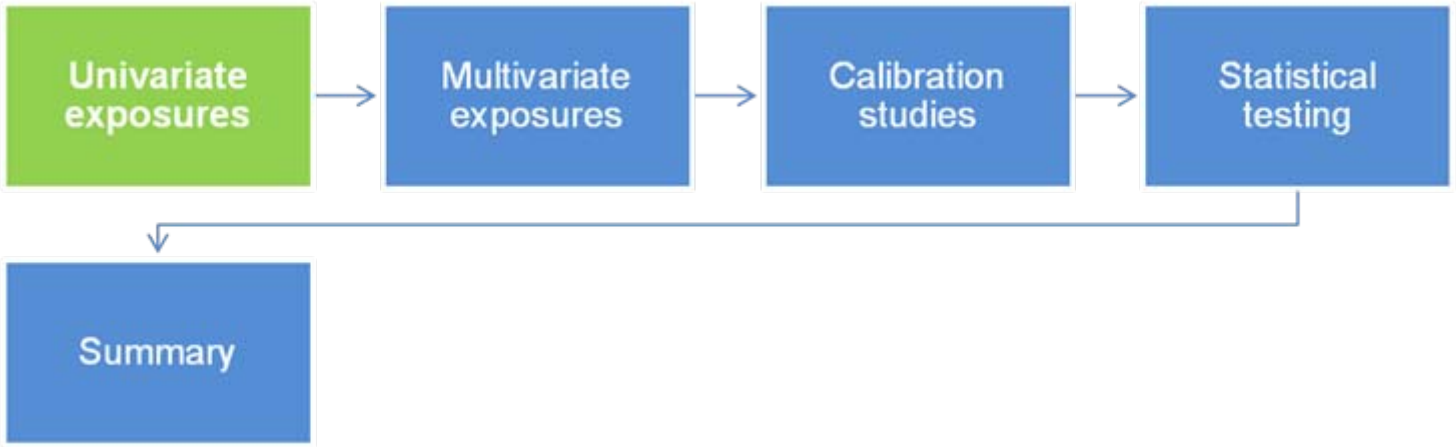
Methods of adjusting the risk estimate

- **Regression calibration**
- SIMEX
- Maximum likelihood
- Multiple imputation
- Moment reconstruction
- … and more!

Slide 7

There are many methods to adjust attenuated risk estimates. They include regression calibration, simulation extrapolation, maximum likelihood, multiple imputation, and moment reconstruction.

We're going to focus on regression calibration, because it is relatively simple to use and performs well in many situations. It's also a method that is commonly used in diet-and-health studies.



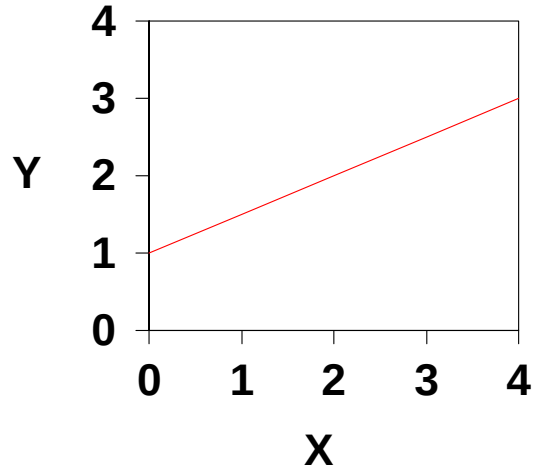
REGRESSION CALIBRATION FOR UNIVARIATE EXPOSURES

Slide 8

We'll begin with the simplest case, which is when there is a **single** dietary variable measured with error.

Linear equations and linear functions

- The equation for a **line** in **two** dimensions is
  - $Y = a_0 + a_1X$



- The equation for a **line** in **k** dimensions is
  - $Y = a_0 + a_1X_1 + a_2X_2 + \ldots + a_kX_k$
- This relationship can be expressed as a function
  - $f(Y) = a_0 + a_1X_1 + a_2X_2 + \ldots + a_kX_k$
- A function of this form is called a **linear function**

Slide 9

But, first, we'll briefly review *linear functions.* You'll remember that the equation for a line in 2-dimensional space is: $Y = a_0 + a_1X$.

Similarly, the equation for a line in (k+1)-dimensional space is: $Y = a_0 + a_1X_1 + a_2X_2$, etc. This linear relationship can be expressed as a function, where we say that $f(Y) = a_0 + a_1X_1$, etc.

A function having this form is called a linear function. We are going to make extensive use of linear functions, so I wanted to remind everyone what they are.

Risk model

- Risk model (logistic regression):

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T T + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

  - Y = health outcome variable (0 or 1)
  - Odds(Y=1) = Prob(Y=1) / Prob(Y=0)
  - T = true dietary intake
  - $Z_1, \ldots, Z_p$ = other variables in disease model
  - $Z = \{Z_1, \ldots, Z_p\}$
  - $\alpha_T, \alpha_{Z1}, \ldots, \alpha_{Zp}$ = regression coefficients = log odds ratios

Slide 10

To assess a diet-health association, we need to specify a diet-health model, which I will call the "risk" model. Risk models relate a health outcome to one or more explanatory variables, sometimes called "covariates." There are different types of risk models for different types of health outcomes. Today we're going to talk about the logistic regression risk model.

In logistic regression, the health outcome is a binary variable, Y, that equals 0 or 1, indicating whether or not some health event has occurred. The "*odds*" for Y is defined as the ratio of the probability that the event has occurred to the probability that it has not occurred. Our logistic regression model assumes that the log of the odds is equal to a linear function of the explanatory variables.

In our example, the explanatory variables are a single dietary exposure called T, and a set of other variables called Z_1, Z_2, etc. For convenience, I'll refer to these other variables collectively as "Z." Z comprises other risk factors that we want to include in our model; for example, age and smoking status. We assume that T is measured with error and that Z is measured exactly.

The parameters α_T, α_{Z1}, etc. are called regression coefficients. They quantify the relationship between the explanatory variable and the health outcome. In logistic regression, the regression coefficients represent log odds ratios. These parameters are unknown, and the goal is to estimate them, using the data from our study.

Risk model

- Risk model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T \cancel{T}^{Q} + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

- Problem:
  - We are unable to measure **true** intake T
  - Instead, we obtain **reported** intake Q which is subject to measurement error
  - If we use Q instead of T in the risk model, the estimate of α_T will be biased (attenuated)

Slide 11

If we could measure true intake, T, in our study, fitting the risk model would be straightforward. The problem is that we can't measure true intake. We can only measure reported intake that is subject to measurement error. In today's talk, we will focus mainly on the case when reported intake is obtained from a food frequency questionnaire, which we call Q.

If we simply use Q instead of T in our risk model, our estimate of the log odds ratio, α_T, will be biased.

The log odds ratios for the other covariates, Z, may also be biased, even though they were measured without error. This results from a phenomenon known as "residual confounding," which was discussed in webinar 6.

This means that even if the main exposure of interest is measured exactly, we still need to be concerned about measurement error in the other variables.

Regression calibration

- Risk model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T T + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

E(T|Q,Z)

- Regression calibration method:
  - Step 1: Calculate E(T|Q, Z) = conditional expectation of T given Q and Z

    E(T|Q, Z) is the "**predicted value**" of T given Q and Z
  - Step 2: Replace T with E(T|Q, Z) in risk model

Slide 12

We don't want biased estimates of our log odds ratios, so we're going to use regression calibration to correct for this bias. Regression calibration is a two-step method.

The first step is to calculate the conditional expectation of true intake, T, given the observed data, Q and Z.

Conditional expectation is a statistical term that may not be familiar to everyone. It can be thought of as a prediction of T based on Q and Z. In fact, the conditional expectation is known to be the *best* predictor of T, in the sense that it has the smallest mean squared error of any predictor that is based on Q and Z.

In this talk, I will use the terms "conditional expectation" and "predicted value" interchangeably.

So, step 1 is to calculate the predicted value of T, and step 2 is to replace T with its predicted value in the risk model and then perform the standard logistic regression analysis.

Regression calibration

- Risk model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T \cancel{T} + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

E(T|Q,Z)

- Regression calibration assumption:
  - Q has "nondifferential error" with respect to disease Y
  - Q has no information about Y beyond that provided by T and Z
- Under this assumption, regression calibration estimates are (approximately) unbiased

Slide 13

Replacing T by its conditional expectation is justified under the assumption that Q has nondifferential error with respect to disease, Y. This means that Q contributes no additional information about disease risk beyond that already provided by T and Z.

This assumption is usually considered reasonable for prospective cohort studies, where the dietary data are collected at the beginning, before any disease has occurred. It is sometimes considered questionable for retrospective case-control studies, where the dietary data are collected after the disease has occurred.

When Q has nondifferential error, regression calibration leads to approximately unbiased estimates of the log odds ratios.

How do we calculate E(T|Q, Z)?

- In order to predict T, we need to develop a "**prediction equation**"
- Example: **linear** prediction equation

$$E(T|Q, Z) = \lambda_0 + \lambda_Q Q + \lambda_{Z1} Z_1 + ... + \lambda_{Zp} Z_p$$

- If T were observable in a sample of participants, could estimate the parameters in prediction equation by regressing T on Q and Z

Slide 14

The obvious question at this point is, “How do we calculate this conditional expectation or predicted value?” Well, we need to develop a “prediction equation.”

The simplest example is a linear prediction equation, where the predicted value is a linear function of Q and Z. If we were able to observe T in sample of participants, we could estimate the parameters in the prediction equation by regressing T on Q and Z.

How do we calculate E(T|Q, Z)?

- Instead of observing T, we observe a "**reference measure**" that we call R
- Assumption: R is **unbiased** for T
  - R = T + e
  - e is random error with mean zero
  - e is uncorrelated with T, Q and Z
- Under this assumption, E(R|Q, Z) = E(T|Q, Z)
- Estimate prediction equation by regressing R on Q and Z in a sample of participants

Slide 15

Since it is not possible to observe T, we need instead a “reference measure,” which we call R. We assume that this reference measure is unbiased for true intake, T, at the individual level. This means that for any individual, R is equal to true intake, T, for that individual plus some random within-person error, where the random error has mean zero and is unrelated to T, Q, and Z.

Under this assumption, the conditional expectation of R given Q and Z is equal to the conditional expectation of T given Q and Z. This means that we can estimate the parameters in the prediction equation by regressing R on Q and Z in a sample of participants.

Calibration studies

- Prediction equation (calibration equation) is developed in a sample on which the reference instrument is measured
- A sample collected for this purpose is called a "**calibration study**"
- We will learn more about calibration studies later in this webinar

Slide 16

The prediction equation, sometimes called the “calibration equation,” is developed in a sample of individuals on which the reference instrument is measured. A sample collected for this purpose is called a “calibration study,” or sometimes a “validation study.”

We will learn more about calibration studies in the third section of this webinar.

Summary of regression calibration

- Regression calibration involves 2 regressions:
  - Step 1: Regress R on Q and Z to get prediction equation E(T|Q, Z)
  - Step 2: Regress health outcome Y on E(T|Q, Z) and Z
- Regression calibration makes 2 assumptions:
  - Q has nondifferential error with respect to Y
  - R is unbiased for T

Slide 17

In summary, regression calibration involves two regressions. In step 1, we regress R on Q and Z to estimate the parameters in the prediction equation. In step 2, we regress the health outcome, Y, on the predicted value of T and Z. In our examples, this second regression is logistic regression.

In addition, regression calibration makes two assumptions. First, it assumes that reported intake Q has nondifferential error with respect to health outcome Y. Second, it assumes that reference measure R is unbiased for true intake T.

Note that this second assumption is a "working assumption" that may or may not be strictly true for any given reference measure. We need to make such an assumption, though, in order to make any progress in correcting for measurement error.

Linear regression calibration

- We will focus on **linear** regression calibration
- In linear regression calibration:
  - Predicted value of T is a **linear** function of Q and Z

$$E(T|Q, Z) = \lambda_0 + \lambda_Q Q + \lambda_{Z1} Z_1 + \ldots + \lambda_{Zp} Z_p$$

  - Parameters in prediction equation (λ_0, …, λ_{Zp}) are estimated by **linear** regression of R on Q and Z

Slide 18

Today, we are focusing on linear regression calibration. In linear regression calibration, the predicted value of T is a linear function of Q and Z. Further, the regression parameters in the prediction equation are estimated by linear regression of R on Q and Z.

The underlying assumption is that the conditional means of both T and R are approximately linear functions of Q and Z. These approximations are often quite adequate for our needs. This is particularly true when the dietary component is consumed nearly every day by nearly everyone, since in this situation we can consider our data to be continuous. There are situations, however, where we may need to consider more complicated nonlinear prediction functions.

Potential problem with linear regression calibration

- Sometimes we want to fit a risk model where T is on a **transformed** scale

  Risk model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T g(T) + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

- Examples of transformation g(T):
  - Log transformation: $g(T) = \log(T)$
  - Square root transformation: $g(T) = \sqrt{T}$

Slide 19

I wanted to mention a potential problem with linear regression calibration. Sometimes researchers want to fit a disease model where T is on a transformed scale. This is particularly true if the data are highly skewed. Here, we show a risk model where T has been transformed using a function, g.

Examples of possible transformations are the log transformation and the square root transformation. The log transformation is the most commonly used transformation and is particularly appropriate if you have reason to believe that the effect of T on Y is multiplicative rather than additive.

Potential problem with linear regression calibration

- Risk model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_T g(T) + \alpha_{Z1} Z_1 + \ldots + \alpha_{Zp} Z_p$$

- Prediction equation is also on transformed scale

$$E\{g(T)|Q, Z\} = \lambda_0 + \lambda_Q g(Q) + \lambda_{Z1} Z_1 + \ldots + \lambda_{Zp} Z_p$$

- Estimate parameters in prediction equation by regressing g(R) on g(Q) and Z

Slide 20

In linear regression calibration, if the risk model is on a transformed scale, then the prediction equation must also be transformed.

In the prediction equation, the predicted value of g(T) is a linear function of g(Q) and Z. We estimate the parameters in the prediction equation by regressing g(R) on g(Q) and Z in the calibration study.

Potential problem with linear regression calibration

- Assumption for reference instruments

$$E(R|Q, Z) = E(T|Q, Z)$$

- After transformation, this equality is only approximate

$$E\{g(R)|Q, Z\} \approx E\{g(T)|Q, Z\}$$

- In practice, approximation is usually assumed good enough for dietary components consumed (nearly) every day

Slide 21

A critical assumption for regression calibration is that the reference instrument and true intake have the same conditional expectation. This allows us to estimate the prediction equation by regressing R on Q and Z. The problem is that after transformation this equality no longer holds. It can be shown that the conditional expectations of g(R) and g(T) are approximately equal, but how good this approximation actually is varies and is subject to many factors.

Nevertheless, in practice, it is usually assumed that this approximation is good enough for dietary components that are consumed every day, and we will proceed under this assumption.

Example: dietary fat and breast cancer

- NIH-AARP Diet and Health Study
- Observational cohort (1995-present)
  - 550,644 participants
  - Food frequency questionnaire (FFQ = Q)
- Calibration sub-study (1996)
  - 1942 participants
  - Two 24-hour dietary recalls (24HR = R)

Slide 22

Now we're going to look at an example using regression calibration in the National Institutes of Health/AARP Diet and Health study, which I'll call the AARP study. The AARP study is a prospective cohort of about 550,000 participants who were administered a food frequency questionnaire at baseline. The study includes a calibration substudy of about 2,000 participants who in addition to the FFQ also completed two nonconsecutive 24-hour dietary recalls, which we are going to use as our reference instrument.

Example: dietary fat and breast cancer

- Thiebaut et al. (J Nat Cancer Inst, 2007)
- Nested case-control analysis
- 3501 invasive breast cancer cases, with 4 matched controls per case:
  - Year of entry (1995, 1996, 1997)
  - Age at entry (+/- 1 year)
  - Person-years at risk (≥ years for case)
  - Hormone use (never/former, current)

Slide 23

In a 2007 article, Anne Thiebaut and her colleagues reported a statistically significant association between dietary fat intake and the risk of invasive breast cancer in women in the AARP study. We will use their analysis as a basis for our examples. We note, however, that we have made many changes and simplifications, and our examples our meant only to illustrate the methods we are discussing.

In the original analysis, the authors used a Cox regression risk model, which is often used to analyze the type of censored data that are obtained from observational studies. To keep our example simple, we are going to use a logistic regression risk model and fit it using a nested case-control analysis. This involves selecting cases and matched controls from our study and then performing logistic regression on that subset.

Our data consisted of 3,501 invasive breast cancers cases, plus 4 matched controls for each case. The controls were matched on a variety of factors, including age and person-years at risk.

Example: dietary fat and breast cancer

- Logistic regression of breast cancer status (Y) on log total fat intake (T)
- Other covariates (Z)
  - Body mass index (< 25, 25-30, ≥ 30)
  - Age at first birth / number of children (nulliparous, <30 / 1-2, <30 / 3+, ≥30 / 1+)
  - Hormone use (never/former, current)
  - Age at entry

Slide 24

We have a logistic regression risk model where the health outcome is breast cancer status and the dietary variable is the log of total fat intake. The other risk factors included in the model are body mass index, age at first birth, number of children, hormone use, and age at entry into the study.

Example: dietary fat and breast cancer

- Regression calibration:
- Step 1: Estimate prediction equation by linear regression of R on Q and Z in calibration study

$$\begin{aligned} E(T|Q, Z) &= 4.5 + 0.28\times Q \\ &\quad + 0.06\times BMI_2 + 0.04\times BMI_3 \\ &\quad + 0.01\times AFB_1 - 0.03\times AFB_2 + 0.05\times AFB_3 \\ &\quad + 0.01\times \text{Hormone} - 0.002\times \text{Age} \end{aligned}$$

- Use prediction equation to predict intake for each subject in main study

Slide 25

We fit the model using regression calibration. Step 1 is to estimate the prediction equation by linear regression of our reference instrument on FFQ-reported log intake of fat and the other covariates in the calibration study. In our case, the reference instrument is the log of the mean reported fat intake on the two 24-hour dietary recalls.

Here is the prediction equation we obtained. We then used this equation to predict intake for all of the subjects in our main study.

Example: dietary fat and breast cancer

- Regression calibration:
- Step 2: Logistic regression of Y on E(T|Q,Z) and Z in main study

$$\log\{\text{Odds}(Y=1)\} = -2.4 + 0.11 \times E(T|Q, Z) + 0.06 \times BMI_2 + 0.17 \times BMI_3 - 0.26 \times AFB_1 - 0.36 \times AFB_2 + 0.01 \times AFB_3 + 0.03 \times \text{Hormone} + 0.004 \times \text{Age}$$

- Estimated log odds ratio: $\hat{\alpha}_T = 0.11$
- Estimated odds ratio: $\exp(\hat{\alpha}_T) = 1.12$

Slide 26

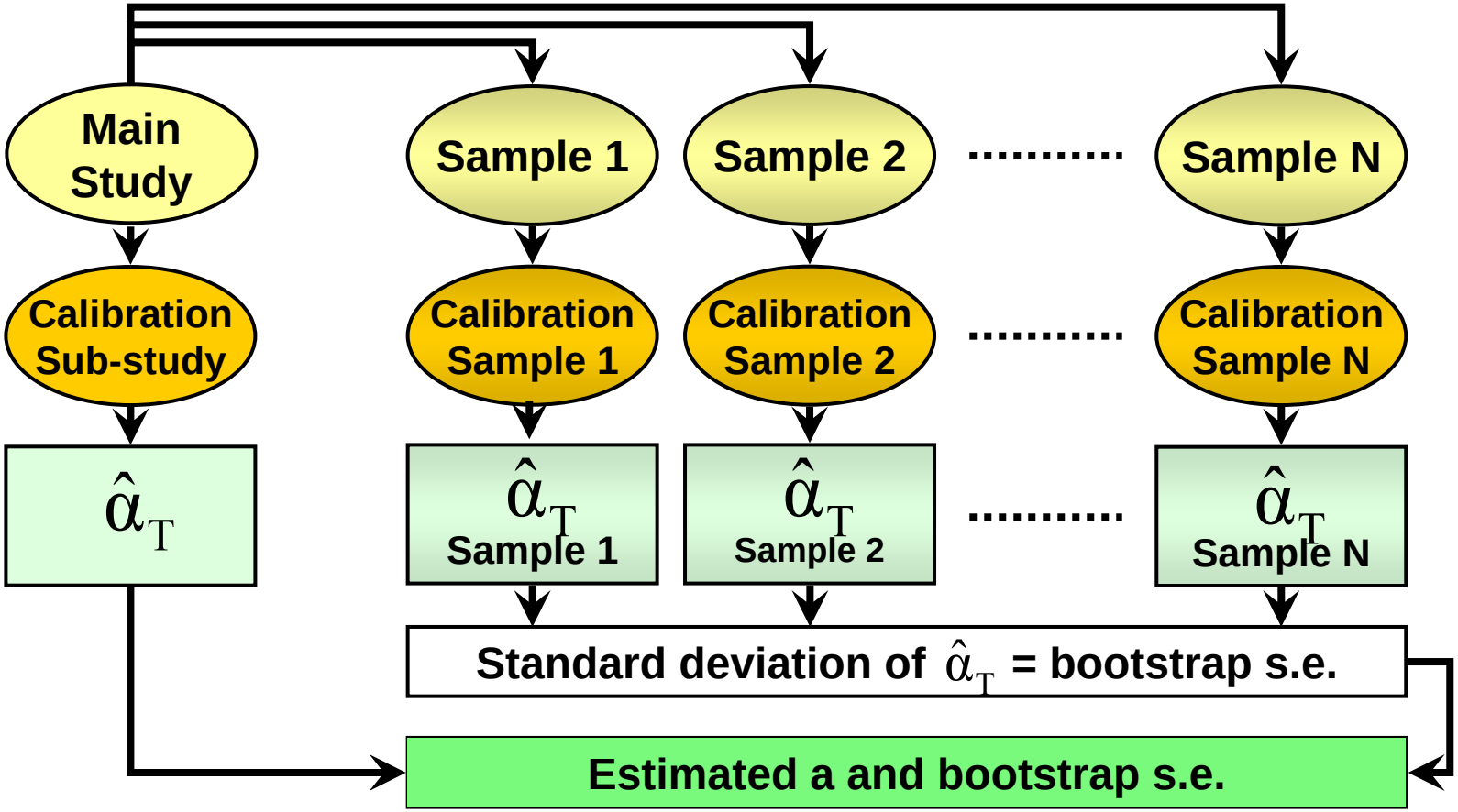
For step 2 of regression calibration, we performed logistic regression of breast cancer status on the predicted value of true log fat intake and other covariates.

Here are the results of that logistic regression. The estimated log odds ratio for fat intake is equal to 0.11. This corresponds to an estimated odds ratio of 1.12. The odds ratio is a measure of the association between fat intake and the risk of breast cancer.

In order to make the odds ratio easier to interpret, we've scaled our estimates so that a unit increase in log fat intake corresponds to a doubling of fat intake on the original scale. Then an odds ratio of 1.12 can be interpreted as a 12 percent increase in the risk of developing breast cancer associated with a doubling of fat intake. This is an example of what is called a "**multiplicative"** effect. Every time intake is multiplied by two, the risk of breast cancer increases by 12 percent.

Bootstrap standard error for log odds ratio

- Bootstrap: sampling with replacement
- Both calibration data and disease model data



Thanks to Anne-Claire Vergnaud

Slide 27

The logistic regression procedure in SAS, or any other statistical software, will print out an estimate of the standard error of the log odds ratio, and a 95 percent confidence interval for the odds ratio. Unfortunately, these standard errors and confidence intervals will not be correct for regression calibration estimates. This is because regression calibration is a two-step procedure, and the logistic regression standard errors in step 2 don't incorporate the uncertainty due to estimating the prediction equation parameters in step 1.

The bootstrap method is a very general and commonly used method for estimating standard errors that can be applied in such two-step procedures. In webinar 4, Kevin Dodd gave a general description of the bootstrap method, and here we're going to describe how it applies to regression calibration.

As before, we sample with replacement from our original data, but this time we need to resample from *both* the main study and the calibration study.

Here we see our analysis of the original data. We used the main study and calibration substudy to estimate the log odds ratio.

To calculate the standard error, we sample with replacement from both the main study and calibration study. These "bootstrapped" samples are meant to mimic data from a new study. We use these new data to calculate a "bootstrapped" estimate of the log odds ratio. We repeat this many times and get many different bootstrapped estimates.

Finally, we calculate the standard deviation of these bootstrapped estimates to obtain the estimated standard error of our log odds ratio.

Note that the estimate of the log odds ratio comes from the original study. We are only using the bootstrapped samples to estimate its standard error.

Confidence intervals

- Estimated log odds ratio: $\hat{\alpha}_T = 0.11$
- Bootstrap s.e. log odds ratio: $\text{s.e.}\left(\hat{\alpha}_T\right) = 0.10$
- Estimated odds ratio: $\exp\left(\hat{\alpha}_T\right) = 1.12$
- 95% confidence interval for log odds ratio:

$$\hat{\alpha}_T \pm 1.96 \times \text{s.e.}\left(\hat{\alpha}_T\right) = \left(-0.09, 0.31\right)$$

- 95% confidence interval for odds ratio:

$$\left(\exp\{-0.09\}, \exp\{0.31\}\right) = \left(0.91, 1.36\right)$$

Slide 28

Once we've calculated the standard error of the log odds ratios, we can calculate confidence intervals.

Since the estimated log odds ratio has an approximately normal distribution, we can calculate an approximate 95 percent confidence interval by adding and subtracting from the estimate a value equal to 1.96 times its standard error.

In order to get 95 percent confidence intervals for the odds ratio, we take the endpoints of the confidence interval for the log odds ratio, and exponentiate them. As you can see, the confidence interval for the odds ratio includes the value 1, which represents no association. This means that we can't reject the null hypothesis that there is no association between fat intake and breast cancer risk in this particular example.

Rosner's method for linear regression calibration

Rosner et al. (Am J Epidemiol, 1990)

From previous lecture (lecture 6):

- Expected log odds ratio estimate is attenuated

$$E\left(\hat{\alpha}_Q\right) = \lambda_Q \alpha_T$$

- α_T = true log odds ratio
  λ_Q = attenuation factor (from prediction equation)

- Solution: divide attenuated log odds ratio by λ_Q

$$\hat{\alpha}_T = \hat{\alpha}_Q / \lambda_Q$$

Slide 29

Now I'd like to look briefly at an alternative method for applying linear regression calibration that was proposed by Bernard Rosner and his colleagues back in 1990. In webinar 6, we saw that when reported intake, Q, is measured with error, the expected value of the estimated log odds ratio is biased by a multiplicative factor called the "attenuation factor." This attenuation factor is equal to the regression coefficient for Q in the prediction equation. So a simple way to "de-attenuate" the attenuated log odds ratio is to divide it by the attenuation factor.

Rosner's method for linear regression calibration

- Step 1: Same as step 1 for regular method
- Prediction equation:

$$E(T|Q, Z) = 4.5 + 0.28 \times Q + 0.06 \times BMI_2 + 0.04 \times BMI_3 + 0.01 \times AFB_1 - 0.03 \times AFB_2 + 0.05 \times AFB_3 + 0.01 \times \text{Hormone} - 0.002 \times \text{Age}$$

- Estimated attenuation factor: $\hat{\lambda}_Q = 0.28$
- Standard error: $\text{s.e.}(\hat{\lambda}_Q) = 0.03$

Slide 30

Rosner's method can also be described as a two-step method. The first step is the same as before—estimate the parameters in the prediction equation using linear regression calibration of the reference measure R on Q and Z.

From this linear regression, we get an estimate of the attenuation factor and also the standard error of this estimate, which will be useful later.

Rosner's method for linear regression calibration

- Step 2: Logistic regression of Y on Q and Z in the main study

$$\log\{\text{Odds}(Y=1)\} = -1.7 + 0.03 \times Q + 0.08 \times BMI_2 + 0.18 \times BMI_3 - 0.26 \times AFB_1 - 0.37 \times AFB_2 + 0.02 \times AFB_3 + 0.03 \times \text{Hormone} + 0.003 \times \text{Age}$$

- Attenuated log odds ratio: $\hat{\alpha}_Q = 0.03$
- Standard error: $\text{s.e.}(\hat{\alpha}_Q) = 0.03$

Slide 31

Step 2 is to fit a logistic regression of health outcome Y on Q and Z in the main study to estimate the attenuated log odds ratio.

Here are the results of that logistic regression. The attenuated log odds ratio is 0.03, with a standard error of 0.03. Again, this standard error will be useful later.

Rosner's method for linear regression calibration

Step 3: Divide attenuated log odds ratio by attenuation coefficient

$$\hat{\alpha}_T = \hat{\alpha}_Q / \hat{\lambda}_Q = 0.03 / 0.28 = 0.11$$

- Rosner's method and regular linear regression calibration: estimates are the same

Slide 32

Finally, we divide the attenuated log odds ratio by the attenuation factor. The resulting estimate is often called the “de-attenuated” log odds ratio. In our example, the de-attenuated log odds ratio equals 0.11.

Notice that this estimate is the same as in our earlier example. This is true in general. Rosner’s method and regular linear regression calibration give exactly the same estimate of the log odds ratio.

Rosner's method for linear regression calibration

- Step 4: Estimate standard error using delta method

$$\text{s.e.}(\hat{\alpha}_T) \approx \sqrt{\left(\frac{\text{s.e.}(\hat{\alpha}_Q)}{\hat{\lambda}_Q}\right)^2 + \left(\frac{\hat{\alpha}_Q \text{s.e.}(\hat{\lambda}_Q)}{\hat{\lambda}_Q^2}\right)^2}$$

$$= 0.09$$

- Bootstrap and delta method standard errors are similar but not exactly the same

Slide 33

Rosner's method uses the delta method to estimate the standard error of the log odds ratio. Like the bootstrap method, the delta method can be used to estimate standard errors in two-step procedures. I won't describe the delta method in general, but in this application it leads to a fairly simple formula based on the estimated values of the attenuated log odds ratio, the attenuation factor, and their standard errors.

I want to note that the bootstrap and delta method standard errors are generally similar, but they are not exactly the same.

Example: dietary fat and breast cancer

Dietary fat intake and breast cancer risk in NIH-AARP

Correction for Measurement Error	Log Odds Ratio (s.e.)	Odds Ratio (95% CI)
Uncorrected	0.03 (0.03)	1.03 (0.98, 1.09)
Regression calibration	0.11 (0.10)	1.12 (0.93, 1.35)
Rosner's method	0.11 (0.09)	1.12 (0.94, 1.33)

Slide 34

Here is a table summarizing our analysis of dietary fat and breast cancer risk. We see that the uncorrected estimate of the odds ratio is much smaller than the regression calibration estimates. We also see that the versions of regression calibration give the same estimate of the odds ratio but give slightly different estimates of the 95 percent confidence intervals. Notice also that none of estimated odds ratios are statistically significantly different from 1, since all the confidence intervals include the value 1.

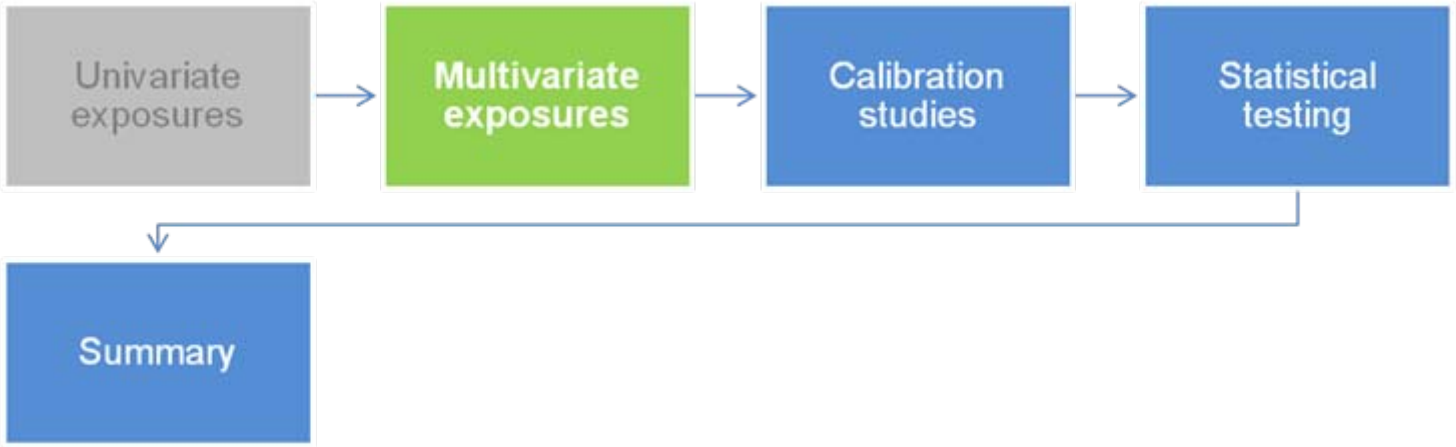
Summary of Rosner's method

- Advantage: standard errors can be computed quickly and easily
- Limitation: Only applies when the regression calibration model is **linear**
- Next webinar (webinar 8) will focus on situation where regression calibration model is **nonlinear**

Slide 35

The main advantage of Rosner's method is that it allows standard errors to be computed quickly and easily. A limitation is that the method only applies when the regression calibration model is linear. The regular method can be extended to nonlinear models, and in next week's webinar we'll see an example where the regression calibration model is nonlinear.

Now, I want to mention that in an earlier version of these slides that was available online, I included a second limitation to Rosner's method, which stated that the method only corrects the odds ratio for the variable measured with error and that the odds ratios for Z are still biased. I realized, however, that this statement was incorrect. There is a slightly more complex version of Rosner's method than the one I described here that does produce unbiased estimates of the odds ratios for Z.



REGRESSION CALIBRATION FOR MULTIVARIATE EXPOSURES

Slide 36

Now I'm going to talk about regression calibration when there are multiple variables measured with error. I'll focus on the case of two variables measured with error, but the methods extend easily to more than two variables.

Motivation: energy-adjusted analysis

- Researchers often perform an “energy-adjusted” analysis by adding total energy intake to model
- Sometimes, the main exposure variable is also modified
  - Example: Percent energy from fat
- Reasons for energy adjustment:
  - Interest in association between dietary composition and health
  - Energy-adjustment often decreases measurement error in reported intake

Slide 37

I'll begin with a motivating example that was also discussed in webinar 6. In nutritional epidemiology, researchers often perform what are called "energy-adjusted" analyses, in which total energy intake is included as a covariate in the risk model. Sometimes the main exposure variable is modified as well. An example of such a modified variable is percent energy from fat, as opposed to absolute fat intake.

There are two main reasons for doing an energy-adjusted analysis. The first is that the researcher may be interested in the association between dietary composition and health. For example, the researcher may be interested in the proportion of fat in the diet rather than the absolute amount. The second reason, which was discussed in webinar 6, is that energy adjustment often decreases the measurement error in reported intake.

Regression calibration with two dietary exposures

- Disease model:

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_{T1}T_1 + \alpha_{T2}T_2 + \alpha_{Z1}Z_1 + \ \ldots \ + \alpha_{Zp}Z_p$$

- Replace T_1 and T_2 with $E(T_1|Q_1,Q_2,\underline{Z})$ and $E(T_2|Q_1,Q_2,\underline{Z})$
- $E(T_1|Q_1,Q_2,\underline{Z})$ is the **predicted value** of T_1 given reported intakes Q_1 and Q_2 and explanatory variables $Z_1, \ldots, Z_p$
- Confidence intervals for odds ratios calculated using the bootstrap method, exactly as described for a single exposure

Slide 38

Here is a risk model that has two dietary exposures, T_1 and T_2, plus other covariates, Z. Regression calibration with two dietary exposures is very similar to regression calibration with a single dietary exposure.

First, calculate the conditional expectations of T_1 and T_2 given Q_1, Q_2, and Z. Then replace T_1 and T_2 in the risk model with their predicted values and perform the standard logistic regression analysis.

Confidence intervals for the estimated odds ratios are calculated using the bootstrap method, exactly as described for a single dietary exposure.

Linear regression calibration

- Prediction equations:

$$E(T_1|Q_1,Q_2,Z) = \lambda_{01} + \lambda_{Q11}Q_1 + \lambda_{Q21}Q_2 + \lambda_{Z11}Z_1 + \ldots + \lambda_{Zp1}Z_p$$

$$E(T_2|Q_1,Q_2,Z) = \lambda_{02} + \lambda_{Q12}Q_1 + \lambda_{Q22}Q_2 + \lambda_{Z12}Z_1 + \ldots + \lambda_{Zp2}Z_p$$

- Reference measures R_1 and R_2 for T_1 and T_2
- Linear regression of R_1 and R_2 on Q_1, Q_2 and Z in calibration study

Slide 39

As before, prediction equations must be developed to calculate the predicted values. Notice that both Q_1 and Q_2 are in the prediction equation for T_1 and that both are in the prediction equation for T_2.

Since we can't observe T_1 or T_2, we need two reference measures, called R_1 and R_2, that we assume are unbiased for the true exposures. We then estimate the prediction equations using linear regression of R_1 and R_2 on Q_1, Q_2 and Z in the calibration study.

Rosner's method for linear regression calibration

$$\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_{T1}T_1 + \alpha_{T2}T_2 + \alpha_{Z1}Z_1 + \ \ldots\ + \alpha_{Zp}Z_p$$

$$E(T_1|Q_1,Q_2,\underline{Z}) = \lambda_{01} + \lambda_{Q11}Q_1 + \lambda_{Q21}Q_2 + \lambda_{Z11}Z_1 + \ldots + \lambda_{Zp1}Z_p$$

$$E(T_2|Q_1,Q_2,\underline{Z}) = \lambda_{02} + \lambda_{Q12}Q_1 + \lambda_{Q22}Q_2 + \lambda_{Z12}Z_1 + \ldots + \lambda_{Zp2}Z_p$$

- From lecture 6, log odds ratios estimated using Q_1 and Q_2 actually estimate:

$$\alpha_{Q1} = \lambda_{Q11} \times \alpha_{T1} + \lambda_{Q12} \times \alpha_{T2}$$

$$\alpha_{Q2} = \lambda_{Q22} \times \alpha_{T2} + \lambda_{Q21} \times \alpha_{T1}$$

contamination

attenuation

Slide 40

Rosner's method also extends to the case where there are multiple variables measured with error. At first, the extension might seem complicated, but I'll try to convince you that it's really pretty simple.

We start with our risk model and two prediction equations. We saw in last week's webinar that if we don't correct for measurement error, the log odds ratios estimated using Q_1 and Q_2 actually estimate a mixture of the true log odds ratios. We have attenuation of the true effect of the dietary exposure, combined with contamination from the effect of the other dietary exposure. The attenuation factors and contamination factors are equal to regression coefficients in the prediction equations.

Rosner's method for linear regression calibration

$$\alpha_{Q1} = \lambda_{Q11} \times \alpha_{T1} + \lambda_{Q12} \times \alpha_{T2}$$

$$\alpha_{Q2} = \lambda_{Q22} \times \alpha_{T2} + \lambda_{Q21} \times \alpha_{T1}$$

- Equations can be written in matrix notation

$$\alpha_Q = \Lambda_Q \alpha_T$$

where:

$$\Lambda_Q = \begin{pmatrix} \lambda_{Q11} & \lambda_{Q12} \\ \lambda_{Q21} & \lambda_{Q22} \end{pmatrix}, \quad \alpha_T = \begin{pmatrix} \alpha_{T1} \\ \alpha_{T2} \end{pmatrix}, \quad \alpha_Q = \begin{pmatrix} \alpha_{Q1} \\ \alpha_{Q2} \end{pmatrix}$$

- Λ_Q is called the "attenuation-contamination" matrix

Slide 41

What we have is a system of linear equations that can be solved using linear algebra. If we write the equations in matrix notation, then α_Q equals Λ_Q times α_T, just like in the univariate case, except now α_Q and α_T are vectors and Λ_Q is a matrix.

Λ_Q is called the “attenuation-contamination” matrix because it consists of the attenuation and contamination factors in our equations.

Rosner's method for linear regression calibration

Univariate case:

- Bias: $\alpha_Q = \lambda_Q \alpha_T$
- Estimate: $\alpha_T = \alpha_Q / \lambda_Q = \lambda_Q^{-1} \alpha_Q$

Bivariate case:

- Bias: $\alpha_Q = \Lambda_Q \alpha_T$
- Estimate: $\alpha_T = \Lambda_Q^{-1} \alpha_Q$
- Standard errors for α_{T1} and α_{T2} can be estimated by multivariate delta method

Slide 42

In the case of a *single* variable measured with error, the attenuated log odds ratio is divided by the attenuation factor to obtain our estimate. Equivalently, we can say we are multiplying the inverse of the attenuation factor by the attenuated log odds ratio.

When we have two or more variables measured with error, we do exactly the same thing, except this time we multiply the inverse of the attenuation-contamination matrix by the vector of attenuated log odds ratios.

Standard errors for our estimates can be obtained using a multivariate version of the delta method.

If this still seems complicated, I do want to mention that there are programs available that will perform all the necessary calculations, so you don't have to do them yourself.

Example: energy-adjusted fat and breast cancer

- NIH-AARP Diet and Health Study
- Logistic regression of breast cancer status (Y) on log total fat intake (T_1) and log non-alcohol energy intake (T_2)
- Substitution effect: adding fat intake while keeping non-alcohol energy constant
- Other covariates (Z) same as for univariate fat and breast cancer example

Slide 43

Let's continue our example analysis of dietary fat and breast cancer.

We are going to perform logistic regression of breast cancer status on log total fat intake and log nonalcohol energy intake. By nonalcohol energy intake, we mean caloric intake from all sources except alcohol.

When we include nonalcohol energy intake in the model, we are estimating what is called a "*substitution effect*" for fat intake. This means we are estimating the effect of substituting fat for other nonalcohol sources of energy in the diet.

Example: energy-adjusted fat and breast cancer

Dietary fat intake and breast cancer risk in NIH-AARP

Correction for Measurement Error	Log Odds Ratio (s.e.)	Odds Ratio (95% CI)
Uncorrected	0.15 (0.05)	1.16 (1.05, 1.29)
Regression calibration	0.29 (0.11)	1.34 (1.09, 1.66)
Alternate method	0.29 (0.10)	1.34 (1.10, 1.64)

- Adjusted for non-alcohol energy intake

Slide 44

Here we see the results of the energy-adjusted analysis. Again, we see that the uncorrected estimate of the odds ratio is much smaller than the regression calibration estimates. The two regression calibration estimates are again equal, but the 95 percent confidence intervals are slightly different.

Notice that none of the 95 percent confidence intervals include the value 1, which means that the odds ratios are statistically significantly different from 1 at the 5 percent significance level. We can therefore reject the null hypothesis of no association between fat intake and breast cancer risk.

Why are we able to reject the null hypothesis in the energy-adjusted analysis, but are unable to do so in the univariate analysis? There are two possible reasons. The first is that we are estimating a different fat-breast cancer relationship. In the energy-adjusted analysis, we are estimating the effect of *substituting* fat for other nonalcohol sources of energy, while in the univariate analysis we are estimating the effect of adding fat without regard to any other change in the diet. The second reason, which I mentioned earlier, is that energy adjustment often reduces the measurement error in reported intake. When the measurement error is reduced, the power to detect an association is increased.

CALIBRATION STUDIES

Slide 45

Now we're going to look at calibration studies in more detail. Calibration studies are a crucial part of regression calibration or any other method to adjust for measurement error.

Calibration studies

- Studies performed to “calibrate” the main study instrument (Q) to a reference instrument (R)
- To calibrate Q means to develop a prediction equation to predict R given Q
- The information from these studies can be used as the basis for regression calibration

Slide 46

Calibration studies are studies that are performed to “calibrate” the main study instrument. To calibrate an instrument means to develop a prediction equation to predict true intake given reported intake. In order to do this, we must observe either true intake or an unbiased reference instrument in the calibration study. The prediction equation can then be used in regression calibration.

Types of calibration studies

- Internal calibration study: random subsample of main study participants
- External calibration study: separate study of participants that are similar to those participating in the main study
- Participants must complete the same study instrument (Q) that is used in the main study
- Internal calibration is preferable

Slide 47

There are two types of calibration studies. An internal calibration study is a random subsample of the main study participants, while an external calibration study is a separate study of participants who are similar to those participating in the main study. In either type of study, participants must complete the same dietary instrument that is being used in the main study.

In general, internal calibration is preferable, because it assures that the participants in the calibration study are similar to those in the main study and that all covariates, including the main dietary instrument, are administered or measured in the same way.

Reference instrument: ideal properties

- Unbiased measure of individual true usual intake
- Errors uncorrelated with true usual intake
- Errors uncorrelated with errors in study instrument

Slide 48

Ideally, the reference instrument in the calibration study should have three properties: First, it should provide an unbiased estimate of true intake at the individual level. Second, error in the reference instrument should be uncorrelated with true intake. Third, error should be uncorrelated with errors in the main study instrument.

Reference instrument: examples

- Doubly labeled water for energy intake
- 24-hour urinary nitrogen for protein intake
- 24-hour urinary potassium for potassium intake

Slide 49

There are a few *biomarkers* of dietary intake that have been shown in feeding studies to have these ideal properties, at least approximately. They include:

- Doubly-labeled water for energy intake
- 24-hour urinary nitrogen for protein intake
- 24-hour urinary potassium for potassium intake.

These three reference biomarkers can be used to develop prediction equations for energy, protein, and potassium. Unfortunately, for the vast majority of dietary components that are of interest to researchers there are no known reference instruments that have these ideal properties.

Reference instrument

- Instruments that are usually used as a reference
  - 24-hour recalls (one or more)
  - Multiple-day food records
- Problems
  - Biased for true intake
  - Errors correlated with true intake
  - Errors correlated with errors in FFQ

Slide 50

In practice, the instruments that are usually used as references in calibration studies are 24-hour dietary recalls and multiple-day food records. These instruments provide estimates of intake for one or more days, and require only that the participant remembers what he/she ate on the previous day or writes down what he/she eats each day.

There is evidence that these instruments are less biased than food frequency questionnaires, but they are not ideal. In studies with reference biomarkers, they have been shown to be biased for true intake and to have errors that are correlated with true intake and with the errors in an FFQ. So the question becomes: Is it better to adjust for measurement error using a reference instrument that is known to be biased or to not correct for measurement error at all?

Performance of 24HR as reference

OPEN Study: Attenuation factors estimated using recovery biomarker or 24HR as reference (Freedman et al., J Nat Cancer Inst, 2011)

Nutrient	Gender	Reference Biomarker	Reference 24HR
Energy	Men	0.08 (0.03)	0.21 (0.04)
	Women	0.03 (0.03)	0.09 (0.05)
Protein Density	Men	0.43 (0.07)	0.35 (0.07)
	Women	0.33 (0.08)	0.45 (0.06)
Potassium Density	Men	0.57 (0.08)	0.59 (0.05)
	Women	0.61 (0.08)	0.62 (0.07)

Slide 51

This question was addressed in a recent commentary by Larry Freedman and colleagues. They based their conclusions on an analysis of the OPEN biomarker study. OPEN is a study of about 500 participants who provided information on dietary intake as measured by a food frequency questionnaire, a 24-hour dietary recall, and reference biomarkers for energy, protein and potassium.

This table, taken from their commentary, shows attenuation factors for energy, protein density, and potassium density, estimated using either a reference biomarker or the 24HR as a reference instrument.

Protein and potassium density are examples of *energy-adjusted* dietary components. Protein density equals protein intake divided by total energy intake, and potassium density is similarly defined.

For energy intake, we see that using 24HR as a reference results in an overestimate of the attenuation factor compared to using a reference biomarker.

If the attenuation factor is overestimated, using it in regression calibration would not completely adjust for the true attenuation in an observed risk estimate. As a result, the regression calibration estimate would still be attenuated, though not as severely as the uncorrected risk estimate.

For protein and potassium density, the attenuation factors estimated using 24HR are not that different from those estimated using a reference biomarker. This is especially true for potassium density, but even for protein density we see no clear pattern of over- or underestimation by the 24HR.

Performance of 24HR as reference

- Attenuation factors appear to be similar for energy-adjusted nutrients
- Freedman et al. (2011) concluded that regression calibration with 24HR improves estimation (on average) compared to no adjustment
- Caveat: conclusion based on only three nutrients: protein, potassium and energy

Slide 52

To summarize, attenuation factors estimated using a reference biomarker or 24HR appear to be similar for energy-adjusted dietary components. In their commentary, Freedman and colleagues concluded that regression calibration with a 24HR improves risk estimation, on average, compared with no adjustment. They included, however, the caveat that their conclusion was based on only three dietary components: protein, potassium, and energy.

Design of calibration studies

- Two important factors:
  - Number of participants in calibration study
  - Number of reference measures per participant

Slide 53

I want to spend a little time discussing two important factors in the design of calibration studies. The first is the number of participants to include in the study, and the second is the number of reference measures to collect from each participant.

Size of calibration study

$$s.e.(\hat{\alpha}_T) \approx \frac{\alpha_Q}{\lambda_Q}\sqrt{\left(\frac{s.e.(\hat{\alpha}_Q)}{\alpha_Q}\right)^2 + \left(\frac{s.e.(\hat{\lambda}_Q)}{\lambda_Q}\right)^2}$$

- 1st term is uncertainty of estimating attenuated log odds ratio in main study of size N
- 2nd term is uncertainty of estimating attenuation factor in calibration study of size n

$$s.e.(\hat{\lambda}_Q) = \sqrt{\frac{\sigma_\varepsilon^2}{n\,\sigma_Q^2}}$$

- σ_ε^2 = residual variance in regression of R on Q

Slide 54

We'll first consider the number of participants in the calibration study. We'll assume that the purpose of the study is to estimate prediction equations for regression calibration. We want the calibration study to be large enough so that we can get precise estimates of the log hazard ratios in a diet-health model.

Here is the formula for the standard error of the de-attenuated log odds ratio. As I mentioned before, the standard error reflects the uncertainty of both steps of the regression calibration method.

The first term inside the square root sign reflects the uncertainty of estimating the attenuated log odds ratio in a main study of size, capital N, while the second term reflects the uncertainty of estimating the attenuation factor in the calibration study of size, small n. It is this second term that is affected by the size of the calibration study. We can see this explicitly by writing the standard error of the attenuation factor as a function of n, the variance of Q, and the residual variance in the regression of R on Q.

Size of calibration study

$$s.e.(\hat{\alpha}_T) \approx \frac{\alpha_Q}{\lambda_Q}\sqrt{\left(\frac{s.e.(\hat{\alpha}_Q)}{\alpha_Q}\right)^2 + \left(\frac{\sigma_\varepsilon^2}{n\lambda_Q^2\sigma_Q^2}\right)}$$

- Choose n so that 2nd term is a small fraction of 1st term, say 1/10th

$$\frac{\sigma_\varepsilon^2}{n\lambda_Q^2\sigma_Q^2} = \frac{1}{10}\left(\frac{s.e.(\hat{\alpha}_Q)}{\alpha_Q}\right)^2$$

- Solve for n

$$n = 10\left(\frac{\sigma_\varepsilon^2}{\lambda_Q^2\sigma_Q^2}\right)\left(\frac{\alpha_Q}{s.e.(\hat{\alpha}_Q)}\right)^2$$

Slide 55

Here again is the standard error of the de-attenuated log odds ratio, this time written as a function of n. We want to choose n so that the second term inside the square root sign contributes only a small proportion of the total standard error. We do this by choosing n so that the second term is a small fraction of the first term, say one tenth. Then, we solve for n to get a formula for choosing the sample size of the calibration study.

Example: size of calibration study

$$n = 10\left(\frac{\sigma_\varepsilon^2}{\lambda_Q^2\sigma_Q^2}\right)\left(\frac{\alpha_Q}{s.e.(\hat{\alpha}_Q)}\right)^2$$

- 95% CI: $\hat{\alpha}_Q \pm 1.96 \times s.e.\left(\hat{\alpha}_Q\right)$
- Choose $s.e.(\hat{\alpha}_Q) = \alpha_Q / 2$

$$n = 40\left(\frac{\sigma_\varepsilon^2}{\lambda_Q^2\sigma_Q^2}\right)$$

For fat intake in AARP

$$\lambda_Q = 0.28, \quad \sigma_Q^2 = 0.25, \quad \sigma_\varepsilon^2 = 0.34$$

- n = 694

Slide 56

Note that the formula for n depends on the ratio of the attenuated log odds ratio to its standard error. This ratio determines whether or not, on average, the 95 percent confidence interval for the attenuated log odds ratio includes 0, which in turn determines whether or not the estimated association is statistically significant.

As we saw earlier, the confidence interval for the attenuated log odds ratio equals the estimated log odds ratio plus or minus 1.96 times its standard error. So if the standard error is equal to half the size of the attenuated log odds, then the 95 percent confidence interval will, on average, not include 0.

This is the point at which we want to control the standard error of the de-attenuated log odds ratio, so that attenuated associations that are statistically significant do not become nonsignificant after de-attenuation.

Setting the standard error of α_Q equal to one half α_Q, we get a simplified formula for n. We can estimate the parameters in this formula using fat intake in the AARP study. Plugging them into the formula for n, we get a required sample size equal to 694. Since the AARP calibration study has 1,000 women, we conclude that it is large enough for estimating the association between fat intake and breast cancer risk.

Number of reference measurements

- Calibration study of sample size n, with k administrations of R per participant
- $\bar{R}_k$ = average of k repeats of R for participant
- Estimate λ_Q by regressing $\bar{R}_k$ on Q

$$s.e.(\hat{\lambda}_Q) = \sqrt{\frac{\sigma^2_{\varepsilon k}}{n\sigma^2_Q}}$$

- $\sigma^2_{\varepsilon k}$ = residual variance in regression of $\bar{R}_k$ on Q

Slide 57

Now we are going to consider how the number of reference measures per person affects the sample size of the calibration study.

Consider a calibration study with n participants and k reference measures per person. Let $\bar{R}_K$ be the average of the k reference measures for each participant. We estimate the attenuation factor by regressing $\bar{R}_K$ on Q. As before, the standard error of the attenuation factor is a function of n, the variance of Q, and the residual variance in the regression of $\bar{R}_K$ on Q.

Number of reference measurements

$$\sigma^2_{\varepsilon k} = \sigma_{RR} + \left(\sigma^2_R - \sigma_{RR}\right) / k - \lambda^2_Q \sigma^2_Q$$

- σ^2_R is the variance of a single R and σ_{RR} is the covariance between repeat R's

$$s.e.(\hat{\lambda}_Q) = \sqrt{\frac{\sigma^2_{\varepsilon k}}{n\sigma^2_Q}} = \sqrt{\frac{\sigma_{RR} + \left(\sigma^2_R - \sigma_{RR}\right) / k - \lambda^2_Q \sigma^2_Q}{n\sigma^2_Q}}$$

- Gain in precision due to additional reference measurements depends on the size of $\left(\sigma^2_M - \sigma_{MM}\right)$

Slide 58

The residual variance in the regression of $\bar{R}_K$ on Q can be expressed as a function of k. In this expression, σ_R^2 is the variance of a single 24HR, and σ_{RR} is the covariance between repeat 24HR.

Plugging this expression into the formula for the standard error of the attenuation factor, we get this formula.

We can see that the gain in precision due to additional reference measures depends on the difference between σ_R^2 and σ_{RR} .

Example: number of reference measurements

$$s.e.(\hat{\lambda}_Q) = \sqrt{\frac{\sigma^2_{\varepsilon k}}{n\sigma^2_Q}} = \sqrt{\frac{\sigma_{RR} + \left(\sigma^2_R - \sigma_{RR}\right)/k - \lambda^2_Q\sigma^2_Q}{n\sigma^2_Q}}$$

- For fat intake in AARP

$$\lambda_Q = 0.28,\ \ \sigma^2_Q = 0.25,\ \ \sigma^2_R = 0.350,\ \ \sigma_{RR} = 0.133$$

$$s.e.(\hat{\lambda}_Q) = \sqrt{\frac{0.454 + 0.868/k}{n}}$$

- s.e. is a function of n and k

Slide 59

Here is the formula for the standard error of the attenuation factor as a function of the various parameters. Again, we can estimate these parameters for fat intake in the AARP study. Plugging them into our formula, we get a simple formula that is a function of n and k. We can use this formula to determine the relative sample size required, depending on the number of repeat measurements.

Example: number of reference measurements

- Relative sample size required in a calibration study with k repeats of M per participant

k	n
1	1000
2	672
3	563
4	508

(Based on fat intake in AARP)

Slide 60

Here is a table showing the relative size required in a calibration study with k reference measures per person, based on fat intake in the AARP study.

If a calibration study with one reference measure per person required 1,000 participants to reach a desired precision, then a study with two measures per person would require 672 participants to reach the same precision, while a study with four measures per person would require about 500 participants.

Note that the sample size, n, is not cut in half when the number of reference measures per person is doubled. This means that the total number of reference measures required in the study increases when the number of reference measures per person increases.

So, if the cost of a calibration study is driven by the cost of administering the reference measure, then it would be most cost-efficient to collect only one reference measure per person. If the cost of administering the reference measure is only one of many costs, then it may be more cost-efficient to collect more reference measures per person.

Minimum number of reference instruments

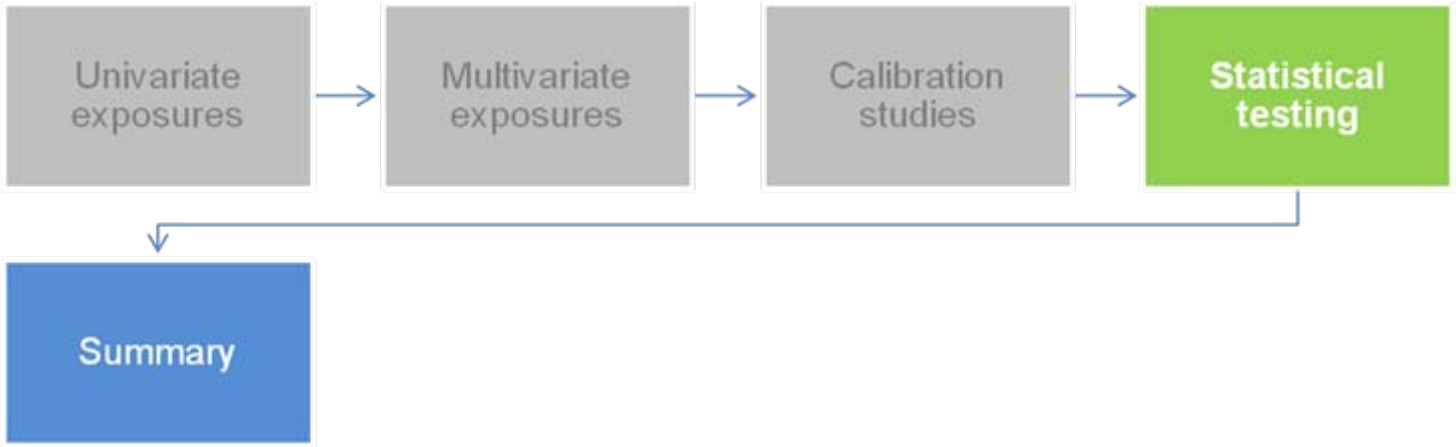
- Performing regression calibration when regression calibration model is linear
  - Minimum number = 1
- Performing regression calibration when regression calibration model is nonlinear
  - Minimum number = 2
- Estimating correlation of Q and T
  - Minimum number = 2
  - Estimating power
  - Assessing the quality of Q (validity)

Slide 61

When deciding how many reference measures to administer to each participant in a calibration study, you also have to keep in mind the minimum number required for different purposes.

For linear regression calibration when an FFQ is the main instrument, the minimum number required is one. For nonlinear regression calibration when an FFQ is the main instrument, the minimum number required is one or two. The slide says two, but it really depends on the type of nonlinearity in the model. For regression calibration when the 24HR is the main instrument, the minimum number required is two. And to estimate the correlation between true and FFQ-reported intake, the minimum number is two. The correlation is needed to estimate the power to test diet-health associations and to assess the quality or "validity" of the FFQ.

Given these considerations, we think that, in general, it is best to collect at least two reference measures for each participant in the calibration study.



STATISTICAL TESTING AND REGRESSION CALIBRATION

Slide 62

In this final section, I'm going to talk a little about statistical testing when dietary components are measured with error.

Hypothesis testing

- Null hypothesis: no association between dietary intake and the health outcome ($\alpha_T = 0$)
- Wald test statistic: estimated log odds ratio divided by its standard error

$$W = \hat{\alpha}_T / \text{s.e.}(\hat{\alpha}_T)$$

- W is approximately normal with
  - mean = $E(\hat{\alpha}_T) / \text{s.e.}(\hat{\alpha}_T)$
  - standard deviation = 1

Slide 63

A major part of statistics involves hypothesis testing. The researcher formulates a hypothesis, called the null hypothesis, and then tests whether or not it be can rejected. In our examples, we will test the null hypothesis that there is no association between dietary intake and the health outcome.

To test a hypothesis, we need a test statistic. We'll use the Wald test statistic, which we'll call W. In logistic regression, the Wald statistic equals the estimated log odds ratio divided by its standard error. The Wald statistic is known to be approximately normal, with expected value equal to the expected value of the estimated log odds ratio divided by its standard error.

Hypothesis testing

- Assumption: under null hypothesis, W has mean = 0
- Wald test: reject null hypothesis if W is too large

$$|W| \geq c$$

- Type I error = reject null when null is true
  Type II error = not reject null when null is false
- Significance level: probability of Type I error
  Power: 1 – probability of Type II error
- Controlling Type I error: choose c so that probability of Type 1 error is small, typically 5%

Slide 64

The main assumption of the Wald test is that, under the null hypothesis, the expected value of the Wald statistic, W, is equal to zero. Under this assumption, large values of W indicate that the null hypothesis is false. So the Wald test rejects the null hypothesis if the absolute value of W is greater than some value, c, which is called the critical value.

Now, there are two types of error associated with any test. Type I error occurs when the null hypothesis is rejected when it's really true. Type II error occurs when the null hypothesis is not rejected when it really is false.

Associated with these two types of error are the concepts of significance level and power. The significance level is the probability of making a Type I error, while the power is equal to 1 minus the probability of making a Type II error.

Both significance level and power depend on the critical value, c, that is chosen for the Wald test. Ideally, one would like to choose c so as to minimize the significance level while maximizing the power. Unfortunately, this isn't possible, since power is maximized when c is small, while the significance level is minimized when c is large.

Since statisticians consider Type I error to be the more serious type of error, the strategy is to control the probability of a Type I error by choosing c so that the significance level is small, typically 5 percent. If the resulting power is small, then one has to be very careful when interpreting the results. In particular, it's important to remember that failing to reject the null hypothesis is not the same as concluding that the null hypothesis is true. It may be that there simply wasn't enough power to detect that it was false. So, whenever you fail to reject the null hypothesis, you have to consider what your power was.

Validity of statistical tests

- A statistical test is valid if probability of a type I error really is at the chosen significance level
- For Wald test, condition holds only if the mean of W = 0 under the null hypothesis
- Since mean of W = $E(\hat{\alpha}_T) / s.e.(\hat{\alpha}_T)$, Wald test is valid only if

$$E(\hat{\alpha}_T) = 0 \text{ whenever } \alpha_T = 0$$

- If estimator is unbiased (i.e., if $E(\hat{\alpha}_T) = \alpha_T$), then Wald test is valid

Slide 65

A critical property of any test is *validity*. A test is valid if the probability of Type I error really equals the chosen significance level. For the Wald test, this condition holds only if the expected value of the test statistic equals zero under the null hypothesis. This means that the Wald test is valid only if the expected value of the estimated log odds ratio equals zero whenever the true log odds ratio is zero.

Note that if the estimated log odds ratio is unbiased, then the Wald test is valid, since then the condition always holds. For biased estimators, however, the Wald test may or may not be valid.

Is the uncorrected test valid?

- Wald test performed on the uncorrected log odds ratio will be called the “uncorrected test”
- For univariate exposures, the uncorrected log odds ratio is biased

$$E(\hat{\alpha}_Q) = \lambda_Q \alpha_T$$

- Nevertheless, if $\alpha_T = 0$, then $E(\hat{\alpha}_Q) = 0$
- Uncorrected Wald test for univariate exposure is valid

Slide 66

Now we ask: Is the uncorrected test valid? By uncorrected test, we mean the Wald test performed on the uncorrected log odds ratio.

For a single dietary exposure measured with error, we know that the expected value of the uncorrected log odds ratio is biased, equal to the attenuation factor times the true log odds ratio. Nevertheless, it's still true that whenever the true log odds ratio equals zero, so does the expected value of the uncorrected log odds ratio. Therefore, the uncorrected Wald test for a univariate exposure is valid.

Is the uncorrected test valid?

- For bivariate exposures, the uncorrected log odds ratio has mean

$$E(\hat{\alpha}_{Q1}) = \lambda_{Q11}\alpha_{T1} + \lambda_{Q12}\alpha_{T2}$$

- When $\alpha_{T1} = 0$,

$$E(\hat{\alpha}_{Q1}) = \lambda_{Q12}\alpha_{T2}$$

- Uncorrected Wald test for bivariate exposures is not valid
- If contamination factor λ_{Q12} is sufficiently small, then uncorrected test is approximately valid

Slide 67

For two exposures measured with error, the case is somewhat different. As we've seen, the uncorrected log odds ratio for one exposure is subject to contamination from the true effect of the other exposure. As a result, the uncorrected log odds ratio may be nonzero even when the true log odds ratio is zero. Therefore, the uncorrected test for two exposures is not valid.

If, however, the contamination factor is sufficiently small, then the probability of a Type I error will be close to the chosen value, and we can say that the test is *approximately* valid.

Is the uncorrected test valid?

OPEN – Estimated Contamination Factors (Freedman et al., J Nat Cancer Inst, 2011)

Nutrient	Gender	Energy	Protein Density	Potassium Density
Energy	Men	-	-0.01 (0.03)	0.13 (0.05)
	Women	-	0.03 (0.05)	0.10 (0.06)
Protein Density	Men	0.08 (0.05)	-	-0.01 (0.09)
	Women	0.06 (0.05)	-	0.00 (0.10)
Potassium Density	Men	0.04 (0.04)	-0.05 (0.06)	-
	Women	-0.04 (0.05)	0.00 (0.07)	-
Total Fat Density	Men	0.05 (0.05)	-0.03 (0.07)	0.00 (0.08)
	Women	-0.07 (0.05)	-0.02 (0.08)	-0.08 (0.10)
Saturated Fat Density	Men	0.10 (0.04)	-0.03 (0.05)	-0.04 (0.07)
	Women	-0.02 (0.04)	-0.01 (0.06)	0.07 (0.08)

Slide 68

Here is another table taken from the recent commentary by Freedman and colleagues. It shows estimated contamination factors in the OPEN study for a model that has energy, protein density, potassium density, and one other nutrient density. We see that the estimated contamination factors are generally small, and only two are statistically significantly different from zero.

Is the uncorrected test valid?

- Contamination factors are generally small and not statistically significant
- Freedman et al. (2011) concluded that statistical tests for uncorrected test with multiple dietary exposures will be approximately valid
- Caveat: conclusion based on only three nutrients: protein, potassium and energy

Slide 69

Freedman and colleagues concluded from this that the uncorrected test for multiple dietary exposures measured with error should be approximately valid. Again, they included the caveat that this conclusion was based on only three nutrients for which reference biomarkers are available.

Statistical power

- Statistical power is the probability of rejecting the null hypothesis when the null hypothesis is false
- Equivalently, the probability of detecting an association as statistically significant
- Power depends on the size of true log odds ratio:
  - For α_T close to 0, power is small
  - As $|\alpha_T|$ increases, power increases

Slide 70

Now we're going to look at the effect of measurement error on statistical power. As we saw earlier, power is the probability to reject the null hypothesis when it is false. Equivalently, we can say it is the probability of detecting an association as statistically significant when there really is an association.

Power depends in part on the size of the true log odds ratios. For log odds ratios close to zero, the power will be small. As the absolute value of the log odds ratio increases, the power increases.

Statistical power for uncorrected test

- Power of Wald test depends on |E(W)| = absolute value of the expected value of Wald statistic W
- For univariate exposures, expected value of W for the uncorrected test is

$$E(W) = Corr(Q,T) \times E(W_T)$$

$E(W_T)$ = expected value of W if T could be measured without error

- Since |Corr(Q,T)| ≤ 1, measurement error always leads to loss of power (unless |Corr(Q,T)| = 1)

Slide 71

The power of the Wald test depends on the expected value of the Wald test statistic, W. The greater the expected value of W, the greater the power to detect the association.

For a single dietary exposure measured with error, the expected value of W for the uncorrected test is equal to the correlation of reported and true intake times what the expected value of W would have been if true intake could have been measured.

Since the correlation of reported and true intake is always between -1 and 1, this means that measurement error in reported intake always leads to a loss of power.

Example: loss of power for uncorrected test

- NIH-AARP Diet and Health Study
- For fat intake, estimated $\rho_{QT} = 0.40$
- In a study with N subjects:
  - If power = 90% using true intake T, then power = 25% using Q instead of T
  - To get 90% power using Q, would need a sample size of $N / \rho_{QT}^2 = 6.25 \times N$
- Similar loss of power for multivariate exposures measured with error

Slide 72

As an example, the estimated correlation between FFQ-reported fat intake and true intake in the AARP study is 0.4.

Consider a study with N subjects. If the study were designed to have 90 percent power to detect a fat-breast cancer association under the assumption that fat intake could be measured exactly, then in reality the study would only have 25 percent power to detect the association using the FFQ to measure intake. And in order to have 90 percent power using the FFQ, one would have needed over six times as many subjects.

There would be a similar loss of power for multiple dietary exposures measured with error.

Statistical power for regression calibration

- Regression calibration adjusts the estimated log odds ratio by dividing by the attenuation coefficient

$$\hat{\alpha}_T = \hat{\alpha}_Q / \hat{\lambda}_Q$$

- However, this adjustment changes the standard error of the estimate

$$se(\hat{\alpha}_T) \approx \sqrt{\left(\frac{se(\hat{\alpha}_Q)}{\hat{\lambda}_Q}\right)^2 + \left(\frac{\hat{\alpha}_Q se(\hat{\lambda}_Q)}{\hat{\lambda}_Q^2}\right)^2} > \frac{se(\hat{\alpha}_Q)}{\hat{\lambda}_Q}$$

Slide 73

We have seen that regression calibration can correct for bias. Can it also recover lost power?

As we've seen, regression calibration increases the attenuated log odds ratio by dividing it by the attenuation factor. However, because of the uncertainty added by having to estimate the attenuation factor, the standard error increases by an even larger amount.

Statistical power for regression calibration

- As a result:
  - Expected value of test statistic W does not increase (it decreases slightly)
  - Wald test based on regression calibration has slightly less power than the unadjusted test
- Regression calibration in its usual form corrects for bias, but does recover power lost due to measurement error

Slide 74

As a result, regression calibration causes a slight decrease in the expected value of the Wald statistic, and so regression calibration has slightly less power than the uncorrected test.

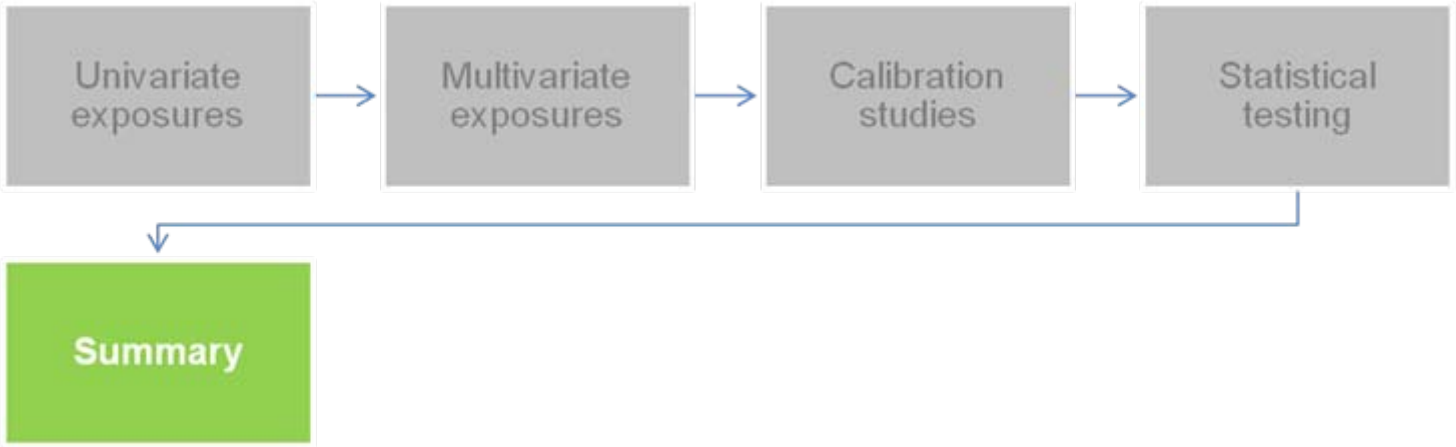
In summary, regression calibration in its usual form does not recover power lost due to measurement error.

Can regression calibration be made more powerful?

- An enhanced version of regression calibration can sometimes be used to recover (some of) the power lost due to measurement error
- Idea: predict T using E(T|Q, Z, C) instead of E(T|Q, Z), where C is a variable that:
  i. Helps to predict true intake, but
  ii. Is not related to disease outcome conditional on true intake and covariates Z
- See lecture 10 in the series for fuller discussion

Slide 75

Is it possible to make regression calibration more powerful? There is an enhanced version that can sometimes be used to recover some of the lost power. The idea is to add an additional variable, C, to the prediction equation for true intake, T. The variable C must have two special qualities: first, it helps predict true intake; and second, it is not related to the health outcome conditional on true intake and covariates, Z. There will be a fuller discussion of this idea in webinar 10, when we consider how to combine different self-report instruments to improve prediction of true intake.


Univariate exposures
Multivariate exposures
Calibration studies
Statistical testing
Summary

SUMMARY

Slide 76

[No notes.]

Summary

1. Measurement error causes attenuation of estimated risk parameters and loss of power to detect diet-health associations.

2. Regression calibration is an accessible method for adjusting these attenuated estimates to remove bias.

3. Calibration studies are needed to provide the information necessary to apply regression calibration.

4. In its usual form, regression calibration does not recover power lost due to measurement error.

5. Statistical tests of uncorrected risk estimates are, on current evidence, approximately valid.

Slide 77

I'd like to summarize the main points in this talk. First, measurement error causes attenuation of estimated risk parameters and loss of power to detect diet-health associations. Second, regression calibration is an accessible method for adjusting these attenuated estimates to remove bias. Third, calibration studies are needed to provide the information necessary to apply regression calibration. Fourth, in its usual form, regression calibration does not recover power lost due to measurement error. Fifth, statistical tests of uncorrected risk estimates are, on current evidence, approximately valid.

QUESTIONS & ANSWERS

Moderator: Amy Subar

Please submit questions using the *Chat* function

Slide 78

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

Measurement Error Webinar 7 Q&A

Question: **Regarding the multivariate regression analysis that was including multiple error-prone variables and the use of the attenuation contamination factor, and is that a term that's commonly used and accepted in the literature? You mention it again for the Freedman paper.**

Yes, it's a term that we use here at NCI in our measurement error group and use it in published papers. Other terms for it would be "residual confounding". I think the idea of an attenuation/contamination matrix is a generally used terminology. *(D. Midthune)*

In the fat-cancer example you had an alcohol-cancer example, what percent of the standard error came from uncertainty of the regression calibration?

Well, I didn't measure it exactly. I would assume that it was not a substantial amount simply because when we looked at the attenuated log odds ratio, divided by its standard error, we got a result that was similar than when we looked at the de-attenuated estimate and its standard error. So the ratio between the two estimates and the standard errors didn't change that much, so I'd say that the increase due to the uncertainty was not that noticeable. *(D. Midthune)*

For the bootstrap example, you used normal theory formulas to get confidence intervals using the bootstrap standard error. Could you also have computed a confidence interval empirically from the bootstrap replicate?

Yes, you can do that. I used the normal theory just because it was a little simpler to explain but an alternative method is to take the bootstrap estimates and find the cutpoints in that empirical distribution. And sometimes, under some circumstances where the normal theory doesn't hold, this empirical method may be superior. *(D. Midthune)*

If 24 hour recalls are used as reference instruments, is there any benefit to modeling multiple prediction equations jointly instead of one at a time? So if two exposures are measured at the same time, like fat and vitamin C, can the temporal similarity provide additional precision?

Well, it can under certain conditions. In this example, we have two prediction equations and both prediction equations include both reported intakes measured with error, so T, true intake 1, was related to both reported intakes 1 and 2. And if that's the case, then you don't really gain any precision by modeling jointly. But if you can, maybe by testing,

determine that the first prediction equation only depends on the first reported intake, and the second prediction equation only depends on the second reported intake, then you can actually get gain by modeling them jointly. This is basically the idea of seemingly unrelated equations, or in this case seemingly unrelated measurement error models. *(D. Midthune)*

This refers to the bootstrap procedure you described, the two-stage bootstrap procedure. So it can be done in two ways: You get the estimate at the first step and then apply the best estimate to the second step, or bootstrapping the two stages simultaneously. Which would you recommend?

Well, I would say it actually depends on the software you have. Like I said, they give exactly the same (parameter) estimates and very similar estimates of the confidence intervals. The only exception is the bootstrap method may work better when some of your underlying assumptions about the normality of the data are violated in an extreme manner. If they are approximately normally distributed, then either method should give similar results. So the what I called the Rosner method is a little simpler in that you don't have to do this bootstrap sampling, so I'd say it's a matter of convenience. *(D. Midthune)*

Other than using at least two reference instruments like two recalls, do you have any general rules for planning the number of measurements and sample size for epi studies?

Actually, we're going to be considering that; actually, I'm giving the tenth webinar, which is about combining dietary instruments, and so in that talk we're going to consider how many reference instruments are ideal, or perhaps not ideal but at least better than otherwise. *(D. Midthune)*

Is there a limit to how small an attenuation factor can be before you would hesitate to use these adjustment methods?

Well, definitely, the smaller the attenuation factor the more uncertain any adjustment is going to be. I don't know if there is a hard limit but, certainly, I would say if the attenuation factor is below .2, .3; you definitely would not want to use it if it gets close to zero, because dividing by zero basically inflates your log odds ratio to infinity. So you want to keep your estimate away from zero. So maybe .3 might be a rule of thumb. *(D. Midthune)*

National Cancer Institute

measurement ERROR webinar series

Next Session

Tuesday, November 8, 2011
10:00-11:30 EST

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

Victor Kipnis
National Cancer Institute

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

Slide 79

Thank you very much, Doug, and thanks to our audience for joining today's webinar. Please join us next week for webinar 8, when Dr. Victor Kipnis will continue the discussion of methods of accounting for measurement error in the assessment of diet and health relationships, but this time with a focus on episodically-consumed dietary components.