National Cancer Institute

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

measurement ERROR webinar series

The problem of measurement error when examining diet-health relationships

Laurence Freedman, PhD, MA
Gertner Institute for Epidemiology

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.

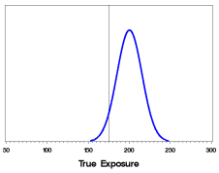
Presenters and Collaborators

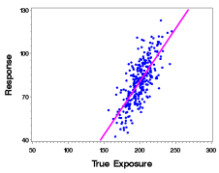
	Sharon Kirkpatrick *Series Organizer*	
Regan Bailey	Laurence Freedman	Douglas Midthune
Dennis Buckman	Patricia Guenther	Amy Subar
Raymond Carroll	Victor Kipnis	Fran Thompson
Kevin Dodd	Susan Krebs-Smith	Janet Tooze

Wake Forest School of Medicine · The Gertner Institute · STATISTICS Texas A&M University

Two main areas of interest

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

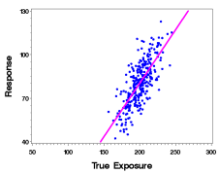




- **Estimating diet-health relationships**: regression coefficients

Two main areas of interest

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

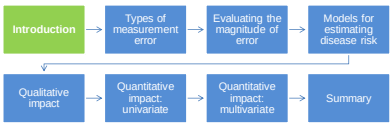


- **Estimating diet-health relationships**: regression coefficients

Objectives

Learning objectives

- Knowing the types and magnitudes of measurement error that occur in dietary data
- Reviewing statistical models for evaluating diet-health relationships
- Understanding the qualitative and quantitative impact of measurement error on studies of diet-health relationships



INTRODUCTION

Types of "analytic" studies

- Animal experiments
- Ecological studies
- Cross-sectional studies
- Case-control studies
- Cohort studies
- Randomized disease prevention trials

Red → Estimated diet-health relationship impacted by dietary measurement error

The exposure (1)

- In these studies we wish to relate:

Dietary Intake → Health Outcome

- The measure of intake thought to be most relevant is:
  - usual intake, i.e., long-term average daily intake

The exposure (2)

- In surveillance studies, "long-term" is often taken to be 1 or 2 years
- In cohort and case-control studies, it is less-well defined but often may be thought of as covering several years

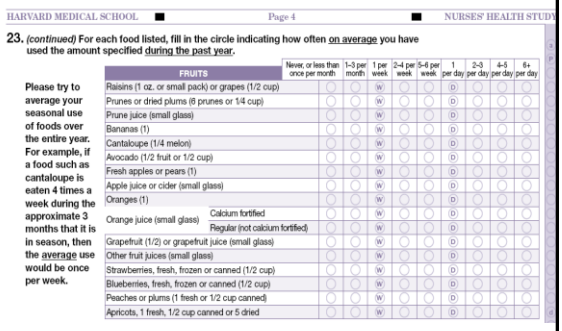
The exposure (3)

- Clearly, to measure an individual's average intake over a long period is a challenging task
- Fortunately, one does not need to measure usual intake exactly in order to make progress

Instruments (1)

- Food Frequency Questionnaires
  - Main instrument for large cohort and case-control studies
  - Inexpensive to administer
  - Aims to measure long-term average intake
- BUT
  - Inaccurate long-term recall
  - Cognitively difficult
  - Conversion to nutrient and food group intakes is difficult

Instruments (2)



HARVARD MEDICAL SCHOOL ■ Page 4 ■ NURSES' HEALTH STUDY

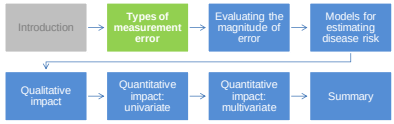
23. *(continued)* For each food listed, fill in the circle indicating how often on average you have used the amount specified during the past year.

Please try to average your seasonal use of foods over the entire year. For example, if a food such as cantaloupe is eaten 4 times a week during the approximate 3 months that it is in season, then the average use would be once per week.

FRUITS	Never, or less than once per month	1–3 per month	1 per week	2–4 per week	5–6 per week	1 per day	2–3 per day	4–5 per day	6+ per day
Raisins (1 oz. or small pack) or grapes (1/2 cup)			W			D			
Prunes or dried plums (6 prunes or 1/4 cup)			W			D			
Prune juice (small glass)			W			D			
Bananas (1)			W			D			
Cantaloupe (1/4 melon)			W			D			
Avocado (1/2 fruit or 1/2 cup)			W			D			
Fresh apples or pears (1)			W			D			
Apple juice or cider (small glass)			W			D			
Oranges (1)			W			D			
Orange juice (small glass) — Calcium fortified			W			D			
Orange juice (small glass) — Regular (not calcium fortified)			W			D			
Grapefruit (1/2) or grapefruit juice (small glass)			W			D			
Other fruit juices (small glass)			W			D			
Strawberries, fresh, frozen or canned (1/2 cup)			W			D			
Blueberries, fresh, frozen or canned (1/2 cup)			W			D			
Peaches or plums (1 fresh or 1/2 cup canned)			W			D			
Apricots, 1 fresh, 1/2 cup canned or 5 dried			W			D			

Instruments (3)

- 24-hour recall
  - Main instrument for surveillance studies
  - Questions what the individual ate over the past 24 hrs
  - More accurate than FFQ (short-term recall only)
  - More detail makes conversion to nutrients easier
- BUT
  - Very expensive to collect and code (up to now)
  - Does not target usual intake



TYPES OF MEASUREMENT ERROR

Types of measurement error (1)

- Additive systematic bias
  - Systematic: the instrument introduces a bias that is common to all individuals

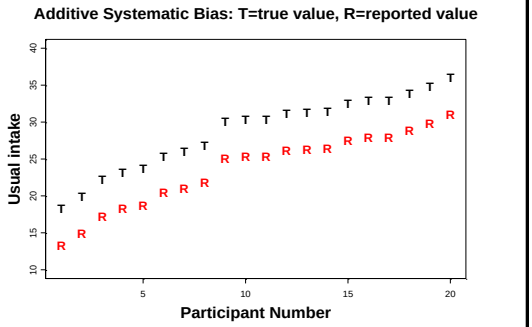
$$R_i = \beta_0 + T_i$$

R_i denotes reported usual intake of individual i

T_i denotes true usual intake of individual i

(All reported values are different by the constant amount β_0 from what they should be)

Types of measurement error (2)



Types of measurement error (3)

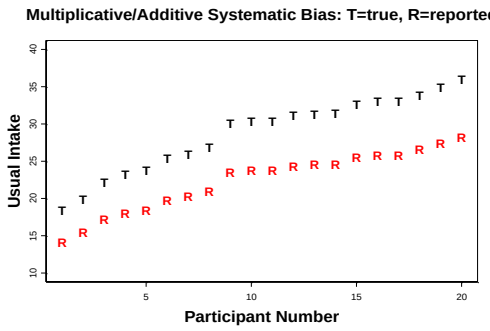
- Multiplicative and additive systematic bias

$$R_i = \beta_0 + \beta_1 \times T_i$$

(In addition to additive bias, the true value is scaled up or down by factor β_1)

Types of measurement error

Types of measurement error (4)



Types of measurement error

Types of measurement error (5)

- Person-specific (random) bias
  - Bias that occurs at the individual level - it is specific to an individual but can differ among individuals

$$R_i = \beta_0 + \beta_1 \times T_i + u_i$$

(In addition to additive and multiplicative systematic bias, there is a bias u_i that varies for each individual i)

Types of measurement error

Types of measurement error (6)

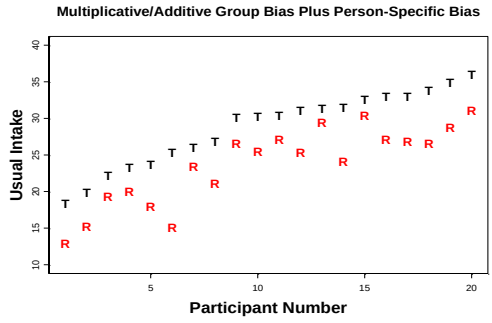
- Person-specific (random) bias

$$R_i = \beta_0 + \beta_1 \times T_i + u_i$$

  - The subject-specific bias u_i is a random term
  - Its magnitude is quantified by SD(u_i), its standard deviation

Types of measurement error

Types of measurement error (7)



Types of measurement error

Types of measurement error (8)

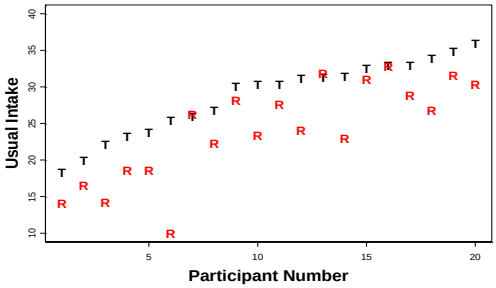
- Within-person random error
  - Variation in reporting by an individual over a series of repeat reports

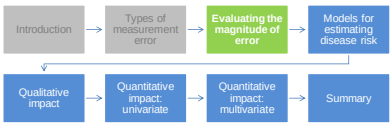
$$R_{ij} = \beta_0 + \beta_1 \times T_i + u_i + \varepsilon_{ij}$$

  - The extra subscript j denotes the sequence number in a series of reports
  - The extra term ε_{ij} is the within-person error that is on average zero
  - Its magnitude is quantified by SD(ε_{ij})

Types of measurement error

Types of measurement error (9)





EVALUATING THE MAGNITUDE OF ERROR

Evaluating the error (1)

- How can we study the errors made in dietary reporting?
  - Validation studies comparing reports with "reference" measures of dietary intake
- Ideal properties of a reference instrument
  i. Unbiased
  ii. Errors uncorrelated to true intake
  iii. Errors uncorrelated with self-report errors

Evaluating the error (2)

- Do we have any ideal "reference" measures?
  - Direct observation (feeding studies)
  - "Recovery" biomarkers: based on recovery of specific biologic products that are directly related to intake and are not subject to substantial inter-individual differences in metabolism
    - Doubly labeled water for energy intake
    - 24-hour urinary nitrogen for protein intake
    - 24-hour urinary potassium for potassium intake
  - "Concentration" biomarkers (e.g., serum lipids) do not share these properties

Evaluating the error (3)

- OPEN (Observing Protein and Energy Study)
  - 261 men; 223 women
  - Adult volunteers residing in Maryland, USA
- Completed:
  - 24-hour recall x 2
  - Food frequency questionnaire x 2
  - 24-hour urinary nitrogen x 2
  - 24-hour urinary potassium x 2
  - Doubly-labeled water x 1 (in 25 persons x 2)

Evaluating the error (4)

Results from OPEN – Means

Sex	Method	Energy (kcal/d)	Protein (g/d)
Men	Marker	2842	105.5
	FFQ	1961	74.7
	24HR	2522	92.2
Women	Marker	2273	77.5
	FFQ	1524	57.2
	24HR	1919	70.9

Evaluating the error? (5)

Results from OPEN – Protein Intake (after log transformation)

Sex	Method	Scaling Factor, β_1	Person-Specific Bias (SD)	Within-Person Error (SD)
Men	FFQ	0.67	0.36	0.19
	24HR	0.70	0.20	0.30
Women	FFQ	0.65	0.33	0.22
	24HR	0.60	0.16	0.35

Evaluating the error? (6)

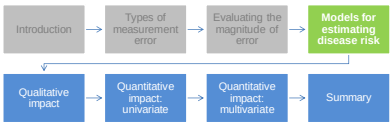
Results from OPEN – Protein Density (after log transformation)

Sex	Method	Scaling Factor, $ß_1$	Person-Specific Bias (SD)	Within-Person Error (SD)
Men	FFQ	0.46	0.13	0.11
	24HR	0.61	0.11	0.24
Women	FFQ	0.37	0.15	0.12
	24HR	0.39	0.11	0.26

Evaluating the error? (7)

Summary of results of OPEN and other large validation studies (AMPM, NBS)

- **Serious under-reporting**
  Energy: FFQ by 30% and 24HR by 10%
- **Food frequency questionnaire (FFQ)**
  Large systematic error, large person-specific bias, small within-person random error
- The biases and random error can be reduced by energy adjustment
- **24-hour recall (24HR)**
  Smaller systematic error, large within-person random error, smaller person-specific bias
- The within-person random error of the 24HR is largely day-to-day variation and can be reduced by using several repeats



MODELS FOR ESTIMATING DISEASE RISK

Estimating disease risk (1)

- Before we study the impact of measurement error on studying diet-health relationships, we need to review measures and statistical models for disease risk
- The two main measures of disease risk are:
  - Relative risk
  - Odds ratio

Estimating disease risk (2)

- When comparing two groups, exposed and unexposed:

$$\text{Relative risk} = \frac{\text{Prob (disease in exposed)}}{\text{Prob (disease in unexposed)}}$$

$$\text{Odds (disease)} = \frac{\text{Prob (disease)}}{1 - \text{Prob(disease)}}$$

$$\text{Odds ratio} = \frac{\text{Odds (disease in exposed)}}{\text{Odds (disease in unexposed)}}$$

Estimating disease risk (3)

- Elements of a nutrition regression model
  1. A health outcome variable (Y)
  2. A set of explanatory variables, (T_1,T_2, Z_1…,Z_p) The T–variables are dietary exposures, and the Z-variables are other exposures, confounders, effect modifiers or intermediate variables
  3. An equation linking the outcome to the explanatory variables

Models for estimating disease risk

Estimating disease risk (4)

For example, logistic regression:

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

- Where:
  Y is a binary variable;
  Y=1 denotes disease ("case")
  Y=0 denotes no disease ("healthy")
- α's are the regression parameters and represent log odds ratios
- Each α represents the increase in the log odds of disease associated with increasing the corresponding variable by 1 unit while keeping the other variables fixed.

Models for estimating disease risk

Estimating disease risk (5)

Estimating an odds ratio: binary exposure
Israeli National Ovarian Cancer Case-Control Study

- Oral Contraceptive use (0=<6m use, 1=6m+ use)
  889 cases; 1747 controls

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 OC$$

- Output from logistic regression program
  Coefficients:

	Value	Std. Error	p value
(Intercept)	-0.65	0.046	<0.0001
ocon1	**-0.13**	**0.10**	**0.19**

- Odds ratio estimate for OC = $\exp(\alpha_1) = \exp(-0.13) = 0.87$

Models for estimating disease risk

Estimating disease risk (6)

Estimating an odds ratio: continuous exposure
Animal Fat intake (kcal/d) from a FFQ

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 afatcal$$

Output from logistic regression program

	Value	Std. Error	p value
(Intercept)	-1.18	0.10	<0.0001
afatcal	**0.0017**	**0.00030**	**<0.0001**

- Odds ratio estimate for increase in animal fat intake of 1kcal/d = $\exp(\alpha_1) = \exp(0.0017) = 1.0017$
- Odds ratio estimate for increase in animal fat of 100kcal/d = $\exp(100\times\alpha_1) = \exp(100\times0.0017) = 1.18$

Models for estimating disease risk

Estimating disease risk (7)

Energy adjustment

- Practical question:
  - A study has been conducted with a FFQ as the main dietary instrument
- When evaluating an association between a FFQ-reported nutrient intake and the health outcome should one adjust for FFQ-reported total energy?

Models for estimating disease risk

Estimating disease risk (8)

Possible reasons for energy adjustment
(see Willett, Howe and Kushi, 1997)

- Energy is a confounder
- The energy-adjusted relative risk is more relevant to public health interests
- **The adjustment increases the precision of the relative risk estimate**

Models for estimating disease risk

Estimating disease risk (9)

Energy adjustment models

- There are several different methods for energy adjustment – we will look at two:
  i. Standard model
  ii. Density model

Estimating disease risk (10)

Energy adjustment models

- **Standard model**:
  Add total energy intake as a second explanatory variable, for example:

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 afatcal + \alpha_2 energy$$

- Meaning of the coefficient α_1 changes:
  The log odds ratio associated with increasing animal fat intake by 1 kcal while keeping total energy intake fixed
- **which means:**
  The log odds ratio associated with **substituting** 1 kcal of animal fat for 1 kcal of other nutrients

Estimating disease risk (11)

Energy adjustment: Standard model

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 afatcal + \alpha_2 energy$$

	Value	Std. Error	p value
(Intercept)	-1.39	0.13	<0.0001
afatcal	**0.00093**	**0.00042**	**0.027**
energy	0.00025	0.000098	0.009

- Odds ratio for 100kcal increase = exp(0.00093 x 100) = 1.10
- Remember that this association is with **substituting** 100kcal of animal fat for 100kcal of other food sources

Estimating disease risk (12)

Energy adjustment models

- **Density model:**
  Nutrient density = 100 x (nutrient intake in kcal / total energy intake in kcal)%
- Express the nutrient as a **nutrient density** and add total energy intake as a second explanatory variable.

  For example:

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 afatdens + \alpha_2 energy$$

- Meaning of the coefficient α_1:
  The log odds ratio associated with increasing animal fat density by 1% while keeping total energy intake fixed

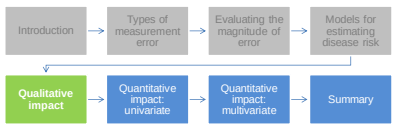
Estimating disease risk (13)

Energy adjustment: Density model

$$\log\{Odds(Y=1)\} = \alpha_0 + \alpha_1 afatdens + \alpha_2 energy$$

	Value	Std. Error	p value
(Intercept)	-1.66	0.20	<0.0001
afatdens	**0.01560**	**0.00760**	**0.041**
energy	0.00041	0.00007	<0.0001

- Odds ratio for 5% increase = exp(0.0156 x 5) = 1.08
- Remember that this association is with increasing animal fat density by 5% while keeping energy intake fixed



QUALITATIVE IMPACT OF MEASUREMENT ERROR

Qualitative impact of error (1)

Additive systematic bias

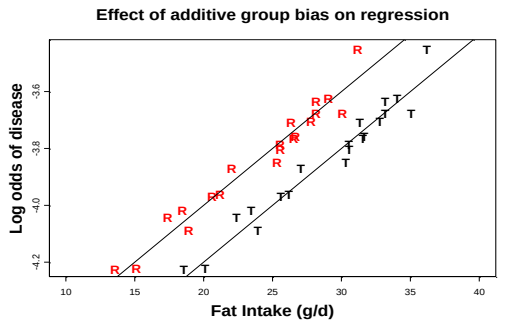
- Suppose we have an instrument with additive systematic bias but no subject-specific bias and no random error.

$$R_i = \beta_0 + T_i$$

- Then:
  a. Log odds ratio estimates are unchanged
  b. Scatter about the regression line is unchanged
  c. Significance tests are unaffected
  d. Study power is unaffected
- Additive systematic bias is not a problem for detecting a relationship!
  But translation to public health message **is** affected

Qualitative impact

Qualitative impact of error (2)



Qualitative impact

Qualitative impact of error (3)

Additive and multiplicative systematic bias

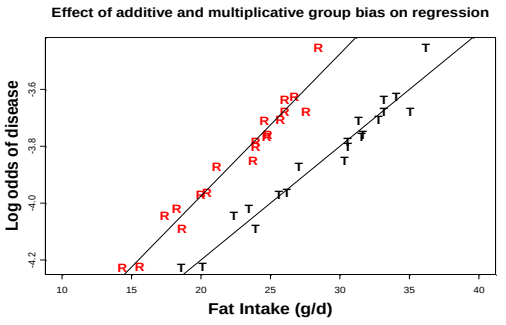
- Suppose we have an instrument with additive and multiplicative systematic bias but no person-specific bias and no random error

$$R_i = \beta_0 + \beta_1 \times T_i$$

- Then:
  - Log odds ratio estimates are scaled by $1/\beta_1$
  - Scatter about the regression line is unchanged
  - Significance tests are unaffected
  - Study power is unaffected
- Systematic bias is not the major problem for detecting a relationship!

Qualitative impact

Qualitative impact of error (4)



Qualitative impact

Qualitative impact of error (5)

Person-specific bias and within-person random error

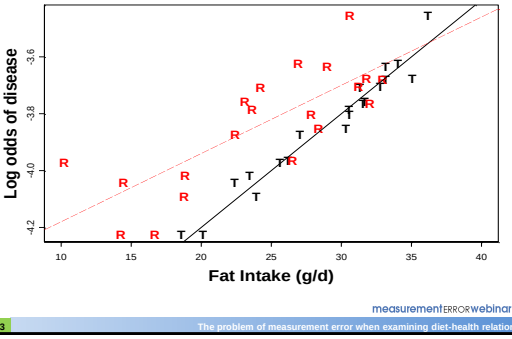
- Suppose we have an instrument with systematic bias and also person-specific bias and within-person random error

$$R_{ij} = \beta_0 + \beta_1 \times T_i + u_i + \varepsilon_{ij}$$

- Then:
  - a) **Log odds ratio estimates are factored down (attenuated)**
  - b) Scatter about the regression line is increased
  - c) Significance tests are less powerful but still valid
  - d) **Study power is decreased**
- Person-specific bias and within-person random error are a major problem for detecting a relationship!

Qualitative impact

Qualitative impact of error (6)

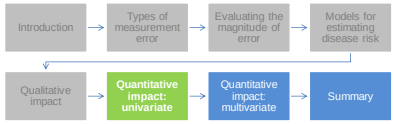


Qualitative impact

Qualitative impact of error (7)

Summary

- When a **single** dietary exposure measured with error is included in a disease outcome regression model:
- Then:
  - a) Log odds ratio estimates are factored down (attenuated)
  - b) Study power is decreased
  - c) Significance tests are less powerful but still valid
- These conclusions seems to hold also for **several** dietary exposures entered together in the same model (e.g., energy-adjustment models) – see later details
- We now quantify the seriousness of these problems



QUANTITATIVE IMPACT OF MEASUREMENT ERROR: UNIVARIATE MODELS

Quantitative impact of error (1)

- We will now quantify the extent of the two main problems:
  a) Log odds ratio estimates are attenuated
  b) Study power is decreased

Quantitative impact of error (2)

Log odds ratio attenuation for
a single continuous dietary intake variable

- Assume we have systematic error, subject-specific bias and random error.
- **Expected log odds ratio estimate = λ × true value,**
- where
  - λ = attenuation factor
    = slope of regression of T (truth) on R (report)
  - λ is nearly always <1 and usually a lot less!
- When the log odds ratio is attenuated, the odds ratio moves towards 1.0

Quantitative impact of error (3)

Log odds ratio attenuation for
a single continuous dietary intake

- OPEN: Attenuation Factors for FFQ and 24HR (Men)
  - (Obtained by regressing recovery biomarker on self-report)

	FFQ	24HR
Energy	0.08	0.18
Protein	0.16	0.20
Protein Density	0.40	0.23

Quantitative impact of error (4)

Implications of these results

- Suppose the attenuation factor λ is **0.16** (as for protein)
- Suppose the true odds ratio between the 90th and 10th percentiles of true intake is **2.5** (i.e., substantial)
  - log OR = log(2.5) = 0.92
  - Expected estimated log OR = 0.92 x 0.16 = 0.147
  - Expected estimated OR = exp(0.147) = **1.16**

Quantitative impact of error (5)

Implications of these results (cont'd)

- **Almost impossible to detect an OR of 1.16 in a case-control or cohort study**
- Reasons:
  a) Enormous sample sizes required to obtain statistical significance (see later)
  b) Cannot eliminate **all** confounding
- The limit of detection for an OR is probably around 1.25

Quantitative impact of error (6)

Implications of these results (cont'd)

- Fortunately, after energy adjustment, attenuation factors with an FFQ are larger (e.g., **0.40** for protein density)
- Suppose the true odds ratio between the 90th and 10th percentiles is **2.5** (i.e., substantial)
  - log OR = log(2.5) = 0.92
  - Expected estimated log OR = 0.92 x 0.40 = 0.368
  - Expected estimated OR = exp(0.368) = **1.44**
- Such an odds ratio is more possible to detect, although still difficult!

Quantitative impact of error (7)

Log odds ratio attenuation for
a single categorized dietary intake

- Assume we categorize our intake into quantiles (e.g., tertiles, quartiles or quintiles)
- The log odds ratio is still attenuated but by a different amount:
- **Expected log odds ratio estimate = $\rho \times$ true value**, where
  ρ = correlation between R (report) and T (truth)
- In other words, for analysis by quantiles, log odds ratios are attenuated by ρ, instead of λ

Quantitative impact of error (8)

Log odds ratio attenuation for
a single categorized dietary intake

OPEN: Correlations with True Usual Intake for FFQ and 24HR (Men)

	FFQ	24HR
Energy	0.20	0.34
Protein	0.32	0.38
Protein Density	0.43	0.38

Quantitative impact of error (9)

Log odds ratio attenuation for categorized variables

- Implications of these results are similar to those stated earlier
- After energy adjustment, the estimated log odds ratios will be greatly attenuated by a factor of about 0.4 for protein density

Quantitative impact of error (10)

Decrease in study power

- Assume we have systematic bias, subject-specific bias and within-person random error

  Effective sample size = Actual sample size x ρ^2
- Where:
  ρ = correlation of R (report) with T (truth)

Quantitative impact of error (11)

Decrease in study power

OPEN: Correlations with 'truth' for FFQ and 24HR (Men)

	FFQ	24HR
Energy	0.20	0.34
Protein	0.32	0.38
Protein Density	0.43	0.38

- **Example: Protein Density**
  FFQ: Effective sample size = 0.43^2 x actual sample size
  = 0.18 x actual sample size
- We effectively lose 82% of our sample size!

Quantitative impact of error (12)

Decrease in study power

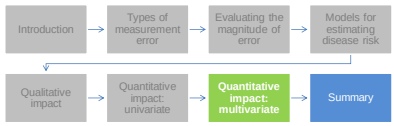
- Suppose that we had calculated a sample size of **50,000** for a cohort study that would give **90%** power for detecting an association the 5% significance level, assuming that we could measure dietary intake exactly
- Then, because of the measurement error we would need $50{,}000/\rho^2 = 50{,}000/0.432^2 =$ **270,000** to preserve the power of 90%

Quantitative impact of error (13)

Decrease in study power

- If we proceeded with the study with sample size 50,000 then the statistical power would be decreased by measurement error from 90% to 28%
  - The formula is given by:

    Power = $\Phi^{-1}(3.24\rho - 1.96)$
  - Where the symbol Φ^{-1} denotes the inverse of the standard normal cumulative distribution function



QUANTITATIVE IMPACT OF MEASUREMENT ERROR: MODELS WITH MULTIVARIATE EXPOSURE

Quantitative impact: multivariate exposures (1)

Two or more dietary variables in the disease regression model

- Typical example: Standard energy-adjustment model

  $\log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_1 \text{afatcal} + \alpha_2 \text{energy}$
- The effects of measurement error in these models is in theory less straightforward:
  i. Estimated log odds may be biased but not attenuated (i.e., inflated)
  ii. Statistical tests may not be valid

Quantitative impact: multivariate exposures (2)

- These problems arise from **residual confounding:**
  - One error-prone exposure and one exactly measured exposure in the same model
  - If the two (true) exposures are correlated, then the exactly measured one will adopt part of the effect of the error-prone exposure
  - When both are measured with error, they will each adopt different fractions of the other's effect!

Quantitative impact: multivariate exposures (3)

- Suppose we have two nutrient intakes.
  There exists an "attenuation-contamination" matrix, as follows:

λ_{11}	λ_{12}
λ_{21}	λ_{22}

- If the true log odds ratios for the two nutrients are α_1 and α_2, then the estimated ones are expected to be:
  $\lambda_{11} \times \alpha_1 + \lambda_{12} \times \alpha_2$ and $\lambda_{22} \times \alpha_2 + \lambda_{21} \times \alpha_1$
- The magnitudes of λ_{12} and λ_{21} tell us how serious is the residual confounding.
  We call them **contamination factors**

Quantitative impact: multivariate exposures (4)

- If λ_{12} and λ_{21} are small, then the only bias in the estimated log odds ratios comes essentially from attenuation, then:
  a) The estimated log odds ratio is attenuated
  b) The significance test is valid
- So we need to know for dietary data, how large are the contamination factors
- We can estimate them from the OPEN study

Quantitative impact: multivariate exposures (5)

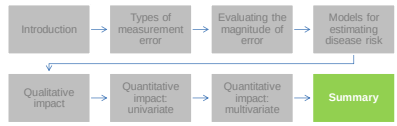
OPEN – Estimated Contamination Factors
(Freedman, Schatzkin, Midthune, Kipnis, J Nat Cancer Inst 2011)

Dietary Component	Gender	Protein Density		Potassium Density		Energy	
Energy	Men	-0.01	(0.03)	0.13	(0.05)	-	
	Women	0.03	(0.05)	0.10	(0.06)	-	
Protein Density	Men	-		-0.01	(0.09)	0.08	(0.05)
	Women	-		0.00	(0.10)	0.06	(0.05)
Potass. Density	Men	-0.05	(0.06)	-		0.04	(0.04)
	Women	0.00	(0.07)	-		-0.04	(0.05)
Total Fat Density	Men	-0.03	(0.07)	0.00	(0.08)	0.05	(0.05)
	Women	-0.02	(0.08)	-0.08	(0.10)	-0.07	(0.05)
Sat. Fat Density	Men	-0.03	(0.05)	-0.04	(0.07)	0.10	(0.04)
	Women	-0.01	(0.06)	-0.07	(0.08)	-0.02	(0.04)

Quantitative impact: multivariate exposures (6)

OPEN: Contamination factors

- Contamination factors generally appear small, meaning that residual confounding does not appear to be a serious problem
  – However, note that OPEN and other recovery biomarker validation studies examine only energy, protein and potassium
  – Similar findings for other nutrients cannot be guaranteed



SUMMARY

Summary

1. Errors in self-reported dietary intake have a complex structure including systematic biases, person-specific biases and within-person random error
2. The person-specific biases and within-person random error have a profound impact on the estimation of disease risk parameters such as the log odds ratio. Estimates of these are severely attenuated
3. For a FFQ, these effects can be partially mitigated by energy-adjustment
4. The same biases and random errors also cause loss of statistical power for detecting diet-health relationships

What's coming next?

- In the next lecture, we will study how we can correct the attenuation in the estimated disease risk parameter
- This will require us to learn about calibration studies and also a neat statistical method known as regression calibration

QUESTIONS & ANSWERS

Moderator: Sharon Kirkpatrick

Please submit questions using the *Chat* function

National Cancer Institute

measurement ERROR webinar series

Next Session

Tuesday, November 1, 2011
10:00-11:30 EDT

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

Douglas Midthune
National Cancer Institute

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