The problem of measurement error when examining diet-health relationships

Laurence Freedman, PhD, MA
Gerstner Institute for Epidemiology

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

Two main areas of interest

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

Objectives

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

Learning objectives
INTRODUCTION

The problem of measurement error when examining diet-health relationships

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
The problem of measurement error when examining diet-health relationships

### Introduction

Instruments (2)

<table>
<thead>
<tr>
<th>Please try to average your reported one of foods over the entire year. For example, if a food such as carrots is eaten 4 times a week during the approximate 3 months that it is in season, then the average seen would be once per week.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Results</strong></td>
</tr>
<tr>
<td><strong>Food</strong></td>
</tr>
<tr>
<td>Carrots (5)</td>
</tr>
<tr>
<td>Peppers (2) or red bell peppers (1)</td>
</tr>
<tr>
<td>Prunes or dried fruit (1) or 1 cup</td>
</tr>
<tr>
<td>Prunes or dried fruit (0.5 cup)</td>
</tr>
<tr>
<td>Baked (1)</td>
</tr>
<tr>
<td>Diced (1) or grated (1)</td>
</tr>
<tr>
<td>Orange (1)</td>
</tr>
<tr>
<td>Other (1) or other small glass</td>
</tr>
<tr>
<td>Rice, beans, pasta, or curd (1)</td>
</tr>
<tr>
<td>Blueberries, fresh, frozen, or canned (1)</td>
</tr>
<tr>
<td>Peaches, or prunes (1 or 1.5 cup canned)</td>
</tr>
<tr>
<td>Apricots, date, fruit (1 or 1.5 cup canned)</td>
</tr>
<tr>
<td><strong>Scale for usual intake per week</strong></td>
</tr>
<tr>
<td>&lt;0.5 portions</td>
</tr>
<tr>
<td>0.5-1 portion</td>
</tr>
<tr>
<td>1-2 portions</td>
</tr>
<tr>
<td>2-4 portions</td>
</tr>
<tr>
<td>&gt;4 portions</td>
</tr>
</tbody>
</table>
| **Additive systematic bias**: 
| $T_i$: true value, $R_i$: reported value |
| $R_i = \beta_0 + T_i$ |
| $R_i$ 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 $\beta_0$ from what they should be) |

### Types of measurement error

#### Types of measurement error (1)

- **Additive systematic bias**
  - Systematic: the instrument introduces a bias that is common to all individuals
  - $R_i = \beta_0 + T_i$

#### Types of measurement error (2)

- **Additive systematic bias**: $T$: true value, $R$: reported value

#### Types of measurement error (3)

- **Multiplicative and additive systematic bias**
  - $R_i = \beta_0 + \beta_1 T_i$
  - (In addition to additive bias, the true value is scaled up or down by factor $\beta_1$)
The problem of measurement error when examining diet-health relationships

Types of measurement error

1. Person-specific (random) bias
   - Bias that occurs at the individual level - it is specific to an individual but can differ among individuals
   - $R_i = \beta_0 + \beta_1 T_i + u_i$
   - The subject-specific bias $u_i$ is a random term
   - Its magnitude is quantified by $SD(u_i)$, its standard deviation

2. Within-person random error
   - Variation in reporting by an individual over a series of repeat reports
   - $R_{ij} = \beta_0 + \beta_1 T_i + u_i + \epsilon_{ij}$
   - The extra subscript $j$ denotes the sequence number in a series of reports
   - The extra term $\epsilon_{ij}$ is the within-person error that is on average zero
   - Its magnitude is quantified by $SD(\epsilon_{ij})$
The problem of measurement error when examining diet-health relationships

EVALUATING THE MAGNITUDE OF ERROR

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
  1. Unbiased
  2. Errors uncorrelated to true intake
  3. Errors uncorrelated with self-report errors

Evaluating the error (2)

- Do we have any ideal "reference" measures?
  1. Direct observation (feeding studies)
  2. "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
  3. "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)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Method</th>
<th>Energy (kcal/d)</th>
<th>Protein (g/d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>Marker</td>
<td>2842</td>
<td>105.5</td>
</tr>
<tr>
<td></td>
<td>FFQ</td>
<td>1961</td>
<td>74.7</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>2522</td>
<td>92.2</td>
</tr>
<tr>
<td>Women</td>
<td>FFQ</td>
<td>1524</td>
<td>57.2</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>1919</td>
<td>70.9</td>
</tr>
</tbody>
</table>

Evaluating the error? (5)

Results from OPEN – Protein Intake (after log transformation)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Method</th>
<th>Scaling Factor, $b_1$</th>
<th>Person-Specific Bias (SD)</th>
<th>Within-Person Error (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>FFQ</td>
<td>0.67</td>
<td>0.36</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>0.70</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td>Women</td>
<td>FFQ</td>
<td>0.65</td>
<td>0.33</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>0.60</td>
<td>0.16</td>
<td>0.35</td>
</tr>
</tbody>
</table>
The problem of measurement error when examining diet-health relationships

Evaluating the magnitude of error

Evaluating the error? (6)

Results from OPEN – Protein Density (after log transformation)

<table>
<thead>
<tr>
<th>Sex</th>
<th>Method</th>
<th>Scaling Factor, ( b_i )</th>
<th>Person-Specific Bias (SD)</th>
<th>Within-Person Error (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>FFQ</td>
<td>0.46</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>0.61</td>
<td>0.11</td>
<td>0.24</td>
</tr>
<tr>
<td>Women</td>
<td>FFQ</td>
<td>0.37</td>
<td>0.15</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>24HR</td>
<td>0.39</td>
<td>0.11</td>
<td>0.26</td>
</tr>
</tbody>
</table>

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:
  - Relative risk = \( \frac{\text{Prob (disease in exposed)}}{\text{Prob (disease in unexposed)}} \)
  - Odds (disease) = \( \frac{\text{Prob (disease)}}{1 - \text{Prob (disease)}} \)
  - 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
The problem of measurement error when examining diet-health relationships

For example, logistic regression:

\[
\log(\text{Odds}(Y = 1)) = \alpha_0 + \alpha_1 T_1 + \alpha_2 T_2 + \alpha_3 Z_1 + \ldots + \alpha_p Z_p,
\]

Where:
- \(Y\) is a binary variable;
- \(Y = 1\) denotes disease (“case”)
- \(Y = 0\) denotes no disease (“healthy”)

\(\alpha_i\)'s are the regression parameters and represent log odds ratios.

Each \(\alpha_i\) 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 (4)

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

\[
\log(\text{Odds}(Y = 1)) = \alpha_0 + \alpha_i \text{OC}
\]

Output from logistic regression program

Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Std. Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.65</td>
<td>0.046</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>oc6</td>
<td>-0.13</td>
<td>0.10</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Odds ratio estimate for OC = exp(\(\hat{\alpha}_1\)) = exp(-0.13) = 0.87

Estimating disease risk (6)

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?

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 (8)

Energy adjustment models

- There are several different methods for energy adjustment – we will look at two:
  i. Standard model
  ii. Density model
The problem of measurement error when examining diet–health relationships

Energy adjustment models

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

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

- Meaning of the coefficient \(\alpha_1\): 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

Models for estimating disease risk

**Estimating disease risk (10)**

<table>
<thead>
<tr>
<th>Value</th>
<th>Std. Error</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-1.39</td>
<td>0.13</td>
</tr>
<tr>
<td>afatcal</td>
<td>0.00093</td>
<td>0.00042</td>
</tr>
<tr>
<td>energy</td>
<td>0.00025</td>
<td>0.000098</td>
</tr>
</tbody>
</table>

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

**Estimating disease risk (11)**

**Density model**: Nutrient density = \(100 \times (\text{nutrient intake in kcal} / \text{total energy intake in kcal})\%\)

- Express the nutrient as a **nutrient density** and add total energy intake as a second explanatory variable. For example:

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

- Meaning of the coefficient \(\alpha_1\): The log odds ratio associated with increasing animal fat density by 1% while keeping total energy intake fixed

**Estimating disease risk (12)**

**Estimating disease risk (13)**

**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**
Log odds ratio estimates are scaled by \( R \). Then:

- Suppose we have an instrument with 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!

These conclusions seem 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.
The problem of measurement error when examining diet-health relationships

Introduction

Types of measurement error

Evaluating the magnitude of error

Models for estimating disease risk

Qualitative impact

Quantitative impact: 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: univariate models

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 = \( \lambda \times \text{true value} \),

  where
  - \( \lambda \) = attenuation factor
  - = slope of regression of T (truth) on R (report)
  - \( \lambda \) 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: univariate models

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)

<table>
<thead>
<tr>
<th></th>
<th>FFQ</th>
<th>24HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td>Protein</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>Protein Density</td>
<td>0.40</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Quantitative impact: univariate models

Quantitative impact of error (4)

Implications of these results

- Suppose the attenuation factor \( \lambda \) 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 \times 0.16 = 0.147
  - Expected estimated OR = \( \exp(0.147) = 1.16 \)

Quantitative impact: univariate models

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: univariate models

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

- Suppose 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 = ρ x 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: univariate models

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)

<table>
<thead>
<tr>
<th></th>
<th>FFQ</th>
<th>24HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.20</td>
<td>0.34</td>
</tr>
<tr>
<td>Protein</td>
<td>0.32</td>
<td>0.38</td>
</tr>
<tr>
<td>Protein Density</td>
<td>0.43</td>
<td>0.38</td>
</tr>
</tbody>
</table>

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: univariate models

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 ρ²

- 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)

<table>
<thead>
<tr>
<th></th>
<th>FFQ</th>
<th>24HR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.20</td>
<td>0.34</td>
</tr>
<tr>
<td>Protein</td>
<td>0.32</td>
<td>0.38</td>
</tr>
<tr>
<td>Protein Density</td>
<td>0.43</td>
<td>0.38</td>
</tr>
</tbody>
</table>

- Example: Protein Density
  - FFQ: Effective sample size = 0.43² x actual sample size = 0.18 x actual sample size
  - We effectively lose 82% of our sample size!
The problem of measurement error when examining diet-health relationships

### 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 at the 5% significance level, assuming that we could measure dietary intake exactly.

- Then, because of the measurement error we would need 
  \[ \frac{50,000}{\sqrt{0.432^2}} = \frac{50,000}{0.432} = 270,000 \]
  to preserve the power of 90%.

### Quantitative impact: univariate models

#### Quantitative impact of error (12)

- **Decrease in study power**

#### 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:
    \[ \text{Power} = \Phi^{-1}(3.24 \cdot 1.96) \]
  - Where the symbol \( \Phi^{-1} \) denotes the inverse of the standard normal cumulative distribution function.

### Quantitative impact: multivariate models

#### Quantitative impact: multivariate exposures (1)

- Two or more dietary variables in the disease regression model
  - Typical example: Standard energy-adjustment model
    \[ \log\{\text{Odds}(Y=1)\} = \alpha_0 + \alpha_1 \text{fatcal} + \alpha_2 \text{energy} \]
  - 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:
  \[ \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{pmatrix} \]
  - If the true log odds ratios for the two nutrients are \( \alpha_1 \) and \( \alpha_2 \), 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 \( \lambda_{12} \) and \( \lambda_{21} \) tell us how serious is the residual confounding.
  - We call them contamination factors.
If $\lambda_{12}$ and $\lambda_{21}$ are small, then the only bias in the estimated log odds ratios comes essentially from attenuation, then:

- The estimated log odds ratio is attenuated
- 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

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

<table>
<thead>
<tr>
<th>Dietary Component</th>
<th>Gender</th>
<th>Protein Density</th>
<th>Potassium Density</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Man</td>
<td>-0.01 (0.03)</td>
<td>0.13 (0.05)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Woman</td>
<td>0.03 (0.05)</td>
<td>0.10 (0.06)</td>
<td>-</td>
</tr>
<tr>
<td>Protein Density</td>
<td>Man</td>
<td>-</td>
<td>-0.01 (0.09)</td>
<td>0.08 (0.05)</td>
</tr>
<tr>
<td></td>
<td>Woman</td>
<td>-0.00 (0.10)</td>
<td>0.06 (0.05)</td>
<td>-</td>
</tr>
<tr>
<td>Potass. Density</td>
<td>Man</td>
<td>-0.05 (0.06)</td>
<td>-</td>
<td>0.04 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Woman</td>
<td>0.00 (0.07)</td>
<td>-</td>
<td>-0.04 (0.05)</td>
</tr>
<tr>
<td>Total Fat Density</td>
<td>Man</td>
<td>-0.03 (0.07)</td>
<td>0.00 (0.06)</td>
<td>0.05 (0.05)</td>
</tr>
<tr>
<td></td>
<td>Woman</td>
<td>-0.02 (0.08)</td>
<td>-0.06 (0.10)</td>
<td>-0.07 (0.05)</td>
</tr>
<tr>
<td>Sat. Fat Density</td>
<td>Man</td>
<td>-0.03 (0.05)</td>
<td>-0.04 (0.07)</td>
<td>0.10 (0.04)</td>
</tr>
<tr>
<td></td>
<td>Woman</td>
<td>-0.01 (0.06)</td>
<td>-0.07 (0.08)</td>
<td>-0.02 (0.04)</td>
</tr>
</tbody>
</table>

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

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
The problem of measurement error when examining diet-health relationships

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

Douglas Midthune
National Cancer Institute