Measurement Error in Nutritional Epidemiology – in Memory of Arthur Schatzkin

Laurence Freedman
Gertner Institute for Epidemiology and Health Policy Research
Israel
Arthur Schatzkin, MD, Dr. P.H.
1948-2011
Employed by US National Cancer Institute from 1988-2011
Some of Arthur’s Attributes

1. An intense belief in social justice
2. A burning desire to improve human existence
3. Intellectual curiosity
4. Intellectual honesty
5. A great sense of humor
6. A quick temper
7. A drive to achieve progress
8. A genuine care for his fellow beings
Arthur’s Interests

1. The role of diet in health

2. Epidemiological methods of studying diet-disease relationships

3. The statistical concepts that underlie the pros and cons of an epidemiologic method

4. How to improve methodology through an interplay of biological and statistical reasoning

5. Advancing diet and physical activity data collection tools through new technology
Designs for Studying Diet-Health Relationships

Main study designs have been:

- Nutritional case-control studies
- Nutritional cohort studies
- Randomized dietary intervention trials
Pros and Cons of These Designs

- **Case-control studies:**
  Inexpensive, but results subject to confounding and several biases such as recall and selection bias

- **Cohort studies:**
  More expensive, and results subject to confounding but not to recall or selection bias
  *(NIH-AARP Diet and Health Study)*

- **Randomized studies:**
  Extremely expensive, but results not subject to confounding
  *(Polyp Prevention Trial – using precursor lesion to reduce the size and shorten the duration)*
Cohort Study Design

Example:
NIH-AARP Diet and Health Study – breast cancer outcome

188,736 women volunteers completed a food frequency questionnaire

Follow-Up for breast cancer

3501 breast cancer diagnoses

Disease-free
Food Frequency Questionnaires

Advantages
- Attempts to be comprehensive
- Inexpensive to administer and use (automated coding)
- Requires single administration over a long time period

Disadvantages
- Cognitively difficult (summarizing long-term intake)
- Conversion to nutrients is difficult
- Limited food list, information on portion size, food preparation

These difficulties lead to measurement error
How to weigh yourself and get the most accurate result. I can't believe I have been doing it wrong all these years!

We must get the word out!
“I’m going to order a broiled skinless chicken breast, but I want you to bring me lasagna and garlic bread by mistake.”
Effects of Dietary Measurement Error

- Downward bias in the estimated diet-disease relationship (attenuation)
- Reduced statistical power to detect the relationship
- Potential invalidity of statistical tests in multiple regression models
Attenuation of the Diet-Disease Relationship
Bias in estimate of relative risk (RR)
Dietary measurement error causes that, on average:

$$\text{Estimated log RR} = \lambda \times \text{True log RR}$$

where usually $\lambda < 1$

We call $\lambda$ the attenuation coefficient

The smaller is $\lambda$, the worse the attenuation
Reduction in Statistical Power

Measurement error reduces the effective sample size from $n$ to $\rho^2 \times n$,

where $\rho$ is the correlation coefficient between reported and true dietary intake

The smaller is $\rho$, the greater the loss of power
How serious are these effects?
Observing Protein & Energy Nutrition (OPEN)

- The first large validation study of dietary assessment instruments with recovery biomarkers
- Conducted by the National Cancer Institute, 1999-2000
- Co-PI’s: Arthur Schatzkin and Amy Subar
- 261 men, 223 women
- Dietary instruments:
  - Food Frequency Questionnaire (twice)
  - 24-hour Recall (twice)
- Recovery Biomarkers:
  - Doubly Labeled Water (for Energy)
  - Urinary Nitrogen (for Protein)
  - Urinary Potassium (for Potassium)
How Serious are These Effects?

Biased Estimation

FFQ attenuation factors, $\lambda$, for selected nutrients (OPEN)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>Protein</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>Protein Density</td>
<td>0.40</td>
<td>0.32</td>
</tr>
</tbody>
</table>

**Note:** The attenuation improves after adjustment for energy
**How Serious are These Effects?**

**Biased Estimation**

<table>
<thead>
<tr>
<th>Attenuation Factor</th>
<th>True Relative Risk</th>
<th>Estimated Relative Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>2.0</td>
<td>1.04</td>
</tr>
<tr>
<td>0.15</td>
<td>2.0</td>
<td>1.11</td>
</tr>
<tr>
<td>0.40</td>
<td>2.0</td>
<td>1.32</td>
</tr>
</tbody>
</table>
# How Serious are These Effects?

**Reduced Power**

**FFQ correlation coefficients, \( \rho \), for selected nutrients (OPEN)**

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>Protein</td>
<td>0.32</td>
<td>0.30</td>
</tr>
<tr>
<td>Protein Density</td>
<td>0.43</td>
<td>0.35</td>
</tr>
</tbody>
</table>
How Serious are These Effects?

Reduced Power

Required sample size if 10,000 are needed when there is no measurement error

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>250,000</td>
<td>1,000,000</td>
</tr>
<tr>
<td>Protein</td>
<td>100,000</td>
<td>110,000</td>
</tr>
<tr>
<td>Protein Density</td>
<td>54,000</td>
<td>84,000</td>
</tr>
</tbody>
</table>
What Can We Do About These Problems?

1. Proper analysis and reporting in the current situation:
   (a) Gather more information about measurement error properties of current instruments
      >> Validation Studies Pooling Project
   (b) Develop guidelines on analysis and reporting
      >> Commentary in the Journal of the National Cancer Institute

2. Ways to improve the measurement of dietary intake:
   (a) Better self-report instruments
   (b) Combining self-report instruments
   (c) Combining self-reports and biomarkers
Validation Studies Pooling Project

**Aims:** To combine data from four large validation studies with recovery biomarkers to answer questions regarding measurement error properties of FFQ’s and 24HR’s.

**Studies (PI’s):**
- OPEN (Arthur Schatzkin and Amy Subar, NCI)
- USDA - AMPM (Alanna Moshfegh, USDA; with Walter Willett, Harvard – FFQ add-on)
- NBS (Ross Prentice and Marian Neuhouser, WHI)
- Energetics (Lenore Arab, UCLA)

To be added: NPAAS (Ross, Prentice, WHI)
## Validation Studies Pooling Project

### Study Populations

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Male</th>
<th>Non-Hispanic White</th>
<th>Age &lt;40y</th>
<th>Age 60y</th>
<th>BMI ≥30</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN</td>
<td>484</td>
<td>54%</td>
<td>83%</td>
<td>0%</td>
<td>24%</td>
<td>29%</td>
</tr>
<tr>
<td>USDA</td>
<td>524</td>
<td>50%</td>
<td>77%</td>
<td>22%</td>
<td>20%</td>
<td>24%</td>
</tr>
<tr>
<td>NBS</td>
<td>544</td>
<td>0%</td>
<td>82%</td>
<td>0%</td>
<td>97%</td>
<td>33%</td>
</tr>
<tr>
<td>Energetics</td>
<td>265</td>
<td>36%</td>
<td>49%</td>
<td>57%</td>
<td>3%</td>
<td>25%</td>
</tr>
</tbody>
</table>
### FFQ Attenuation Factors for Females

<table>
<thead>
<tr>
<th>Study</th>
<th>Energy</th>
<th>Protein</th>
<th>Protein Density</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN</td>
<td>0.04</td>
<td>0.13</td>
<td>0.32</td>
</tr>
<tr>
<td>USDA</td>
<td>0.05</td>
<td>0.17</td>
<td>0.42</td>
</tr>
<tr>
<td>NBS</td>
<td>0.05</td>
<td>0.22</td>
<td>0.41</td>
</tr>
<tr>
<td>Energetics</td>
<td>0.11</td>
<td>0.05</td>
<td>0.45</td>
</tr>
</tbody>
</table>

More on this in Wednesday’s afternoon session
Developing Guidelines for Analysis and Reporting

**Aims:** To make investigators aware of the problem of dietary measurement error and provide reporting standards that address this problem


**Dealing With Dietary Measurement Error in Nutritional Cohort Studies**

Laurence S. Freedman, Arthur Schatzkin†, Douglas Midthune, Victor Kipnis
Developing Guidelines for Analysis and Reporting

An important recommendation:

Use energy adjustment for relative risk estimation

The improvements in attenuation factors and correlation coefficients that we see for protein density in validation studies with recovery biomarkers confirms previous recommendations.
Developing Guidelines for Analysis and Reporting

Two other practical questions:

1. Should relative risk estimates adjusted for measurement error be presented?

2. How should we test hypotheses of no association between diet and disease?
Developing Guidelines for Analysis and Reporting

Should relative risk estimates adjusted for measurement error be presented?

Remember that, on average:

\[ \text{Estimated log RR} = \lambda \times \text{True log RR} \]

So to estimate True log RR we can take:

\[ \text{Estimated log RR} / \lambda \]

To do this we have to estimate \( \lambda \)
Should Relative Risk Estimates Adjusted for Measurement Error be Presented?

- To estimate $\lambda$ we need a validation study in which we compare a reference instrument determination of intake with the FFQ determination.

- When the reference instrument is a recovery biomarker, we get an unbiased estimate of $\lambda$.

- But recovery biomarkers are not available for most nutrients and foods. So we have to use instead a detailed self-report instrument such as a 24h recall or food record.

Question: Is the estimate of $\lambda$ that we get using a 24h recall or food record good enough?
Should Relative Risk Estimates Adjusted for Measurement Error be Presented?

Estimated attenuation factors using recovery biomarker or 24HR as the reference (OPEN)

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Gender</th>
<th>Recovery biomarker-based</th>
<th>24HR-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>Men</td>
<td>0.07</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.04</td>
<td>0.13</td>
</tr>
<tr>
<td>Protein Density</td>
<td>Men</td>
<td>0.45</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.35</td>
<td>0.44</td>
</tr>
<tr>
<td>Potassium Density</td>
<td>Men</td>
<td>0.59</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.59</td>
<td>0.56</td>
</tr>
</tbody>
</table>
Should Relative Risk Estimates Adjusted for Measurement Error be Presented?

**Conclusion**
- For energy-adjusted analyses the estimate of the attenuation factor using a validation study with a detailed self-report instrument is likely to be sufficiently good.

- Adjusting the relative risk estimate in this way is better than no adjustment.

- Therefore, it is recommended to present measurement error adjusted estimates of relative risk.
How Should We Test Hypotheses of No Association Between Diet and Disease?

Conventional statistical wisdom

- **Univariate disease model:**
  One error-prone (dietary) exposure
  Although the relative risk estimate is attenuated
  the usual test of the null hypothesis is valid

- **Multivariate disease model:**
  Several error-prone (dietary) exposures
  Relative risk estimates may be attenuated or inflated or even change sign
  The usual significance test could be invalid
How Should We Test Hypotheses of No Association between Diet and Disease?

- This problem arises 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!
Residual Confounding

- The fraction of the other factor’s effect that is adopted has been termed the contamination factor.

- If the contamination factor is small, then the only bias in the estimated log odds ratios comes from attenuation, and:
  
  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.
How Should We Test Hypotheses of No Association between Diet and Disease?

**OPEN – Estimated Contamination Factors**
(Freedman, Schatzkin, Midthune, and Kipnis. J Nat Cancer Inst 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>Energy</td>
<td>Men</td>
<td>-0.01</td>
<td>0.13</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.03</td>
<td>0.10</td>
<td>-</td>
</tr>
<tr>
<td>Protein Density</td>
<td>Men</td>
<td>-</td>
<td>-0.01</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>-</td>
<td>0.00</td>
<td>0.06</td>
</tr>
<tr>
<td>Potass. Density</td>
<td>Men</td>
<td>-0.05</td>
<td>-</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>0.00</td>
<td>-</td>
<td>-0.04</td>
</tr>
<tr>
<td>Total Fat Density</td>
<td>Men</td>
<td>-0.03</td>
<td>0.00</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>-0.02</td>
<td>-0.08</td>
<td>-0.07</td>
</tr>
<tr>
<td>Sat. Fat Density</td>
<td>Men</td>
<td>-0.03</td>
<td>-0.04</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Women</td>
<td>-0.01</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
</tbody>
</table>
How Should We Test Hypotheses of No Association between Diet and Disease?

Conclusion

The contamination (residual confounding) that occurs with two or more dietary components measured by a FFQ is probably small enough to ignore.

Therefore, it is recommended to conduct hypothesis tests of no association between diet and disease in the usual manner without any adjustment for measurement error.

Note: This conclusion may not hold for diet and physical activity measures included in the same model.
What Can We Do about These Problems?

1. Proper analysis and reporting in the current situation:
   (a) Gather more information about measurement error properties of current instruments
   (b) Develop guidelines on analysis and reporting

2. Ways to improve the measurement of dietary intake:
   (a) Better self-report instruments
   (b) Combining self-report instruments
   (c) Combining self-reports and biomarkers
What Can We Do about These Problems?

Ways to improve the measurement of dietary intake:

*Review*

Observational Epidemiologic Studies of Nutrition and Cancer: The Next Generation (with Better Observation)

Arthur Schatzkin,1 Amy F. Subar,2 Steven Moore,1 Yikyung Park,1 Nancy Potischman,2 Frances E. Thompson,2 Michael Leitzmann,1 Albert Hollenbeck,4 Kerry Grace Morrissey,5 and Victor Kipnis3

What Can We Do about These Problems?

Ways to improve the measurement of dietary intake:

(a) Better self-report instruments
   e.g., automated 24h recalls (such as ASA-24), mobile phone diet records

(b) Combining self-report instruments
   e.g., Carroll RJ et al. Taking advantage of the strengths of two different dietary instruments to improve intake estimates for nutritional epidemiology. Am J Epidemiol, 2012

(c) Combining self-reports and biomarkers
   e.g., Freedman LS et al. Using regression calibration equations that combine self-reported intake and biomarker measures to obtain unbiased estimates and more powerful tests of dietary associations. Am J Epidemiol, 2011
Rabbi Tarphon said:
It is not your duty to complete the task, but you are not free to exempt yourself from it.

*Ethics of the Fathers, Chapter 2, Mishna 21.*