Nutrition Biomarkers in Chronic Disease Prevention Research

Laurence Freedman
Gertner Institute for Epidemiology, Tel Hashomer, Israel
Acknowledgments

- Statisticians:
  - Victor Kipnis
  - Raymond Carroll
  - Douglas Midthune
  - Kevin Dodd
  - Janet Tooze

- Epidemiologists/Nutritionists:
  - Arthur Schatzkin
  - Amy Subar
  - Nancy Potischman
  - Natasa Tasevska
  - Julie Mares
  - Lesley Tinker
Content

• Brief overview
• Effects of measurement error in self-report instruments
• Validation studies with recovery biomarkers and their implications
• Combining self-reports and biomarkers
Background I

- Interest in measuring dietary intake and relating this to health outcomes
- Main studies in chronic disease prevention have been:
  - Case-control studies
  - Cohort studies
  - Randomized dietary intervention trials
The role of dietary intake measurement is very different in these studies:

- **Case-control studies & Cohort studies:**
  Primary role, since it is the main exposure measure

- **Randomized dietary intervention trials**
  Secondary role, since it is a measure of compliance and a potential mediator
Background III

- Dietary intake has mainly been measured through self-report instruments e.g., Food Frequency Questionnaire (FFQ)
- Inconsistent results across studies, and failure to find evidence for apparently strong hypotheses (e.g. fat and breast cancer)
- Consequent doubts about the accuracy of measurements from such instruments
- Search for biomarkers of dietary intake
Background III

Role of nutrition biomarkers

(i) Assessment of compliance in randomized trials

(ii) Mediation analyses in randomized trials

(iii) Validation of self-report instruments for cohort studies

(iv) Adjusting for the bias in estimated risks in cohort studies

(v) Recovering lost power in cohort studies
Types of nutrition biomarker

- Recovery biomarkers

Give an essentially unbiased estimate of intake over a given period:

(i) Doubly labelled water – energy intake
(iii) 24 hour urinary nitrogen – protein intake
(iv) 24 hour urinary potassium – potassium intake

Useful for validation of self-report instruments.
Very expensive or difficult to collect.
Types of nutrition biomarker II

- **Concentration biomarkers**

  Subject to complex metabolic pathways in their regulation: correlated to intake but not an unbiased measure:
  
  e.g., serum carotenoids, lipids, vitamins, etc.

  Less useful for validation, but could be useful for prediction of some dietary intakes.
  Often less expensive and easier to collect.
**Dietary measurement error I**

**Setting:** Cohort study of diet and disease

**Exposure:** Usual dietary intake, $X$

**Outcome:** Disease (often quite rare), $Y$

**Interest:** To estimate the relationship between $X$ and $Y$

$$h(E(Y)) = \alpha_0 + \alpha_1 X$$

**Problem:** We observe not $X$, but a self-report, $W$, that has some error $\delta$:

$$W = X + \delta$$
Dietary measurement error II

The measurement error in W causes two major problems:

- **Bias in the estimated relationship**
- **Loss of statistical power to detect the relationship**
Example: classical measurement error model

\[ W = X + \delta \]
\[ E(\delta) = 0 \]
\[ \delta \perp X \]

Then,

\[ E(\hat{\alpha}_{1w}) = \lambda \alpha_1 \]

where

\[ \lambda = \text{var}(X) / \text{var}(W) < 1 \]

so the estimated coefficient is attenuated.
Dietary measurement error IV

Statistical power

The effective sample size is reduced from $n$ to $\rho^2 n$, where $\rho$ is the correlation between $W$ and $X$.

Validation studies with recovery biomarkers allow us to estimate $\lambda$ and $\rho$, and thereby gauge how serious is the problem.
The first large validation study with recovery biomarkers

The OPEN Study

- Conducted by the National Cancer Institute, 1999-2000
- 261 men, 223 women
- Dietary instruments: 24HR (twice), FFQ (twice)

- Biomarkers: Doubly Labeled Water (for Energy)
  Urinary Nitrogen (for Protein)
  Urinary Potassium (for Potassium)

- These biomarkers have been shown in previous studies to give unbiased measures of these intakes
How serious are the problems? I

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 problems? II

Average estimated RR when true RR = 2:

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>Protein</td>
<td>1.12</td>
<td>1.10</td>
</tr>
<tr>
<td>Protein Density</td>
<td>1.32</td>
<td>1.25</td>
</tr>
</tbody>
</table>

Note:

It is generally thought that uncontrolled confounding precludes reliably detecting RR < 1.25 in a cohort study.
FFQ correlations with true usual intake, $\rho$, for selected nutrients (OPEN):

<table>
<thead>
<tr>
<th>Nutrient</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>0.08</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>
1. Even after adjustment for energy there is serious attenuation of estimated RRs due to the measurement error in a FFQ.

2. There is also serious loss of power to detect diet-disease relationships.

Caveat: OPEN and similar studies can examine only protein, energy and potassium. We can only extrapolate to other nutrients and foods.
Traditional remedy to the problem of RR attenuation

Regression Calibration:
\( Y = \text{disease}; \ X = \text{true dietary intake}; \ W = \text{self-reported intake} \)

Instead of regressing \( Y \) on \( W \) (leading to bias) we can regress \( Y \) on \( E(X|W) \).

(How we determine \( E(X|W) \) is another story!)

This leads to (nearly) unbiased estimates, but no gain in power
Remedy to these problems using (concentration) biomarkers

Enhanced Regression Calibration:

We also have a biomarker $M$ for $X$;

So regress $Y$ on $E(X|W,M)$ instead of $E(X|W)$

Under certain circumstances, we can obtain unbiased estimates and gains in power.

Predicted sample size reduction is by factor:

$$\frac{\text{var}(E(X|W))}{\text{var}(E(X|W,M))}$$
Remedy to these problems using markers

**Example:**
Carotenoids and Eye Disease Study (CAREDS)
Relation between dietary lutein/zeaxanthin and eye cataracts

- $X = \log (\text{true usual lutein/zeaxanthin intake})$
- Self-report instrument, $W = \log (\text{FFQ})$
- Biomarker, $M = \log (\text{serum lutein/zeaxanthin})$
- Outcome, $Y = \text{Eye cataracts (yes/no)}$

- Analysis = logistic regression of $Y$ on explanatory variables
  
  Models: 1: $W$
  2: $E(X|W)$
  3: $E(X|W, M)$
Remedy to these problems using markers

Example:

To evaluate $E(X|W)$ and $E(X|W, M)$, we need a measurement error model relating these measures to true intake.

We developed such a model based on feeding studies and validation studies reported in the literature (Freedman et al, Epidemiol Persp Innov, 2010):

$$W = 0.35 + 0.71 X + \varepsilon_W; \quad \text{var}(\varepsilon_W) = 0.36$$

$$M = 5.29 + 0.60 X + \varepsilon_M; \quad \text{var}(\varepsilon_M) = 0.15$$

$$\text{var}(X) = 0.25$$

$$\frac{\text{var}(E(X|W))}{\text{var}(E(X|W,M))} = 0.53$$
Combining self-report with biomarker

Carotenoids in Eye Disease Study (WHI)
Analysis of relationship between dietary lutein and eye cataracts

<table>
<thead>
<tr>
<th>Model</th>
<th>log OR*</th>
<th>(SE)</th>
<th>z-value</th>
<th>Sample size ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>W=log(FFQ)</td>
<td>-0.165</td>
<td>(0.080)</td>
<td>-2.07</td>
<td>1.00</td>
</tr>
<tr>
<td>E(X</td>
<td>W)</td>
<td>-0.464</td>
<td>(0.225)</td>
<td>-2.07</td>
</tr>
<tr>
<td>E(X</td>
<td>W,M)</td>
<td>-0.506</td>
<td>(0.161)</td>
<td>-3.15</td>
</tr>
</tbody>
</table>

* Adjusted for age and smoking
Combining self-report with biomarker

Caveats

1. Using $E(X|W,M)$, although the **Wald test is valid**, the log OR estimate (-0.506) is **biased** (inflated). However, there is a way to obtain an unbiased estimate.

2. There is always the lurking danger of unknown **confounders** that are involved in the complex metabolic pathways that determine the biomarker level and are also associated with the disease.
1. Prentice et al (Stat Biosci, 2009)
   - Large feeding study (150 women)
   - Provided with a personalized diet for several weeks
   - Measure many biomarkers in blood and urine at end of period
   - Develop regression prediction equations for true intakes based on the array of biomarkers
   - Use these in place of reported intake

2. Challenge: to incorporate self-report into these prediction equations
1. Dietary biomarkers have proved extraordinarily useful in quantifying the extent of the problems caused by measurement error in self-reports.

2. The time has now come to invest in developing their use to solve the problems.

3. This development should go hand-in-hand with attempts to improve self-report instruments for cohort studies.