Measurement Error in Nutritional Epidemiology: Impact, Current Practice for Analysis, and Opportunities for Improvement

Pamela Shaw
shawp@upenn.edu
on behalf of STRATOS TG4

10 May, 2017
9th EMR-IBS, Thessaloniki
Outline

• Background
• Motivating examples showing impact of Measurement error
• Regression calibration
• Literature survey: methodology and results
• Conclusions
STRATOS TG4: Measurement Error and Misclassification

MEMBERSHIP

• Laurence Freedman, Gertner/IMS, Co-Chair
• Victor Kipnis, NCI, Co-Chair
• Raymond Carroll, Texas A&M U
• Veronika Deffner, Munich, LMU
• Kevin Dodd, NCI
• Paul Gustafson, U. British Columbia
• Ruth Keogh, London School of Hygiene
• Helmut Kuechenhoff, Munich, LMU
• Pamela Shaw, U. Pennsylvania
• Janet Tooze, Wake Forest School of Medicine
TG4 Projects

1. Literature Survey for how measurement error is addressed in 4 types of epidemiological studies

2. Guidance paper for nutritional epidemiologists

3. Guidance paper for biostatisticians
TG4 Literature Survey

• There have been many statistical advances to address in measurement error in the past few decades

• TG4 was interested in assessing the current practice for acknowledging and addressing measurement error in epidemiologic/observational studies
  – Want to identify knowledge gaps and opportunities for improvement

• We conducted a literature survey focused on types of epidemiologic studies with exposures that are well known to be subject to measurement error
Example 1: Classical Measurement error

- Classical measurement error (CME) is random, mean zero error
- Covariate $X^*$ with CME can be written as: $X^* = X + u$, where $u$ is mean 0 error term independent of $X$ and $Y$
- Suppose $Y = \beta_0 + \beta_1 X + \varepsilon$, then regressing $Y$ on $X^*$ will estimate slope $\beta_1^* \neq \beta_1$
- $\beta_1^*$ will be attenuated toward 0

\[ \beta_1^* = \lambda \beta \] , where
\[ \lambda = \frac{\text{var}(X)}{\text{var}(X) + \text{var}(u)} \]
So $0 < \lambda < 1$
Example 2: Measuring Dietary Intake

• Measuring dietary intake is of interest in epidemiology as there are a number of diseases for which dietary factors are thought to be important risk factors, including cancer, heart disease and diabetes.

• Dietary intake is a complex exposure to measure
  – Made up of many nutrients obtained from a variety of foods
  – Contains day-to-day variability, possibly also temporal variability

• There are several prevailing dietary assessment methods
  – Self-report: Food frequency questionnaire, 24hour recall, daily food record
  – Objective biomarkers: recovery or concentration markers
Measuring Energy Intake

Biomarker Energy

Visit 2

Visit 1

\( \rho = .72 \)

FFQ Energy

Visit 2

Visit 1

\( \rho = .70 \)
## Energy Intake vs Body Mass Index

Neuhouser et al AJE 2008

### APPENDIX TABLE.

Estimates of energy intake (kcal/day) obtained by self-reported food frequency questionnaire, a biomarker (total energy expenditure), and a calibrated food frequency questionnaire, according to body mass index category, Women’s Health Initiative Nutritional Biomarkers Study, 2004–2005*

<table>
<thead>
<tr>
<th>Body mass index† category</th>
<th>Self-reported FFQ‡</th>
<th>Total energy expenditure</th>
<th>Calibrated FFQ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Geometric mean</td>
<td>IQR‡</td>
<td>Geometric mean</td>
</tr>
<tr>
<td>Normal (&lt;25.0)</td>
<td>1,407</td>
<td>1,157–1,759</td>
<td>1,894</td>
</tr>
<tr>
<td>Overweight (25.0–29.9)</td>
<td>1,462</td>
<td>1,196–1,837</td>
<td>2,043</td>
</tr>
<tr>
<td>Obese (≥30)</td>
<td>1,454</td>
<td>1,161–1,897</td>
<td>2,213</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2,247</td>
</tr>
</tbody>
</table>

* Note that the difference between FFQ energy intake (self-report) and total energy expenditure (biomarker) increases as body mass index increases. The biomarker-calibrated estimates, for the same women, correct for the measurement error using the model shown in table 4.

† Weight (kg)/height (m)².

‡ FFQ, food frequency questionnaire; IQR, interquartile range (25th–75th percentiles).
Regression Calibration:  
A simple approach to adjust for ME

Prentice Biometrika 1982

• Suppose true intake: \( X \)
• Error-prone measure: \( X^* \) (FFQ intake)
• Objective biomarker: \( X^{**} = X + u \)
• Predicted \( X = E(X^{**} | X^*, Z) = E(X+u | X^*, Z) = E(X | X^*, Z) \)
  \[= a_1 + a_2 X^* + a_3 Z + a_4 Z X^*\]

Regression calibration: Regress outcome \( Y \) on predicted intake, other covariates \( Z \)
HR for Uncalibrated vs Calibrated Energy Intake
Prentice, Shaw et al AJE 2009
Survey Areas

Each of four topic areas had its own literature search

• Nutritional intake cohort studies  (Pamela Shaw/Ruth Keogh)
• Dietary intake population surveys  (Kevin Dodd)
• Physical activity cohort studies (Janet Tooze)
• Air pollution cohort studies (Veronika Deffner/Helmut Kuechenhoff)
Overall Approach

• Focused on error-prone variable as exposure in analysis

• For cohort studies, literature search done in two stages
  – **Search A**: Survey recent articles to assess how often articles acknowledged and/or addressed measurement error
  – **Search B**: Survey recent articles that adjusted for measurement error to describe methods in current practice

• Questionnaires filled out for each reviewed article

• Excluded reviews, cross-sectional studies, case-control studies and meta-analyses

• Each topic area conducted a quality control review
  – 20% re-reviewed by independent reviewer
Nutritional Epidemiology
Cohort Studies: Survey Methodology

• Date Range A: Feb 2014-Jun 2015; B: Jan 2001-Jul 2015

• Limited search to three common diseases with dietary risk factors: cancer, heart disease and diabetes
  – Restricted date range to find about 50 articles from Search A and 30 articles from Search B

• Search B: added (measurement error OR misclassification to Search A
  – Not many articles, so did additional key word searches including: (measurement error OR misclassification) AND nutritional epidemiology

• Physical activity and pollution cohort methodology similar, except relied on date range and random sampling to reduce number of articles reviewed
## Number of Articles Reviewed*

<table>
<thead>
<tr>
<th>Category</th>
<th>Search A</th>
<th>Search B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nutritional Epidemiology cohort studies</td>
<td>51</td>
<td>27</td>
</tr>
<tr>
<td>Dietary Intake Population Survey</td>
<td>67</td>
<td>N/A</td>
</tr>
<tr>
<td>Physical Activity cohort studies</td>
<td>30</td>
<td>40</td>
</tr>
<tr>
<td>Air Pollution cohort studies</td>
<td>50</td>
<td>25</td>
</tr>
</tbody>
</table>

* Number in table excludes articles that were identified by search terms but upon closer examination did not meet inclusion criteria
### Search A Survey Results

<table>
<thead>
<tr>
<th></th>
<th>Nutritional Epi Cohort N= 51</th>
<th>Phys activity Cohort N=30</th>
<th>Diet Intake Survey N=67</th>
<th>Pollution Cohort N=50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mention ME as potential problem n(%)</strong></td>
<td>48 (94%)</td>
<td>17 (57%)</td>
<td>53/67 (79%)</td>
<td>20 (40%)</td>
</tr>
<tr>
<td><strong>Used a method to adjust for ME N (%)</strong></td>
<td>5 (10%)</td>
<td>0 (0%)</td>
<td>19/67 (28%)</td>
<td>3 (6%)</td>
</tr>
<tr>
<td><strong>% categorizing exposure</strong></td>
<td>Any 50/51(98%)</td>
<td>Primary exposure</td>
<td>Any 53/67 (79%)</td>
<td>Any 20 (40%)</td>
</tr>
<tr>
<td></td>
<td>Exclusively 27/51 (53%)</td>
<td></td>
<td>Exclusively 19/67 (28%)</td>
<td>Exclusively 3 (6%)</td>
</tr>
<tr>
<td><strong>Statistic of main interest N (%)</strong></td>
<td>HR 45 (88%)</td>
<td>HR 11 (37%)</td>
<td>Mean 51 (76%)</td>
<td>HR 11 (37%)</td>
</tr>
<tr>
<td></td>
<td>OR 3 (6%)</td>
<td>OR/RR 9(30%)</td>
<td>Median 28(42%)</td>
<td>OR 3 (6%)</td>
</tr>
<tr>
<td></td>
<td>RR 2 (4%)</td>
<td>GLM 5 (17%)</td>
<td>%-tiles 21(31%)</td>
<td>RR 2 (4%)</td>
</tr>
<tr>
<td></td>
<td>Slope 5(10%)</td>
<td>Other 5 (17%)</td>
<td>Quality 31(46%)</td>
<td>Slope 5(10%)</td>
</tr>
</tbody>
</table>
## Methods to Address Measurement Error

<table>
<thead>
<tr>
<th>Nutritional Epi Cohort</th>
<th>Phys Activity Cohort</th>
<th>Dietary Intake Pop. Survey</th>
<th>Pollution Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>N= 27*</td>
<td>N=40</td>
<td>N=67</td>
<td>N= 25</td>
</tr>
<tr>
<td>Regression Calib. 26 (96%)</td>
<td>Regression Calib. 1 (50%)</td>
<td>NCI 10(53%)</td>
<td>Sens Analysis 4 (80%)</td>
</tr>
<tr>
<td>SIMEX 1 (4%)</td>
<td>Other 1 (50%)</td>
<td>Means 7(37%)</td>
<td>Instr Variables 1 (20%)</td>
</tr>
<tr>
<td>Other 1 (4%)</td>
<td>Other 1 (50%)</td>
<td>ISU 1 (5%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSM 1 (5%)</td>
<td></td>
</tr>
<tr>
<td>Search A: None 90%</td>
<td>Search A: None 95%</td>
<td>Search A: None 72%</td>
<td>Search A: None 94%</td>
</tr>
</tbody>
</table>

- Number excludes articles that were identified by search terms but upon review did not use a method to correct for error.
- Row percents do not add to 100% due to use of multiple methods.
Other Observations from Diet and Physical Activity Cohort Surveys

• Common in the cohort studies to have multiple covariates with error: eg diet + physical activity, smoking, and/or alcohol intake
  – Many adjust for both diet+ PA, only 1 article adjusted for error in both physical activity (Zhang et al, AJE 2014)
  – Errors in smoking/alcohol not addressed

• Most categorized the continuous exposures
  – Impacts of categorizing an exposure subject to error are ignored
  – Common belief: categorization will lower impact of measurement error in the analysis

• Most people who mentioned error as a problem made an incomplete/incorrect claim
  – Many only mentioned attenuation in found associations
  – Some claimed no bias in associations since prospective subject recall
  – Some claimed no bias since instrument was validated
Other observations from Dietary Intake Population Surveys

• Most studies (80%) used 24HR as primary instrument
  – 31/53 used only 1 24HR, rest had repeats on at least a subsample
  – 8/31 (26%) reported percentiles subject to bias

• 16/31 papers with 1 24HR mentioned that usual intake or adjustment for within-person variation was needed

• 8/11 (73%) of papers using multiple 24HRs to report medians/percentiles, used a complex method (NCI/MSM)
Other Observations from the Air Pollution Cohort Survey

• Statements about the measurement error are often vague
  – The origin of the measurement error is often not clearly specified
  – The size and the impact of the measurement error is often not stated

• Measurement error is often mentioned but rarely addressed in detail or corrected
  – The majority of the studies use daily and spatially aggregated data
  – The often prevailing Berkson error (through temporal and spatial aggregation) is not or only insufficiently described and its implications are not discussed
  – Errors originating from staying in different microenvironments are often neglected or only poorly considered

• Many different exposure measures are analyzed separately or jointly; a homogeneous procedure is lacking
Conclusions

• In cohort studies: measurement error acknowledged, but implications not fully understood and commonly not addressed in statistical analysis
  – Very few used methods to adjust for measurement error
  – For PA studies, little motivation to adjust for error since the naïve associations are generally aligned with a priori hypotheses
  – Many studies had multiple variables measured w/error

• In dietary intake population surveys: minority corrected for measurement error
  – Majority of those that did apply a correction method were taking advantage of software (e.g. NCI method)

• Regression calibration most common method to address measurement error in diet and PA studies
More work is needed....

• Identify the various sources of measurement error
• Disseminate ideas of measurement error correction
  – Discussion of software in guidance documents, tutorials in clinical journals, talks at epi and clinical conferences
• Correct misconceptions, such as:
  – Random error won’t cause bias in associations
  – Attenuation is the only possible direction of bias
  – Categorization reduces the effect of measurement error
  – Validated questionnaires don’t have bias
  – Software is not available
References

Regression Calibration


Simex

References (2)

Iowa State University Method (ISU)

Multiple Source Method (MSM)

NCI Method
Dietary Intake Population Studies: Survey Methodology

• Date range: Jan 2012 – May 2015

• Term “Measurement error” not typically referred to in dietary intake surveys
  – Understood as variance around usual intake
  – Conducted Search A only
Physical Activity Cohort Studies: Survey Methodology

• Date range: Jan 2012 – Sep 2015

• Search A: Very broad search terms: N=8760 from search; randomly selected N=610; N=51 from abstract review

• SEARCH B: Added "measurement error" OR misreport* OR misclassif* OR bias OR attenuat* OR calibrat*
  – N=610 from search; N=86 from abstract review
Air Pollution Cohort Studies: Survey Methodology

• Date range: Jan 2012 – Dec 2014
• Search A broad search within „Web of Science“:
  – Search B Additional keywords: "measurement error", "measurement uncertainty", misclassif*, attenuat*
  – A: 4595 hits, B: 386 hits
• After abstract review: A: 431 hits, B: 32 hits
• Random selection: Search A: 50/Search B:25