

Statistical Methods to Address Measurement Error in Observational Studies: Current Practice and Opportunities for Improvement

Pamela Shaw

on behalf of STRATOS TG4

CEN-IBS Vienna, 31 August, 2017

Outline

- Background
- Motivating Example: The problem and a fix
- Literature survey methodology
- Literature survey 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

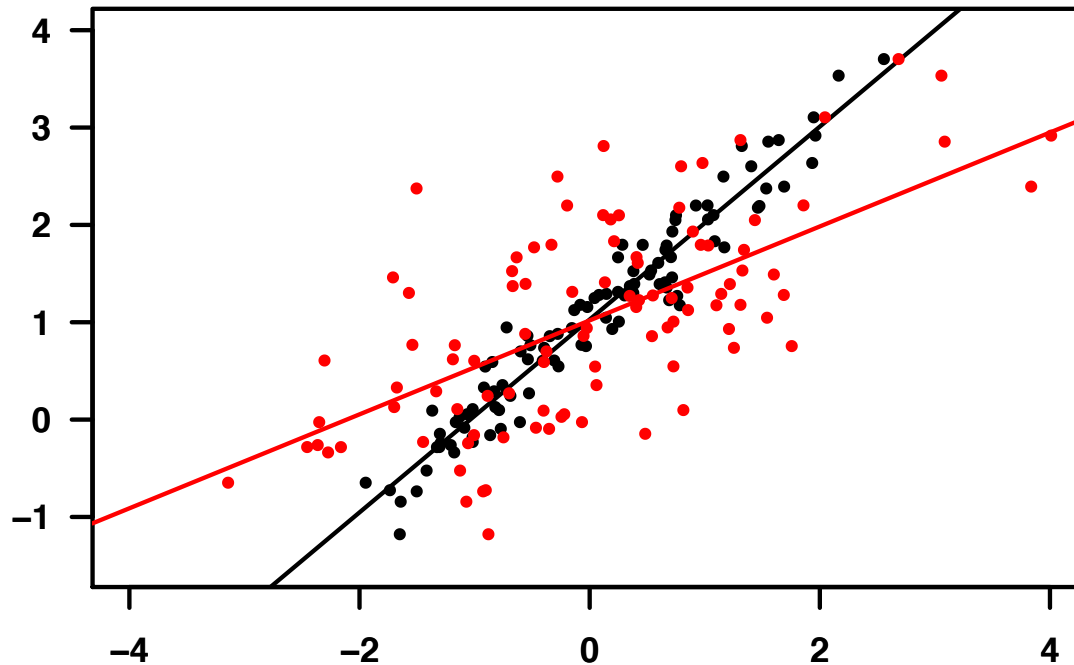
TG4 Projects

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

Background

- There have been many statistical advances to address measurement error in the past few decades (e.g. see Carroll et al 2006)
- 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

The Problem with Covariate Measurement Error...



Regression Calibration:

A simple approach to adjust for ME

Prentice *Biometrika* 1982

- Suppose true intake: X
- Error-prone measure: $X^* = b_0 + b_1X + b_2Z + b_3XZ + e$
- Objective biomarker: $X^{**} = X + u$
- Predicted X : $\hat{X} = E(X \mid X^* Z) = E(X^{**} \mid X^*, Z)$
$$= a_1 + a_2 X^* + a_3 Z + a_4 Z X^*$$
- **Regression calibration:** Regress outcome Y on predicted intake \hat{X} and other covariates Z . Obtain bootstrap SE to account for extra variance of estimating \hat{X}

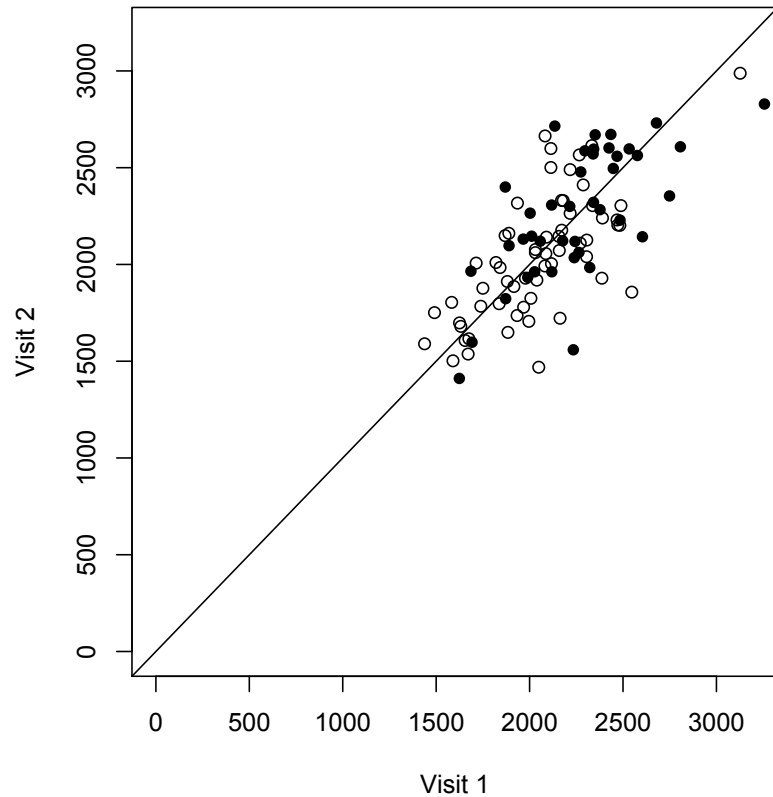
Regression Calibration: A few notes

- Regression calibration (RC) a very practical method applied many settings: Cox model, logistic regression, linear regression, etc.
- Regression calibration provides an unbiased estimate of β_X for linear regression of Y on X and other variables
 - generally not unbiased for non-linear regression models
 - RC can outperform other unbiased methods (Shaw et 2013) in terms of mean-squared error
- In the case of classical ME: $X^* = X+u$

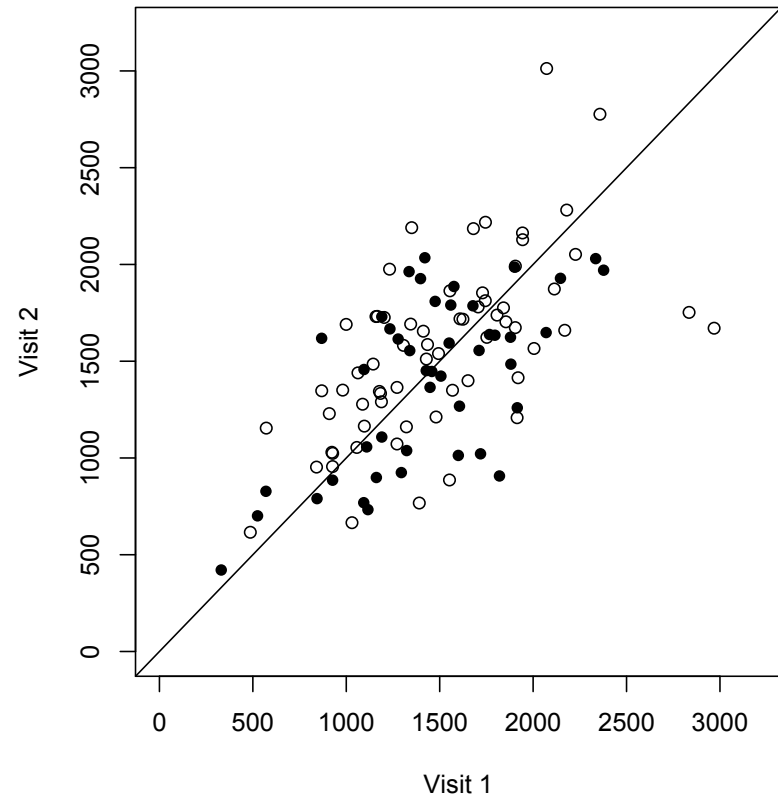
$$\beta_{X^*} = \lambda \beta_X, \text{ where } \lambda = \frac{\text{var}(X)}{\text{var}(X) + \text{var}(U)}$$

Nutritional Epidemiology Example: Measuring Energy Intake

Biomarker Energy



FFQ Energy



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*

Body mass index† category	Self-reported FFQ‡		Total energy expenditure		Calibrated FFQ	
	Geometric mean	IQR‡	Geometric mean	IQR	Geometric mean	IQR
Normal (<25.0)	1,407	1,157–1,759	1,894	1,714–2,083	1,912	1,853–1,980
Overweight (25.0–29.9)	1,462	1,196–1,837	2,043	1,904–2,232	2,028	1,962–2,103
Obese (≥30)	1,454	1,161–1,897	2,213	2,034–2,415	2,247	2,156–2,338

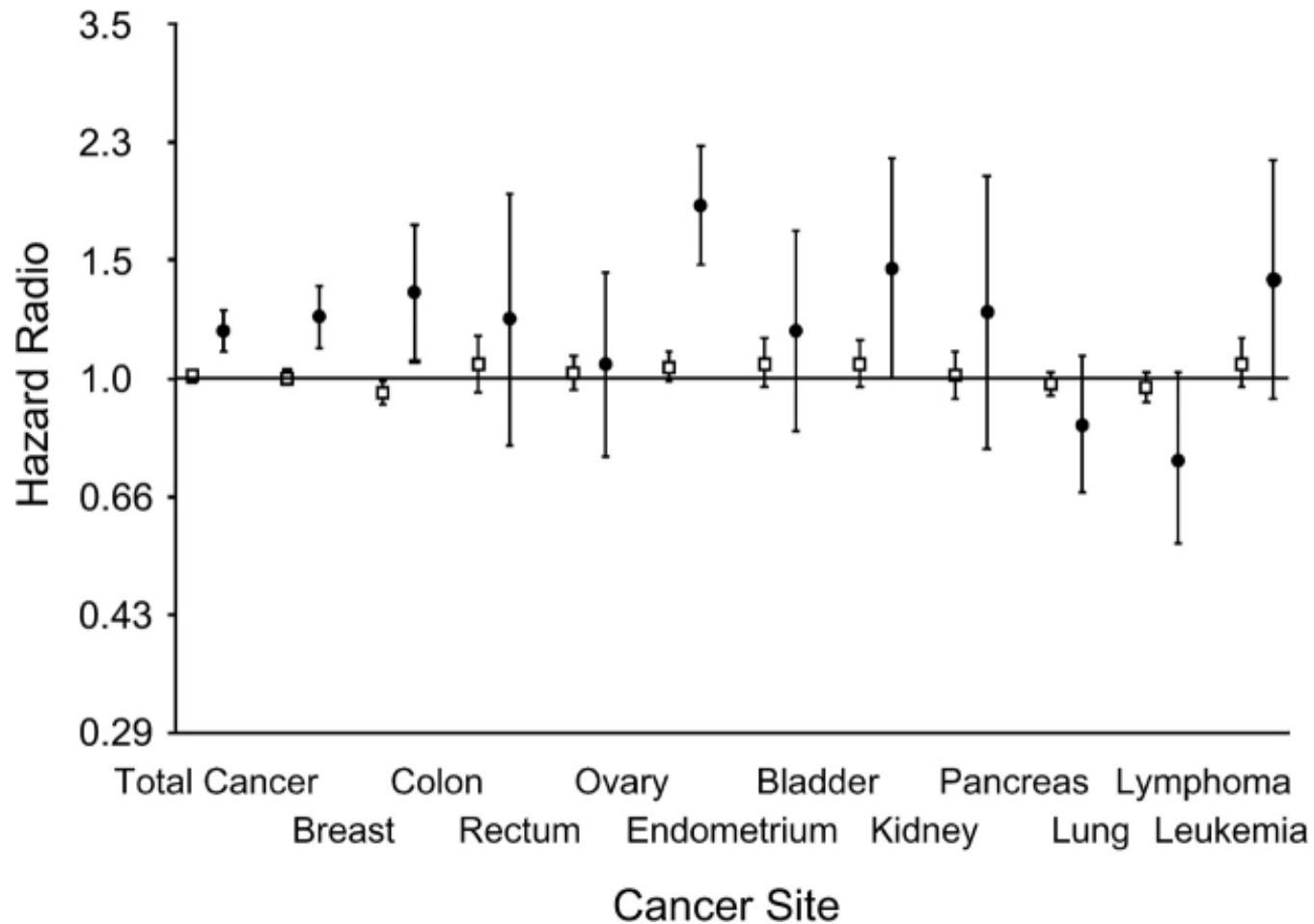
* Note that the difference between FFQ energy intake (self-report) and total energy expenditure (biomarker) increases as body mass index increases. The biomarker-calibrated estimates, for the same women, correct for the measurement error using the model shown in table 4.

† Weight (kg)/height (m)².

‡ FFQ, food frequency questionnaire; IQR, interquartile range (25th–75th percentiles).

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

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

Number of Articles Reviewed*

	Search A	Search B
Nutritional Epidemiology cohort studies	51	27
Dietary Intake Population Survey	67	N/A
Physical Activity cohort studies	30	40
Air Pollution cohort studies	50	25

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

Search A Survey Results

	Nutritional Epi Cohort N= 51	Phys activity Cohort N=30	Diet Intake Survey N=67	Pollution Cohort N=50
Mention ME as potential problem n(%)	48 (94%)	17 (57%)	53/67 (79%)	20 (40%)
Used a method to adjust for ME N (%)	5 (10%)	0 (0%)	19/67 (28%)	3 (6%)
% categorizing exposure	Any 50/51(98%) Exclusively 27/51 (53%)	Primary exposure 21/30 (70%)		
Statistic of main interest N (%)	HR 45 (88%) OR 3 (6%) RR 2 (4%) Slope 5(10%)	HR 11 (37%) OR/RR 9(30%) GLM 5 (17%) Other 5 (17%)	Mean 51 (76%) Median 28(42%) %-tiles 21(31%) Quality 31(46%)	

Methods to Address Measurement Error

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

- 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 (1)

General

- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). Measurement error in nonlinear models: a modern perspective. CRC press.
- Shaw PA and Prentice RL. Hazard Ratio Estimation for Biomarker-Calibrated Dietary Exposures. *Biometrics*, 68(2), 397-407.

Regression Calibration

- Prentice RL, Shaw PA, Bingham SA, Beresford, SA, Caan B, Neuhaus M L, Patterson R E, Stefanick ML, Satterfield S, Thomson CA, et al. (2009). Biomarker-calibrated energy and protein consumption and increased cancer risk among postmenopausal women. *American Journal of Epidemiology* **169**, 977–989.
- Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*. 1982 Aug 1;69(2):331-42.
- Zheng C, Beresford SA, Van Horn L, Tinker LF, Thomson CA, Neuhaus ML, Di C, Manson JE, Mossavar-Rahmani Y, Seguin R, Manini T, LaCroix AZ, Prentice RL. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *Am J Epidemiol*. 2014 Sep 1;180(5):526-35.

Simex

- Cook J and Stefanski LA. A simulation extrapolation method for parametric measurement error models. *Journal of the American Statistical Association* 1995; 89: 1314-1328.
- Küchenhoff, H, Mwalili SM, Lesaffre E. A general method for dealing with misclassification in regression: the misclassification SIMEX. *Biometrics*. 2006; 62(1): 85-96.

References (2)

Iowa State University Method (ISU)

- Nusser, S. M., Carriquiry, A. L., Dodd, K. W., and Fuller, W. A. 1996. A Semiparametric Approach to Estimating Usual Intake Distributions. *Journal of the American Statistical Association*, 91:1440-1449.

Multiple Source Method (MSM)

- Harttig U, Haubrock J, Knüppel S, Boeing H. 2011 The MSM program: web-based statistics package for estimating usual dietary intake using the Multiple Source Method. *Eur J Clin Nutr*. 65 S1:S87-9
- Haubrock J, Nöthlings U, Volatier JL, Dekkers A, Ocké M, Harttig U, Illner AK, Knüppel S, Andersen LF, Boeing H; European Food Consumption Validation Consortium. Estimating usual food intake distributions by using the multiple source method in the EPIC-Potsdam Calibration Study. *J Nutr*, 141, 914-20

NCI Method

- Tooze JA, Midthune D, Dodd KW, Freedman LS, Krebs-Smith SM, Subar AF, Guenther PM, Carroll RJ, Kipnis V. A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution<image001.gif>. *J Am Diet Assoc* 2006 Oct;106(10):1575-87.
- Tooze JA, Kipnis V, Buckman DW, Carroll RJ, Freedman LS, Guenther PM, Krebs-Smith SM, Subar AF, Dodd KW. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method<image001.gif>. *Stat Med* 2010 Nov 30;29(27):2857-68.