Berkson measurement error in epidemiology

Laurence Freedman

Gertner Institute for Epidemiology and Health Policy Research

Israel

STRATOS Task Group 4
Measurement Error and Misclassification

Aim

To increase:

(i) awareness of measurement error issues in observational epidemiology and

(ii) use of statistical methods to adjust for such error

STRATOS Task Group 4
Measurement Error and Misclassification

Hendriek Boshuizen	Netherlands
Raymond Carroll	USA
Veronika Deffner	Germany
Kevin Dodd	USA
Laurence Freedman	Israel
Paul Gustafson	Canada
Ruth Keogh	UK
Victor Kipnis	USA
Helmut Küchenhoff	Germany
Pamela Shaw	USA
Anne Thiébaut	France
Janet Tooze	USA
Michael Wallace	Canada

STRATOS Task Group 4
Publications

Shaw et al 2018
Epidemiologic analyses with error-prone exposures: review of current practice and recommendations.
Ann Epidemiol 2018, in press.

Freedman & Kipnis
Introducing TG4
Biometric Bulletin 2018 Vol 35, Issue 1

Submitted (in two parts):
STRATOS TG4 membership:
STRATOS Guidance Paper on measurement error and misclassification

Impact of Measurement Error on Study Results

Depends on:

- The amount of error
- The nature of the error –

  measurement error model
- What is being estimated

Content of this talk

Focus on the Berkson error model:

- Its definition
- Examples of when it occurs
- Impact on various estimates
- How to adjust for Berkson error
- Examples in epidemiology

To put all this in context, I will contrast it with the classical measurement error model

Classical Measurement Error Definition

$$X^* = X + e$$

X^* is the measurement that has error

X is the true (unknown) value

e is the (additive) error in measurement X^*

e has mean zero (X^* is unbiased)

e is independent of X

Classical Measurement Error Examples

- Average short term blood pressure
- Average short term serum cholesterol

In each of the above, error is due to:

- ❖ laboratory error
- ❖ biological variation and
- ❖ fluctuations over time

Berkson Measurement Error Definition

$$X = X^* + e$$

X^* is the measurement that has error

X is the true (unknown) value

e is the (additive) error in measurement X^*

e has mean zero

e is independent of X^*

Berkson Measurement Error
Some history

Joseph Berkson (1899-1982)

- Physicist, Physician and Biostatistician
- Headed the Biometry Unit at the Mayo Clinic from 1934-64
- Discussed “Berkson” measurement error in a 1950 paper in “Are there two regressions?”

J Am Stat Assoc 1950; 45:164–180.
doi:10.2307/2280676. JSTOR 2280676

Berkson Measurement Error Examples

- Berkson's example:

  Volume of preparation pipetted into a test tube in a laboratory experiment

- Exposure level in occupational medicine studies: groups of individuals classified according to average exposure

- Values obtained from a prediction equation:

  e.g. Schofield's equation for resting energy expenditure based on age, sex and weight

Berkson Measurement Error Prediction Equations

Table 2 Predictive equations for basal metabolic r

Age (years)	Schofield n^{**}	Schofield Equations
Men	4809	
0–2.9	162	0.249 × Wt − 0.127
3.0–9.9	338	0.095 × Wt + 2.110
10.0–17.9	734	0.074 × Wt + 2.754
18.0–29.9	2879	0.063 × Wt + 2.896
30.0–59.9	646	0.048 × Wt + 3.653
≥ 60.0	50	0.049 × Wt + 2.459
60.0–69.9		
≥ 70.0		
Women	2364	
0–2.9	137	0.244 × Wt − 0.130
3.0–9.9	413	0.085 × Wt + 2.033
10.0–17.9	575	0.056 × Wt + 2.898
18.0–29.9	829	0.062 × Wt + 2.036
30.0–59.9	372	0.034 × Wt + 3.538
≥ 60.0	38	0.038 × Wt + 2.755
60.0–69.9		
≥ 70.0		

Each age group prediction equation is a regression of the form:

$REE = b_0 + b_1 \times Wt + e$, with e independent of predicted value

Impact on estimates

Types of Estimate:

- ❑ Percentiles of X: observe X*

- ❑ Coefficient of X in regression of Y on X

  Y is measured exactly, but observe X*, not X
- ❑ Coefficient of X in regression of Y on X

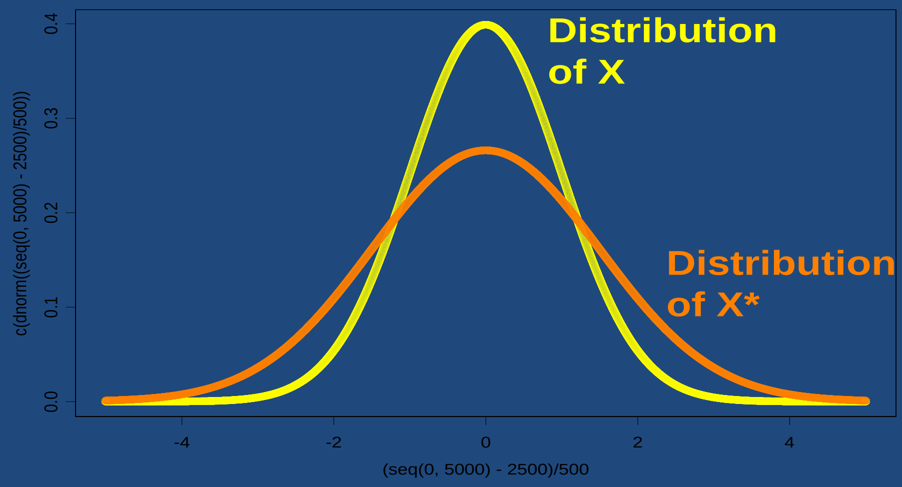
  X is measured exactly, but observe Y*, not Y

Impact on Estimates

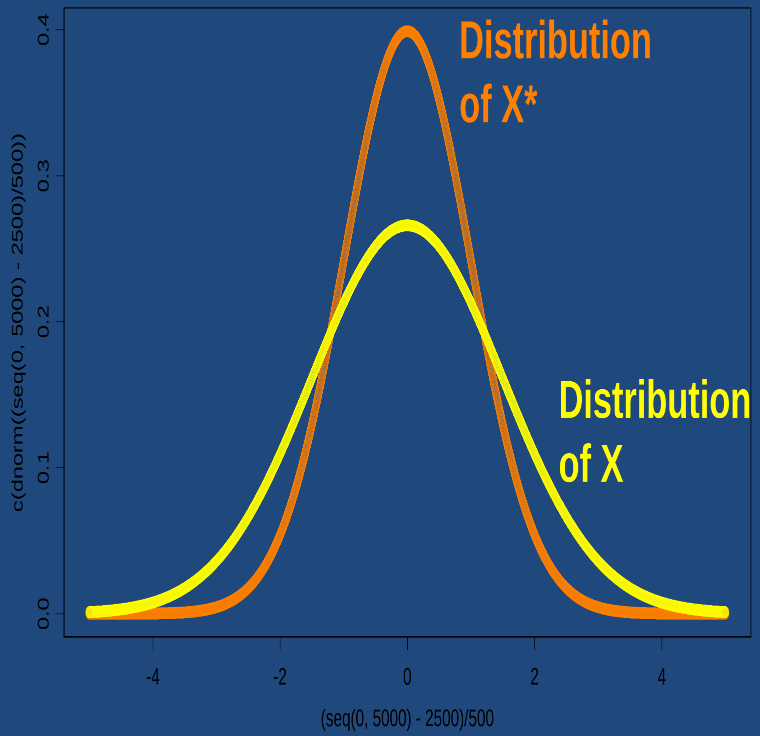
The impacts of classical and Berkson errors on these estimates are opposite!

Percentiles of X

Classical error



Berkson error



Percentiles of X

Estimate	Classical	Berkson
Upper percentile	Overestimate	Underestimate
Lower percentile	Underestimate	Overestimate

Method of Adjustment for Berkson Error

Berkson error: $X = X^* + e$

- The unadjusted estimate forms a distribution of the X^* values

- Instead, use moment reconstruction (MR):
- Form a new variable X_{MR}

  $X_{MR} = (1\text{-}w)\overline{X^*} + wX^*$, where

  $w = \mathrm{SD}(X)/\mathrm{SD}(X^*)$: note that $w > 1$

  $\mathrm{E}(X_{MR}) = \mathrm{E}(X)$; $\mathrm{var}(X_{MR}) = \mathrm{var}(X)$

- w is estimated from external information
- Form the distribution using the X_{MR} values

Example from the OPEN dietary reporting validation study

Potassium intake (K)

K_{FFQ} = Food Frequency Questionnaire report of K

The study also included a urinary determination of K

Calibration (prediction) equation:

$$\ln(K) = 5.895 + 0.271*\ln(K_{FFQ}) - 0.193*\text{sex} + 0.00035*\text{age}$$

This equation for ln(K) has Berkson error

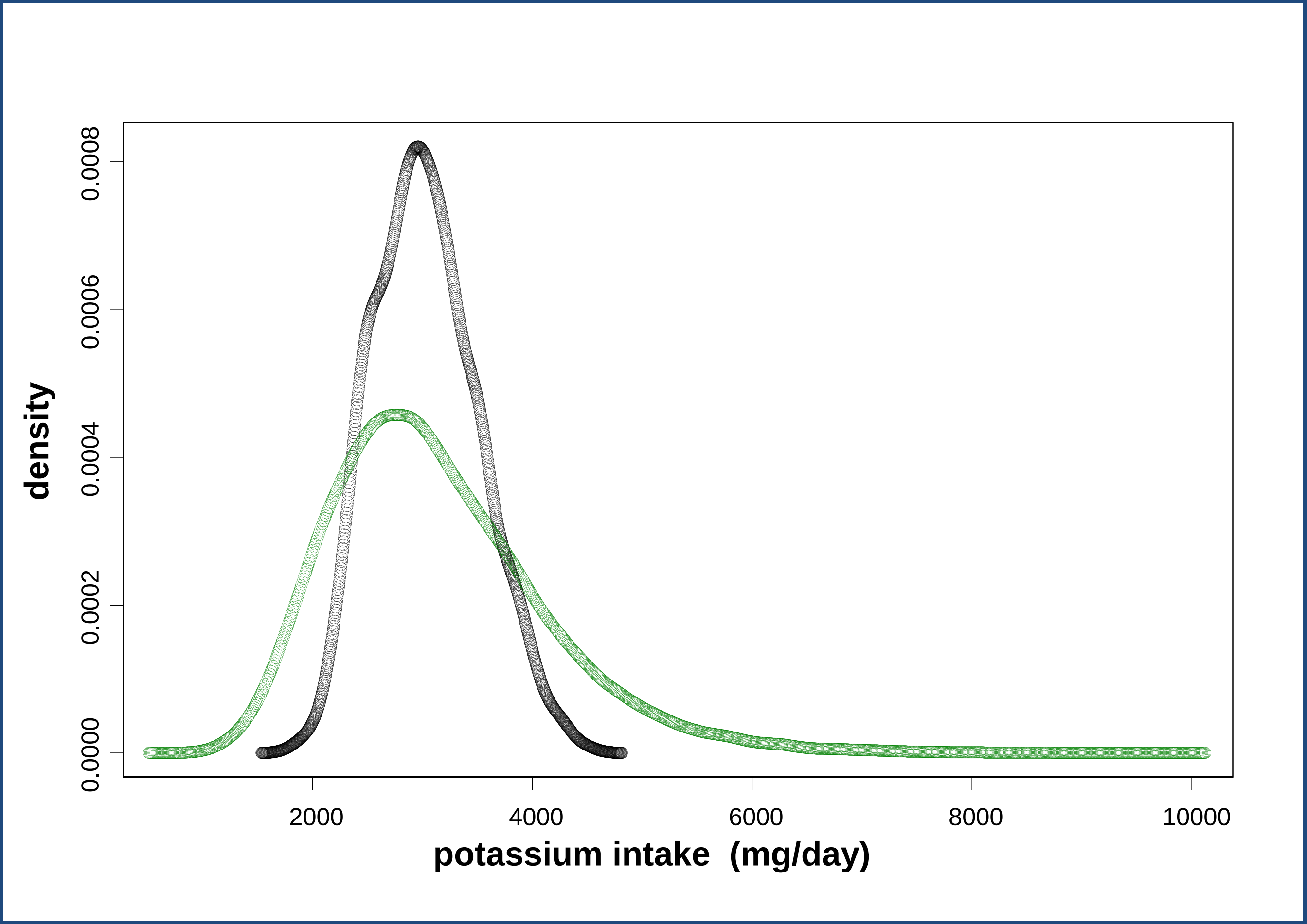
Var(predicted ln(K)) = 0.0239

Var(prediction residual) = 0.0682

MR method: $w = \sqrt{\{(0.0239+0.0682)/0.0239\}} = 1.96$

Results

Black = Empirical distribution of predicted potassium intake
Green = Adjusted for Berkson error



Impact on estimates

Types of Estimate:

- ❑ Percentiles of X: observe X*

- ❑ Coefficient of X in regression of Y on X

  Y is measured exactly, but observe X*, not X

- ❑ Coefficient of X in regression of Y on X

  X is measured exactly, but observe Y*, not Y

An extra assumption

- The errors are non-differential

- For the case where X is measured with error, this means:

  X* and Y are independent conditional on X

- For the case where Y is measured with error, this means:

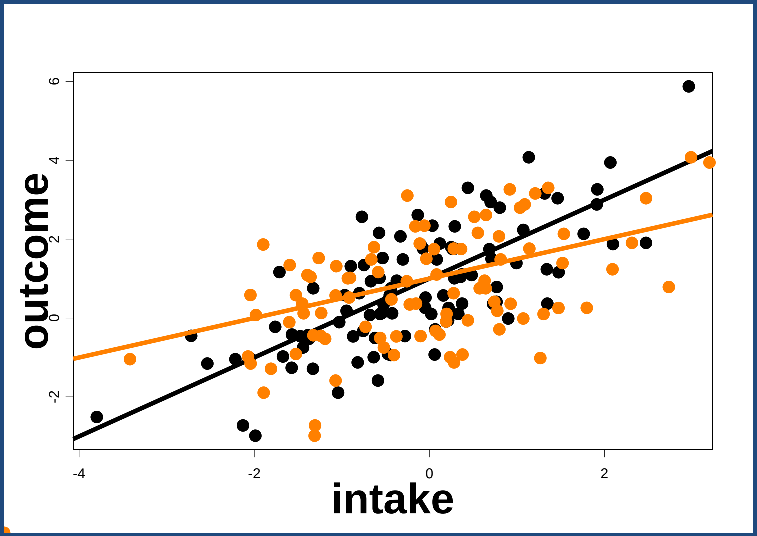
  Y* and X are independent conditional on Y

Impact on Estimates

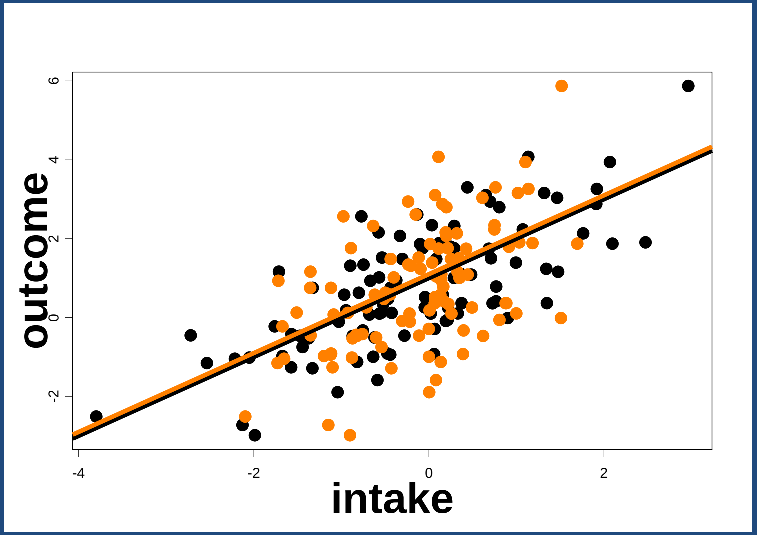
The impacts of classical and Berkson errors on these estimates are opposite!

Impact on Estimates of Regression Coefficients

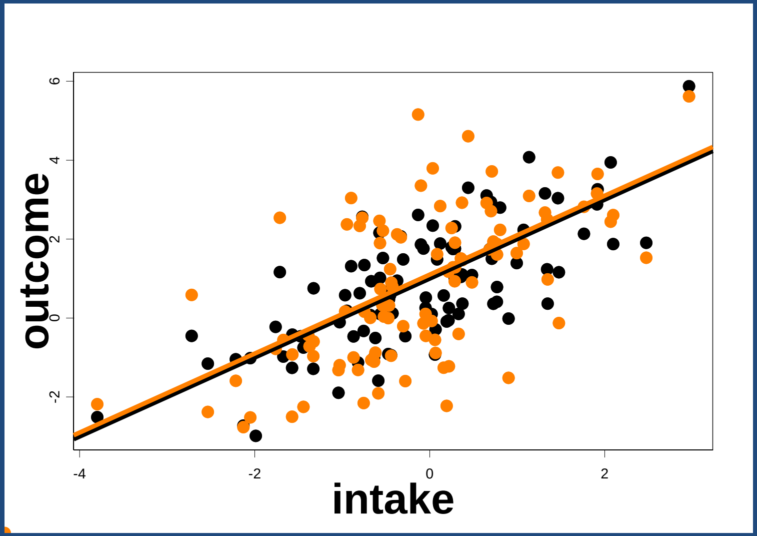
Classical Error in X



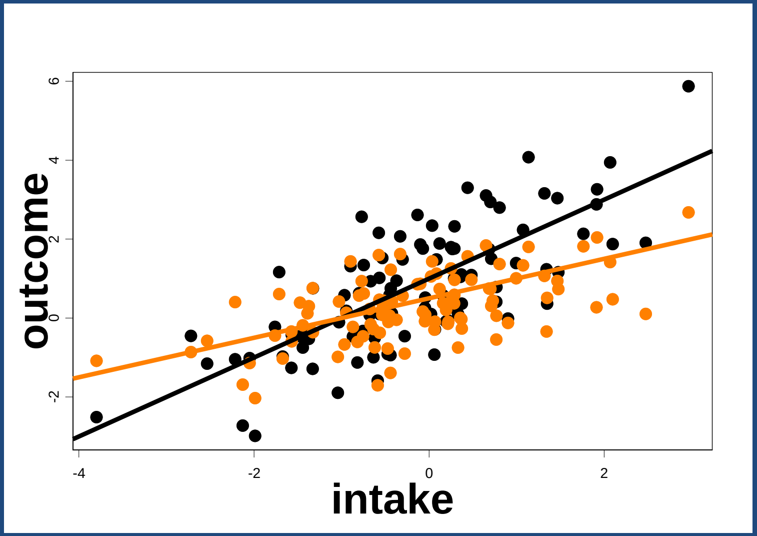
Berkson Error in X



Classical Error in Y



Berkson Error in Y



Regression coefficient of X in regression of Y on X

Variable measured with error	Estimate	Classical	Berkson
X	Regression coefficient	Attenuated	Unbiased
Y	Regression coefficient	Unbiased	Attenuated

How to adjust for classical error in X?

Regression calibration:

- ❑ Obtain a "calibration" equation: E(X|X*)
- ❑ Substitute E(X|X*) for X in regression of Y on X

Why does it work?

E(X|X*) has Berkson error as an estimate of X.

Berkson error in a covariate does not cause bias in estimation.

How to adjust for Berkson error in Y?

"Inverse regression calibration":

- ❑ Invert the Berkson measurement error model $Y=Y^*+U$ to:
  $Y^* = \alpha_0 + \alpha_1 Y + U^*$
- ❑ Form $Y^{est} = (Y^* - \alpha_0)/\alpha_1$ (Buonaccorsi, 1991)
- ❑ Substitute Y^{est} for Y in regression of Y on X

Why does it work?

Y^{est} has classical error as an estimate of Y.

Classical error in Y does not cause bias.

Example from OPEN: does potassium density intake vary with educational level?

Potassium density (mg/kcal)

Calibration (prediction) equation:

$\ln(Kden) = -0.385 + 0.480*\ln(Kden_{FFQ}) - 0.029*sex + 0.00602*age$

This equation for ln(Kden) has Berkson error

Var(predicted ln(Kden)) = 0.0203

Var(prediction residual) = 0.0696

Inverse regression calibration:

Value of Y^{est} to be entered into model of Y on X:

$(\ln(Kden) - 0.124) / 0.226$

Example from OPEN: does potassium density intake vary with educational level?

1. Run regression of ln(Kden) on education, sex and age.
2. Estimate median levels of Kden (mg/1000 kcal) for women, aged 50y, according to educational level

Educational level	Using Y = predicted ln(Kden)	Inv. reg. calib.	Unbiased estimate
High school	1093	856	996
College	1113	933	1095
Post-grad	1159	1110	1216

Summary

- With the increasing use of prediction and calibration equations in medicine, Berkson error will be encountered more and more
- The commonly assumed adage that Berkson error does not cause bias in estimates is wrong.
- Awareness of the effects of Berkson error and methods to adjust for it need more attention

Acknowledgements

Raymond Carroll	USA
Veronika Deffner	Germany
Kevin Dodd	USA
Paul Gustafson	Canada
Ruth Keogh	UK
Victor Kipnis	USA
Helmut Küchenhoff	Germany
Pamela Shaw	USA
Janet Tooze	USA
Michael Wallace	Canada