Barriers and challenges to the use of statistical methods for addressing errors in the measurement and classification of outcome and explanatory variables in observational studies

Laurence Freedman

Gertner Institute for Epidemiology, Israel

and

Information Management Services, Rockville, MD

on behalf of STRATOS TG4

STRATOS
STRengthening Analytical Thinking for Observational Studies

TG4

Topic Group 4

Measurement Error and Misclassification

STRATOS TG4 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

The central problem

1. Errors commonly occur in the measurement and classification of many variables that are used in epidemiological observational studies,

BUT

2. In many fields of epidemiological research the impact of such errors is either not appreciated or is ignored.

Aim of the Presentation

1. Document the extent of the problem

2. Examine the factors contributing to the problem

3. Suggest first steps towards improvement

Content

1. A historical example
2. TG4 literature surveys
3. Barriers to using methods that adjust for measurement error
4. Steps to tackling the problem

Historical example – adjusting for measurement error

Blood pressure, stroke, and coronary heart disease

Part 1, prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias

STEPHEN MACMAHON RICHARD PETO JEFFREY CUTLER
RORY COLLINS PAUL SORLIE JAMES NEATON ROBERT ABBOTT
JON GODWIN ALAN DYER JEREMIAH STAMLER

The Lancet 1990; 335: 765-774

Abstract

Previous analyses have described the uncorrected associations of DBP measured just at "baseline" with subsequent disease rates. But, because of the diluting effects of random fluctuations in DBP, these substantially underestimate the true associations of the usual DBP (ie, an individual's long-term average DBP) with disease. After correction for this "regression dilution" bias, prolonged differences in usual DBP of 5, 7·5, and 10 mm Hg were respectively associated with at least 34%, 46%, and 56% less stroke and at least 21%, 29%, and 37% less CHD. These associations are about 60% greater than in previous uncorrected analyses. (This regression dilution bias is quite

What is correction for the "regression dilution" bias?

True model: $Y = \beta_0 + \beta_1 X + \varepsilon$
Instead of X we observe X*
Calibration model: $X = \lambda_0 + \lambda_1 X^* + \delta$

Then in a model of Y on X* the coefficient of X* is $\lambda_1\beta_1$
To correct, we need to divide the estimated coefficient of X* by λ_1

In most epidemiologic problems $0 < \lambda_1 < 1$.

What is correction for the "regression dilution" bias?

This correction is a simple version of a more general approach to correction known as "regression calibration", in which we use the calibrated value of X, instead of X* in the regression.

Impact

As a result of this and later efforts, estimates of the contribution of blood pressure, cholesterol and smoking to the incidence of heart disease in the UK were re-assessed.

The attributable fraction was found to be around 80% rather than the previously estimated 50%, providing new impetus to heart disease prevention programs.

STRATOS TG4 Preliminary Question

What methods of statistical adjustment are being used today in areas of epidemiology where measurements are made with error?

TG4 Literature Surveys

- Nutritional cohorts (Pamela Shaw and Ruth Keogh)
- Dietary surveys (Kevin Dodd)
- Physical Activity cohorts (Janet Tooze)
- Air Pollution studies (Veronika Deffner and Helmut Kuechenhoff)

First three areas rely primarily on self-reported values of exposure (linear measurement error)

Air pollution studies rely mostly on “fixed-site” estimates of exposure (Berkson error)

TG4 Literature Surveys

Main questions

1. What proportion of study reports

(a) mention measurement error as a potential problem?

(b) use a statistical method to adjust for biases caused by measurement error?

[2. Among those study reports where an adjustment method is used, what was the method?]

TG4 Literature Surveys

Main Approach

- Papers published between 2012-15
- ~50 papers per area
- Information extracted by the main investigator/s
- 20% quality control by a second extractor

TG4 Literature Surveys

Results

	Nutritional Cohorts (N=51)	Physical Activity Cohorts (N=30)	Dietary Surveys (N=67)	Air Pollution Studies (N=48)
Mentioned ME as a problem	94%	57%	79%	40%
Used a statistical method of adjustment	10%	0%	28%	6%
Categorized the exposure	53%	70%	-	-

TG4 Literature Surveys

Main Conclusions

1. In half or more of the reports the investigators were aware that measurement error was a potential problem.
2. However, very few tried to adjust for it.

Why?

Barriers to adjusting for measurement error

1. Lack of, or problems with, data for determining the type and magnitude of the error
2. Inadequate standards of “validation” of measurement instruments
3. Lack of appreciation of the biases and the loss of precision resulting from measurement error
4. Lack of awareness of software for implementing methods of adjustment; incompleteness of software and lack of user friendliness
5. Lack of methods for dealing with categorized data

Examples from nutrition

1. Problems with determining the ME model

Notation: X = true dietary intake

X* = self-reported dietary intake

a) One needs a sub-study in which both X* and X are measured. If X is unavailable, an unbiased measurement of X is sufficient.
b) Often even an unbiased measurement of X is unavailable. So one has to compromise and use a measure that has some bias but less than X*.
c) Sub-studies need to be of sufficient size to provide accurate estimates of the ME parameters – we may be talking about n = several hundreds or even 1000. So is it really worth it?

Examples from nutrition

2. Inadequate standards of validation

a) Investigators often write that their questionnaire has been “validated”.
b) In practice this validation is a comparison of X^* with another imperfect measure of X (X^{**})
c) The validation study has reported that the correlation between X^* and X^{**} is some value between 0.4 and 0.6.
d) This is considered sufficient evidence to use X^* as a measure of X (!)

3. Lack of appreciation of biases due to ME

a) Biases depend upon the type of measurement error (e.g. classical, Berkson) and the type of estimate (e.g. regression coefficient, percentile of a distribution)
b) There is lack of awareness of what the potential biases are.

3. Lack of appreciation of biases due to ME

Classical measurement error

$X^* = X + e$; $E(e)=0$ and e independent of X

Happens quite often with laboratory or clinical measurements

Berkson measurement error

$X = X^* + e$; $E(e)=0$ and e independent of X^*

Happens with predicted values derived from a regression model; or assigning a group value to an individual

3. Lack of appreciation of biases due to ME
Effects of Classical and Berkson error

Variable	Estimate	Classical	Berkson
X	Regression coefficient	Attenuated	Unbiased
Y	Regression coefficient	Unbiased	Attenuated
Y	Upper percentile	Overestimate	Underestimate
Y	Lower percentile	Underestimate	Overestimate

3. Lack of appreciation of biases due to ME
Effects of Berkson error on Percentiles

Estimating the Population Distribution of Usual 24-Hour Sodium Excretion from Timed Urine Void Specimens Using a Statistical Approach Accounting for Correlated Measurement Errors

J Nutrition

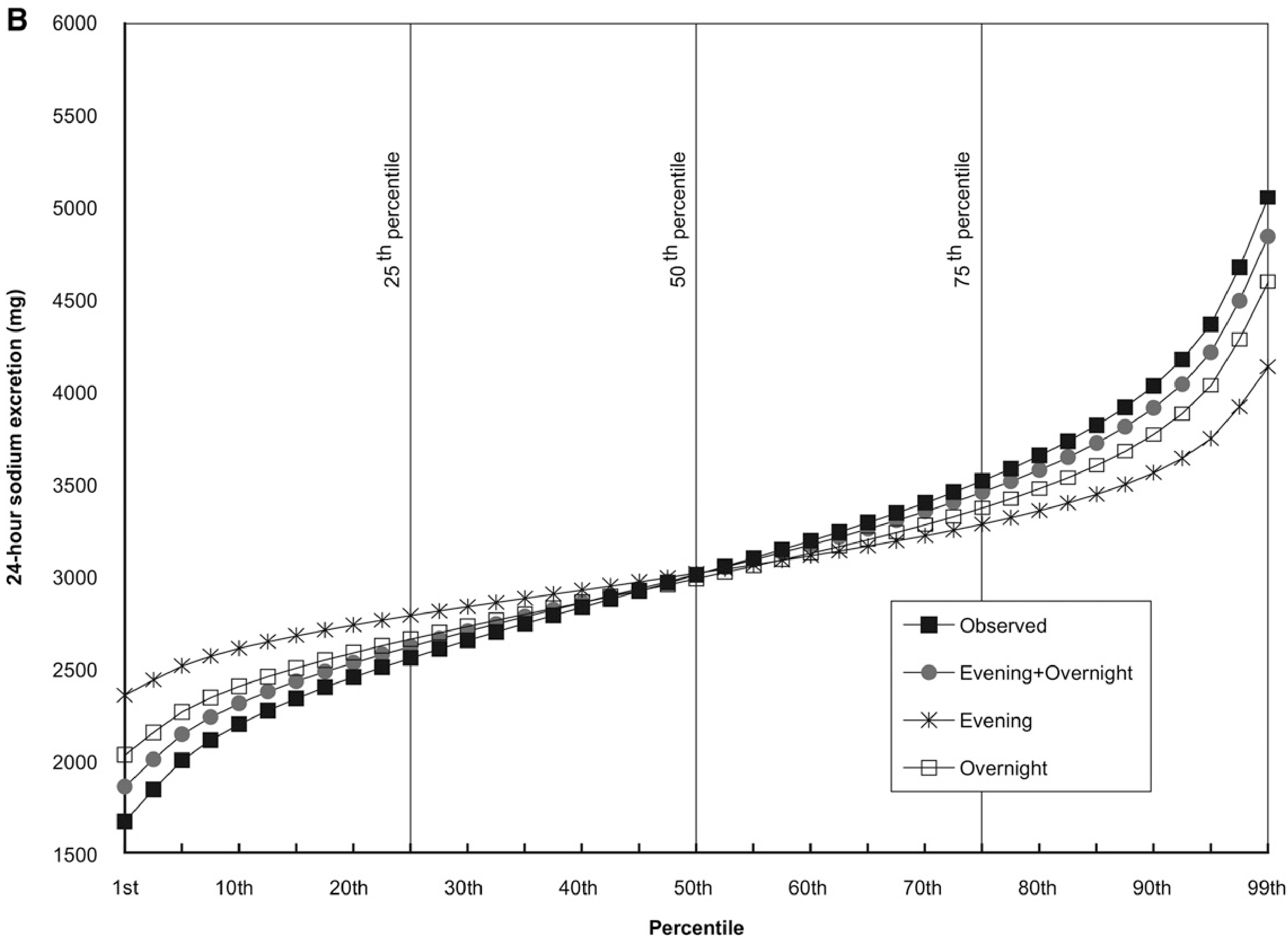
3. Lack of appreciation of biases due to ME Effects of Berkson error on Percentiles

- Urinary excretion of sodium over a 24 hour period is a very good measure of sodium intake over the previous 24 hours
- However, collecting 24 hours of urine from an individual is awkward
- In this study the authors examined whether collecting a limited number of timed urine voids during the 24 hour period would suffice

3. Lack of appreciation of biases due to ME
Effects of Berkson error on Percentiles

- They developed a prediction equation for the full 24 hour value using values from a limited number of timed void values.
- They then compared the distribution of the predicted values with the distribution of the true observed values.
- They forgot that when we use prediction (or calibration) equations, the predictions usually have Berkson error.

Estimated percentiles: True values in bold



3. Lack of appreciation of biases due to ME
Effects of Berkson error on Percentiles

- The authors erroneously concluded that timed voids could not be used in place of a full 24-hour urine collection.
- We are now working with them to revisit this question with appropriate methodology.

4. Lack of appreciation of software available, and incompleteness of software

- STATA programs for regression calibration (Carroll)
- NCI programs for adjusting estimates of distributions

  see website:

http://epi.grants.cancer.gov/diet/usualintakes/method.html

- But there is much room for expansion and improvement

TG4 Literature Surveys

Results

	Nutritional Cohorts (N=51)	Physical Activity Cohorts (N=30)	Dietary Surveys (N=67)	Air Pollution Studies (N=48)
Mentioned ME as a problem	94%	57%	79%	40%
Used a statistical method of adjustment	10%	0%	28%	6%
Categorized the exposure	**53%**	**70%**	-	-

5. Categorized exposures

Exposure X is continuous, but the investigator prefers to analyze it in categories, such as quintiles.

In the simplest case, a single exposure measured with error, Kipnis & Izmirlian (Am J Epid, 2002) provide a method of correction.

Currently there is no available general method for adjusting relative risk estimates for ME when continuous exposures are categorized.

What can be done to overcome these barriers?

- Publication of papers highlighting problems caused by ME and the barriers to overcoming them
- Workshops/courses at statistical and epidemiologic meetings on these topics
- Education of and encouragement of young statisticians to work in this area
- Advertisement of software for measurement error adjustment and further development
- Development of methods for categorized data

- This program is the basis for the future work of TG4