

Epi-nomics

Applying lessons from epidemiology research on
misclassification to economics policy evaluation

Patrick William Koval¹ Daniel Schwab²

¹Boston University School of Public Health, Department of Biostatistics

²College of the Holy Cross, Department of Economics and Accounting

October 1, 2026



Intro & Outline

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- 1 Key question in economics: what is the causal effect of a policy?
 - Outcome variable is often not measured perfectly.
 - How much bias does misclassification cause?
- 2 Ex Ante Quantitative Bias Analysis (QBA): simulations with synthetic data
- 3 Ex Post QBA: simulate misclassification with real data
 - Example with a published paper

Takeaways

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

Results

Discussion

References

References

- Nondifferential (uniform) misclassification causes bias toward the null, but relatively small
- If misclassification varies depending on whether policy is in place, potentially much larger bias
 - If treated areas have higher sensitivity than treated (less undercounting), upward bias.
 - In the reverse case, downward bias.
 - Worse with increasing disparity, bias up to double the given effect size.
- These findings are consistent across use of both simulated and real-world data.

Misclassification in Economics I

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- Outcome variables used in economics/policy analysis are at risk of misclassification/underreporting.
- Earnings and unemployment.
- Crime rates.
- Health behaviors and outcomes.

Misclassification in Economics II

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- Some literature addresses misclassification/underreporting.
- Mostly introducing new estimators or *ex post* adjustments to specific data types.
- Want to:
 - Study bias patterns.
 - Provide researchers with tool to analyze effects of misclassification.

Progress in Public Health

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- Foundations and development largely in public health context, largely applied to tests and diagnosis.
 - 1954: Simplest version formalized mathematically.¹
 - 1977: “Nondifferential misclassification biases toward the null” heuristic.²
 - 1996: Foundations of QBA.³
- Largely now focused on using computational power to produce sensitivity analyses.⁴
 - We’ll call this “*Ex Post* QBA.”

¹Bross, 1954

²Copeland et al., 1977

³Greenland, 1996

⁴Fox et al., 2021; Fox et al., 2022

Misclassification I: Terms

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

Sensitivity

The probability of correctly classifying a positive, or

$$\mathbb{P}(A_{classified} = 1 | A_{true} = 1).$$

Specificity

The probability of correctly classifying a negative, or

$$\mathbb{P}(A_{classified} = 0 | A_{true} = 0).$$

Differential Misclassification

Misclassification whose occurrence (sensitivity, specificity) changes depending on some variable/characteristic.

Misclassification II: Measurement Error

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

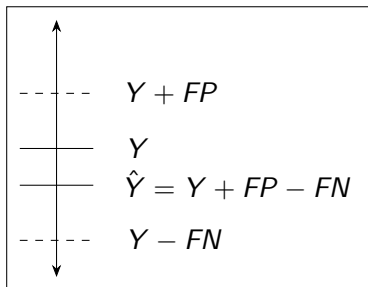
Results

Discussion

References

References

Figure: Conceptualization of Misclassification vs. Measurement Error



$Y/\hat{Y}$: True/Observed Outcome FP: False Positives FN: False Negatives

- Distribution of false positive and false negative matters.
- Information potentially lost when only thinking about measurement error.

Ex Ante QBA: Overview

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

Results

Discussion

References

References

- Synthetic data based on plausible. . .
 - prevalences,
 - effect estimates, and
 - sensitivity and specificity.
- Allows for perfect control to see what hidden biases might look like.

Ex Ante QBA: Causal Structure and Simulation

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

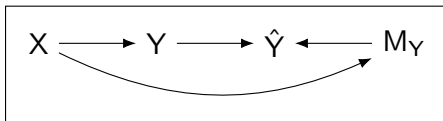
Results

Discussion

References

References

Figure: DAG—Individual-Level Causal Structure



X: Treatment Y/ $\hat{Y}$: True/Observed Outcome M_Y : Outcome Misclassification

- Considering case of policy treatment.
 - Stochastic enactment with 50% probability.
 - Once enacted, stays enacted. Everyone considered treated.
- General Steps
 - 1 Simulate individual-level data for each region-year.
 - 2 Misclassify across outcome categories.
 - 3 Collapse to region-year true and observed outcome counts.

Ex Ante QBA: Specifications

Table: Ex Ante QBA Simulation Specifications

Parameter	Value
Treatment	Policy Enactment
Population (Pop.)	3,500,000/region (state)
Time Frame	20 periods (years)
Baseline Y Incidence	0.020 (20 per 1,000)
Pop. at Risk of FP	0.060 (60 per 1,000)
Correlation Structure	$AR(1), \rho = 0.3, \sigma_{\varepsilon} = 0.004$
Bias Measure	Additive Bias $(\hat{RD} - RD)$
Risk Difference (RD)	▪ $RD \in (-8 \text{ per } 1,000, 8 \text{ per } 1,000)$
Plot Parameters	▪ $Se = 0.80, 1.00, Sp = 0.98, 1.00$
Bias Plot	▪ $RD = 4 \text{ per } 1,000$
Parameters	▪ $Se \in (0.80, 1.00), Sp \in (0.96, 1.00)$

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

Ex Ante QBA: False negatives, no false positives

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

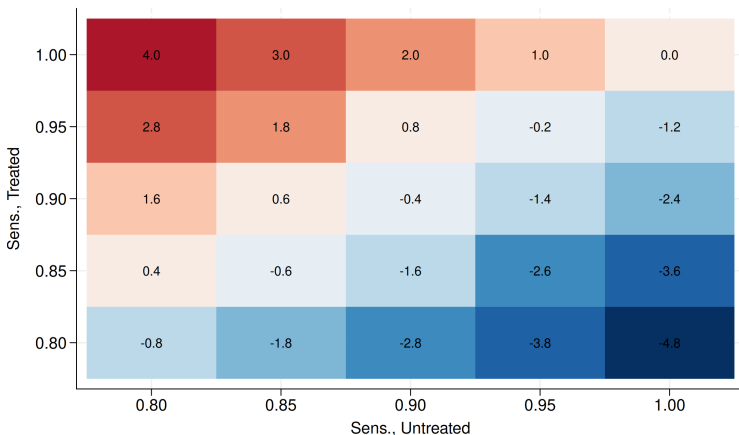
Overview & Methods
Results

Discussion

References

References

Figure: Varying Sensitivity (Risk difference = 4 per 1,000)⁵



⁵Cells represent bias in risk difference per 1,000.

Ex Ante QBA: False positives, no false negatives

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

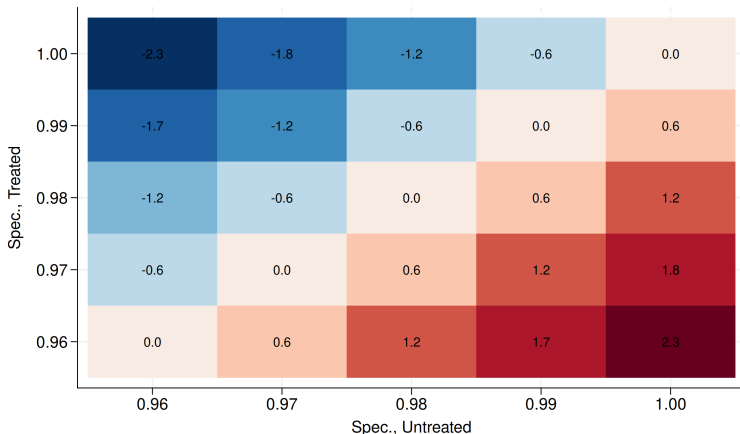
Results

Discussion

References

References

Figure: Varying Specificity (Risk difference = 4 per 1,000)⁶



⁶Cells represent bias in risk difference per 1,000.

Ex Ante QBA: Varying Risk Differences

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

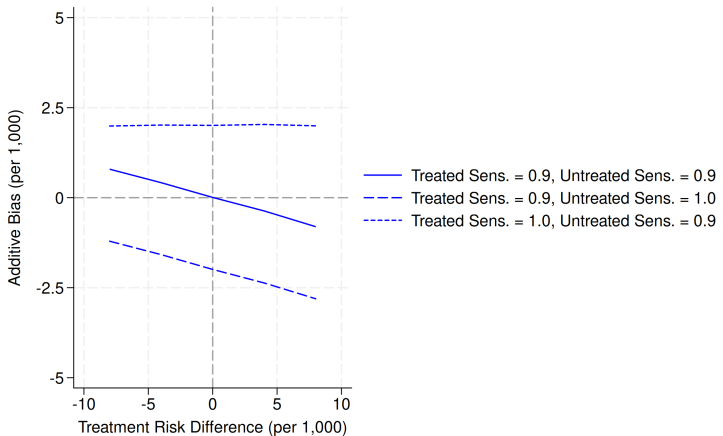
Results

Discussion

References

References

Figure: Assume perfect specificity (no false positives)



Ex Post QBA: Overview

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- We are developing Stata program to run sensitivity analyses alongside regression, grateful for feedback.
- Establishes degree of bias due to misclassification.
- Specify plausible bias parameters to get general sense of uncertainty introduced by potential misclassification.

Ex Post QBA: Example Context

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- Earlier work with Matthew S. Johnson: we showed laws protecting whistleblowers reduced workplace injuries.⁷
- We observe OSHA accident inspections, which are triggered by workplace with 3+ hospitalizations.
- Ideal data: number of accidents where 3+ people were seriously injured.
 - False negatives are possible: injured workers may avoid hospital.
 - False positives are less plausible
- We assume perfect specificity, measure bias under different levels of sensitivity.
 - Sensitivity could vary in states with and without whistleblower protections.

⁷ Johnson, Schwab, Koval 2024

Ex Post QBA: Example Results II

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

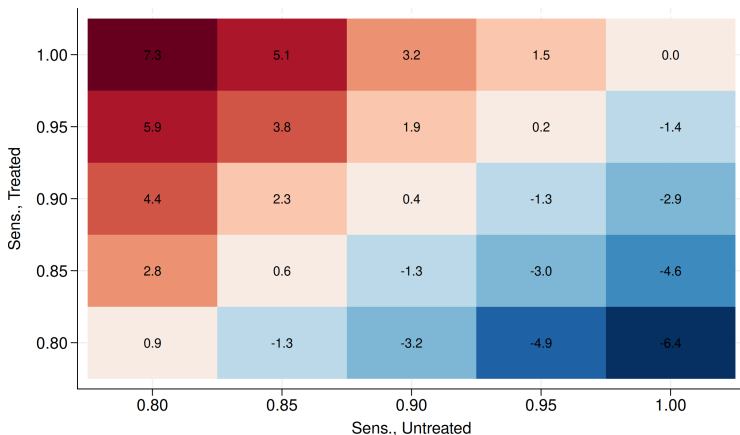
Overview & Methods
Results

Discussion

References

References

Figure: Varying Assumed Sensitivity (Observed Coefficient = -3.9)⁸



⁸Cells represent bias in accidents per 1 million.

Discussion

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- *Ex Ante* QBA
 - Significant and varied effects of risk.
 - Specificity has little effect at plausible ranges.
- *Ex Post* QBA
 - Undercounting can cause substantial bias, especially when it varies widely between treated/untreated groups.
 - Bias direction dependent on which group has more accurate reporting.
- Stata program in progress to analyze effects of misclassification on policy analysis.

References

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods

Results

Ex Post QBA

Overview & Methods

Results

Discussion

References

References

Bross, I. (1954, December). Misclassification in 2 X 2 Tables. Biometrics 10(4), 478.

Copeland, K. T., H. Checkoway, A. J. McMICHAEL, and R. H. Holbrook (1977, May). BIAS DUE TO MISCLASSIFICATION IN THE ESTIMATION OF RELATIVE RISK. American Journal of Epidemiology 105(5), 488–495.

Fox, M. P., R. F. MacLehose, and T. L. Lash (2021). Applying Quantitative Bias Analysis to Epidemiologic Data (2nd ed.). Statistics for Biology and Health. Cham: Springer International Publishing.

Fox, M. P., R. Nianogo, J. E. Rudolph, and C. J. Howe (2022, June). Illustrating How to Simulate Data From Directed Acyclic Graphs to Understand Epidemiologic Concepts. American Journal of Epidemiology 191(7), 1300–1306.

Greenland, S. (1996, December). Basic Methods for Sensitivity Analysis of Biases. International Journal of Epidemiology 25(6), 1107–1116.

Thank You! Questions?

Epi-nomics

Patrick Koval,
Daniel Schwab

Introduction

Background

Ex Ante QBA

Overview & Methods
Results

Ex Post QBA

Overview & Methods
Results

Discussion

References

References

- Questions?
- Suggestions?
- Other Feedback?
- Reach us at:
 - patrickkoval@pm.me
 - dschwab@holycross.edu