# Answering questions by combining data sources

Paul Zivich

Institute of Global Health and Infectious Diseases
University of North Carolina at Chapel Hill

June 12, 2023

# Acknowledgements

Supported by NIH T32-AI007001

pzivich@unc.edu pzivich

# Why Multiple Data Sources?

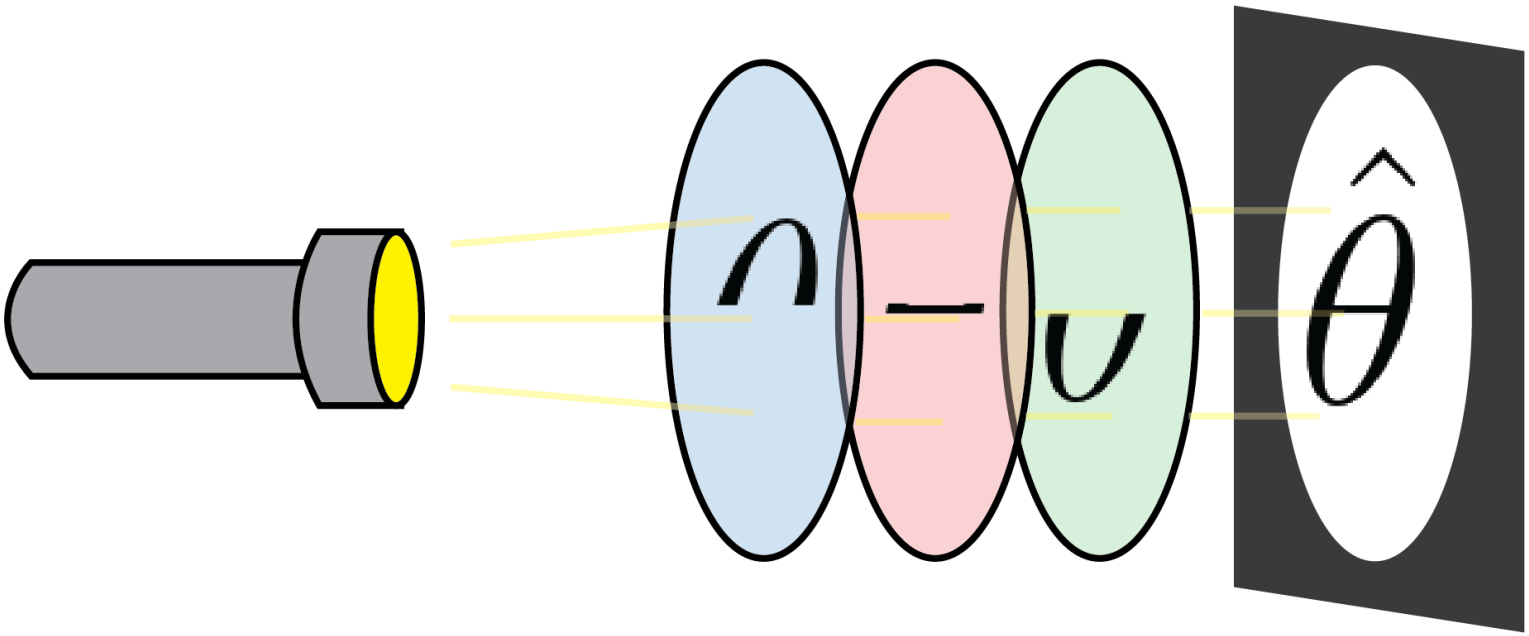
A single data source may be unable to adequately address pressing public health questions

- Measurement error, convenience sample

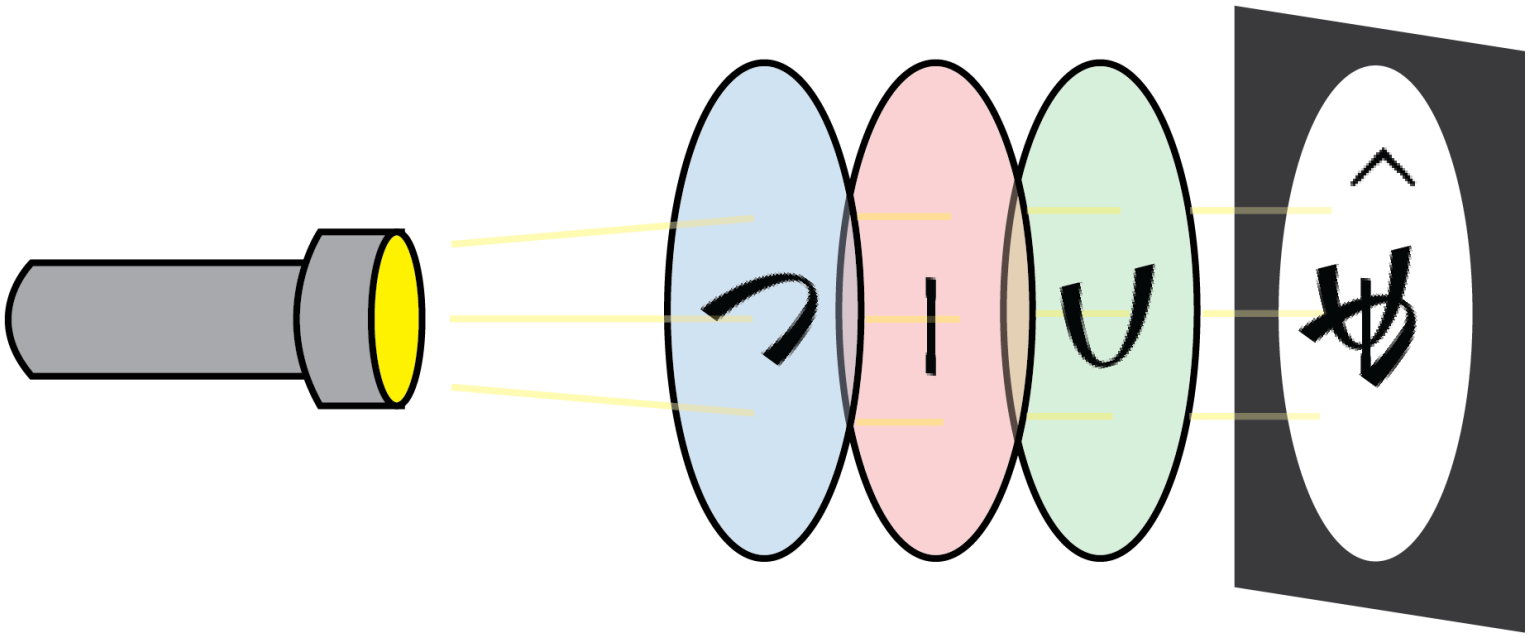
To address these biases

- Integrate information across multiple sources of information

# An Analogy



# An Analogy



# Historical Perspective

Use of multiple data sources has a long history

- Case-control studies[1]
- Validation studies for measurement error[2]
- Meta-analysis[3]
- Two-stage studies[4]
- Mathematical modeling[5]

[1] Lane-Claypon (1926) Reports on Public Health and Medical Subjects 32
[2] Rogan & Gladen (1978) *Am J Epidemiol*
[3] Glass (1978) *Educ Researcher*
[4] Zhao & Lipsitz (1992) *Stats in Med*
[5] Bernoulli (1776) *Mem Math Phy Acad Roy Sci Paris*

# Modern Perspective

Recently, approaches have been sharpened

- Lens of causal inference
- Generally proceeded under the generalizability[6] and transportability[7] framework
  - Acknowledgment of conditions to combine data sources
  - Alignment in flashlight analogy

[6] Cole & Stuart (2010) *Am J Epidemiol*

[7] Bareinboim & Pearl (2016). *PNAS*

# Purpose of the Session

Learn about recent developments in the integration of information across sources from a variety of perspectives

- Hear from different perspectives
- Lead to future improvements

# Speakers

Stephen Cole

- Combining sources of information using a data fusion perspective

Issa Dahabreh

- Combining information from multiple sources to learn about a target popoulation: meta-analysis and beyond

Eric Lofgren

- 'Just Bring Me Everything' Data Fusion in Mathematical Model Fitting and Parameterization

Leon Di Stefano

- Careful hierarchical Bayesian modelling for pooling information across trials

Summary & Panel Discussion