What would it take to change your inference? Tools to better communicate your statistical findings

Qinyun Lin, PhD

FMS 2025 spring annual meeting

Challenges when communicating our statistical findings (1)

- Shared understanding of robustness of evidence among varied stakeholders/different audiences
  - Many different research designs
  - Particularly difficult for small studies
- Unobserved/unmeasured confounding variables in non-experiments (observational study, quasi-experimental study)
- Missing data
- Measurement error

etc….

Challenges when communicating our statistical findings (2)

Some scenarios to consider:

- You receive a major revision for your manuscript and one reviewer asked about a potential confounding variable that you do not have data.

- You are in a conference where somebody presents some findings based on a small RCT. You wonder how strong the evidence is based on the RCT.

Smoking and lung cancer: recent evidence and a discussion of some questions*

Jerome Cornfield,[1] William Haenszel,[2] E. Cuyler Hammond,[3] Abraham M. Lilienfeld,[4] Michael B. Shimkin[5] and Ernst L. Wynder[6]

History of Sensitivity/Robustness Analyses

A dialogue with the public!

Summary This report reviews some of the more recent epidemiologic and experimental findings on the relationship of tobacco smoking to lung cancer, and discusses some criticisms directed against the conclusion that tobacco smoking, especially cigarettes, has a causal role in the increase in broncho-genic carcinoma. The magnitude of the excess lung-cancer risk among cigarette smokers is so great that the results can not be interpreted as arising from an indirect association of cigarette smoking with some other agent or characteristic, since this hypothetical agent would have to be at least as strongly associated with lung cancer as cigarette use; no such agent has been found or suggested. The consistency of all the epidemiologic and experimental evidence also supports the conclusion of a causal relationship with cigarette smoking, while there are serious inconsistencies in reconciling the evidence with other hypotheses which have been advanced. Unquestionably there are areas where more research is necessary, and, of course, no single cause accounts for all lung cancer. The information already available, however, is sufficient for planning and activating public health measures. – *J. Nat. Cancer Inst.* **22**:173–203, 1959.

"The first sensitivity analysis in an observational study was conducted by Cornfield, et al. [6] for certain observational studies of cigarette smoking as a cause of lung cancer; see also [10]. Although the tobacco industry and others had often suggested that cigarettes might not be the cause of high rates of lung cancer among smokers, that some other difference between smokers and nonsmokers might be the cause, Cornfield, et al. found that such an unobserved characteristic would need to **be a near perfect predictor of lung cancer and about nine times more common among smokers than among nonsmokers**. While this sensitivity analysis does not rule out the possibility that such a characteristic might exist, it does clarify what a scientist must logically be prepared to assert in order to defend such a claim."

Rosenbaum, P. R. (2005). Sensitivity analysis in observational studies. Encyclopedia of statistics in behavioral science, 4, 1809-1814.

Our Approach: What would it take to change an inference? (led by Dr. Kenneth A. Frank)

https://konfound-it.org/

KonFound-it!

Quantify the Robustness of Causal Inferences

Sensitivity analyses that quantify the robustness of inferences to concerns about omit

KONFOUND-IT WEBSITE

Powered by version 1.0.3 of the konfound R package.

Home | Resources

Specification

Step 1

Select type of outcome:

- Dichotomous
- Continuous

Step 2

Step 3

Step 4

Text Output

Graphic Outp

The Stata Journal: Promoting communications on statistics and Stata

Stata Press

Impact Factor: 4

Free access | Research article | First published online September 20, 2019

konfound: Command to quantify robustness of causal inferences

Ran Xu, Kenneth A. Frank, [...], and Joshua M. Rosenberg (+1) View all authors and affiliations

Volume 19, Issue 3 | https://doi.org/10.1177/1536867X19874223

ELSEVIER

Journal of Clinical Epidemiology

Volume 134, June 2021, Pages 150-159

JCE

Original Article

Hypothetical case replacement can be used to quantify the robustness of trial results

Kenneth A. Frank [a #], Qinyun Lin [b #], Spiro Maroulis [c #], Anna S. Mueller [d #], Ran Xu [e], Joshua M. Rosenberg [f], Christopher S. Hayter [c], Ramy A. Mahmoud [g], Marynia Kolak [b], Thomas Dietz [h], Lixin Zhang [i]

Show more

Educational Evaluation and Policy Analysis
December 2013, Vol. 35, No. 4, pp. 437–460
DOI: 10.3102/0162373713493129

What Would It Take to Change an Inference? Using Rubin's Causal Model to Interpret the Robustness of Causal Inferences

Kenneth A. Frank
Michigan State University

Spiro J. Maroulis
Arizona State University

Minh Q. Duong
Pacific Metrics Corporation

Benjamin M. Kelcey
University of Cincinnati

Overview of our approaches

- Sensitivity/Robustness analysis quantifies what it would take to change an inference based on hypothetical data or conditions
  - Unobserved covariates
  - Unobserved samples
  - ****Important**: researchers have tried their best to minimize the potential bias before conducting the sensitivity/robust analysis (no matter research design, model specification etc.)

Overview of our approaches

1. **Uncontrolled confounding variables:** How strong the **confounder(s)** needs to be?
   - Impact threshold for a confounding variable (ITCV)
2. **Non-random selection into a sample:** How different the **data** need to be? (Rubin's causal model and the counterfactual)
   - Robustness of an Inference to Replacement (RIR)

Overview of our approaches

1. Impact threshold for a confounding variable (ITCV)

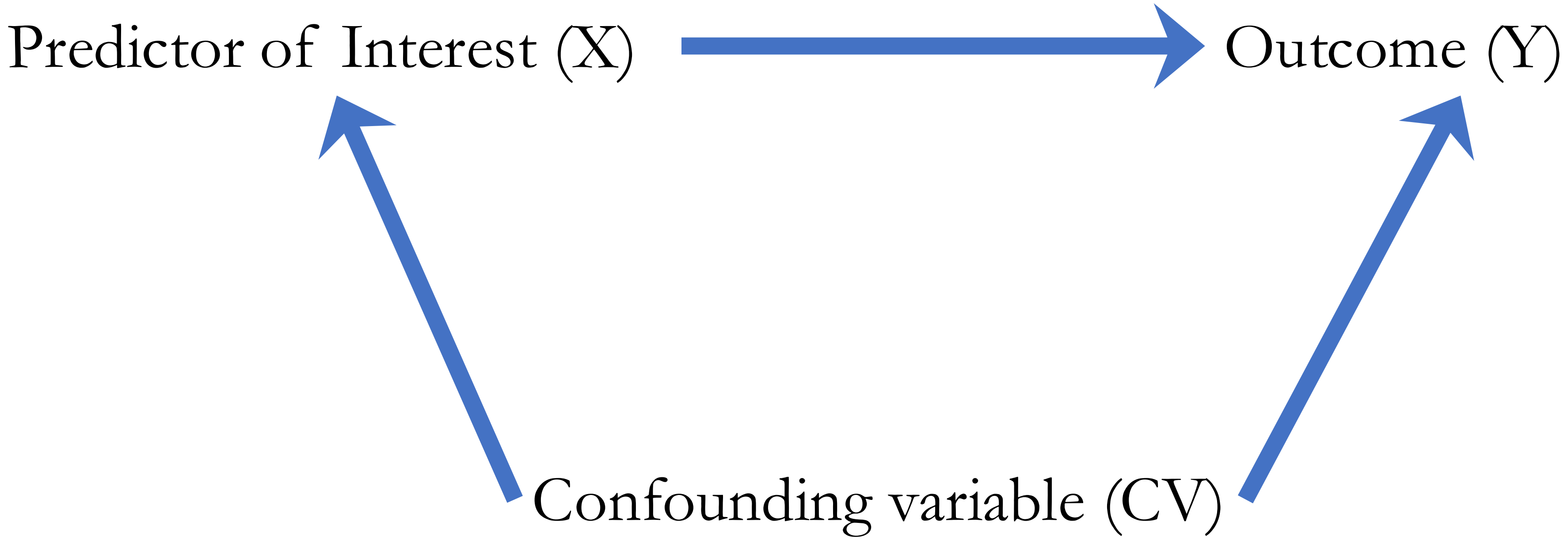
- Could be applied to: linear regression, mediation in a traditional framework
- “An omitted variable would have to be correlated at ___ with the predictor of interest and with the outcome to change the inference.”

Overview of our approaches

2. Robustness of an Inference to Replacement (RIR)

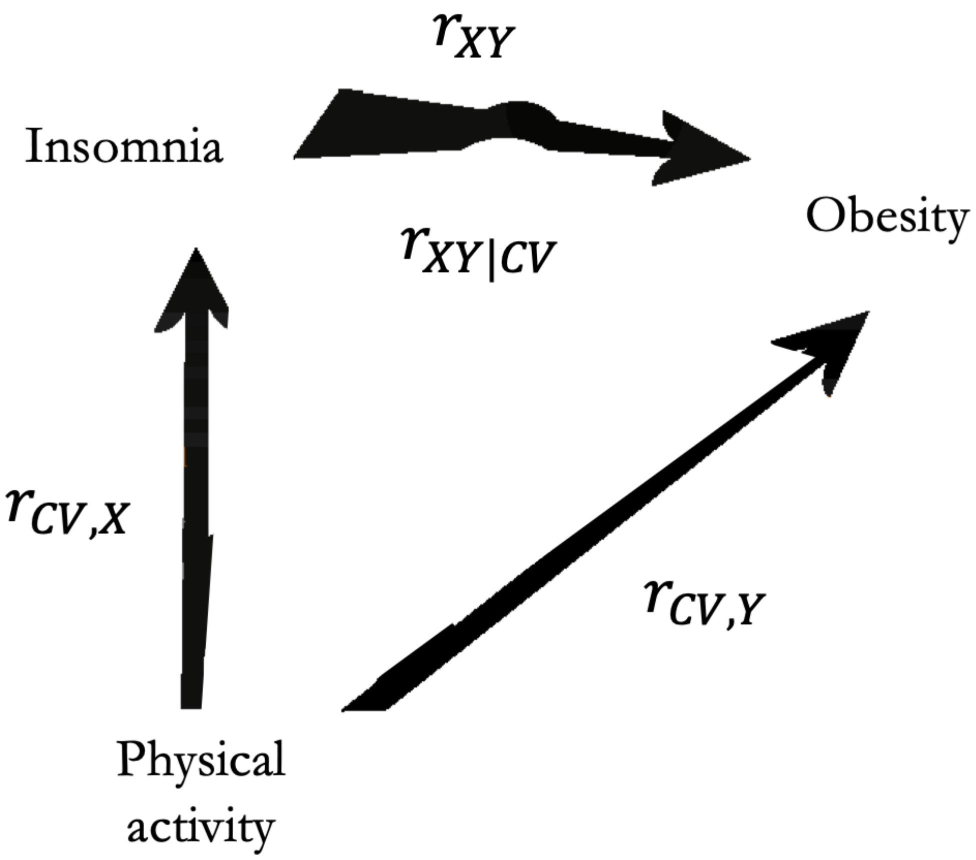
- Could be applied to: linear & logistic regression (including 2 by 2 table), moderation/interaction, spillover, mediation in a modern framework, survival analysis (work in progress)
- "To nullify the inference, __% of the data would have to be replaced with counterfactual data points for which the treatment had no effect."

1a. Understand confounding variables



1b. How Regression Works:

Impact of a Confounding Variable on a Regression Coefficient



Impact weights the relationship between CV and Y by the relationship between CV and X:

- the stronger the relationship between CV and Y, the more important the relationship between CV and X.
- Vice versa.

Impact of a confounder: to invalidate your inference and omitted variable would have to be correlated at __ with your predictor and outcome.

Note: we assume $r_{CV,X} = r_{CV,Y}$ to maximize the impact & favor the challenger of the inference (more conservative). Could also look at a curve where you allow these two correlations to be different.

Let's try this out in Rshiny.

Specification

Step 1

Select type of outcome:

- ○ Dichotomous
- ● Continuous

Step 2

Select source of data:

- ● Estimates from a linear model

Step 3

Select type of analysis:

- ● **ITCV:** Impact Threshold for a Confounding Variable *(Basic Analysis)*
- ○ **RIR:** Generalized Robustness of Inference to Replacement *(Basic Analysis)*
- ○ Preserve Standard Error *(Advanced Analysis)*
- ○ Coefficient of Proportionality *(Advanced Analysis; in beta)*

Step 4

Enter these values:

Estimated Effect

2

Standard Error

0,4

Number of Observations

100

Number of Covariates

3

Note that decimals must be denoted with a period, e.g., 2.1

RUN

Quick Example

Economic Connectedness and Upward Mobility

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > articles > article

Article | Open Access | Published: 01 August 2022

Social capital I: measurement and associations with economic mobility

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole & Nils Wernerfelt

Nature (2022) | Cite this article

Abstract

Social capital—the strength of an individual's social network and community—has been identified as a potential determinant of outcomes ranging from education to health[1,2,3,4,5,6,7,8]. However, efforts to understand what types of social capital matter for these outcomes have been hindered by a lack of social network data. Here, in the first of a pair of papers[9], we use data on 21 billion friendships from Facebook to study social capital. We measure and analyse three types of social capital by ZIP (postal) code in the United States: (1) connectedness between different types of people, such as those with low versus high socioeconomic status (SES); (2) social cohesion, such as the extent of cliques in friendship networks; and (3) civic engagement, such as rates of volunteering. These measures vary substantially across areas, but are not highly correlated with each other. We demonstrate the importance of distinguishing these forms of social capital by analysing their associations with economic mobility across areas. The share of high-SES friends among individuals with low SES—which we term economic connectedness—is among the strongest predictors of upward income mobility identified to date[10,11]. Other social capital measures are not strongly associated with economic mobility. If children with low-SES parents were to grow up in counties with economic connectedness comparable to that of the average child with high-SES parents, their incomes in adulthood would increase by 20% on average. Differences in economic connectedness can explain well-known relationships between upward income mobility and racial segregation, poverty rates, and inequality[12,13,14]. To support further research and policy interventions, we publicly release privacy-protected statistics on social capital by ZIP code at https://www.socialcapital.org.

Quick Example: with [Konfound-it](): [Economic Connectedness and Upward Mobility]()

The share of high-SES friends among individuals with low SES—which we term economic connectedness—is among the strongest predictors of upward income mobility identified to date[10,11]

Step 4

Enter these values (Note that decimals must be denoted with a period, e.g., 2.1):

Estimated Effect

0.548

Standard Error

0.038

Number of Observations

24165

Number of Covariates

1

RUN

Table 2 | Associations between upward income m

EC versus median income and poverty rates	
Dependent variable	**Upward income mobility**
	(6)
Median income	0.209***
	(0.029)
Poverty rate	–
	–
Economic connectedness	0.548***
	(0.038)
Observations	24,165
R^2	0.496

Results

Text Output

Impact Threshold for a Confounding Variable (ITCV):

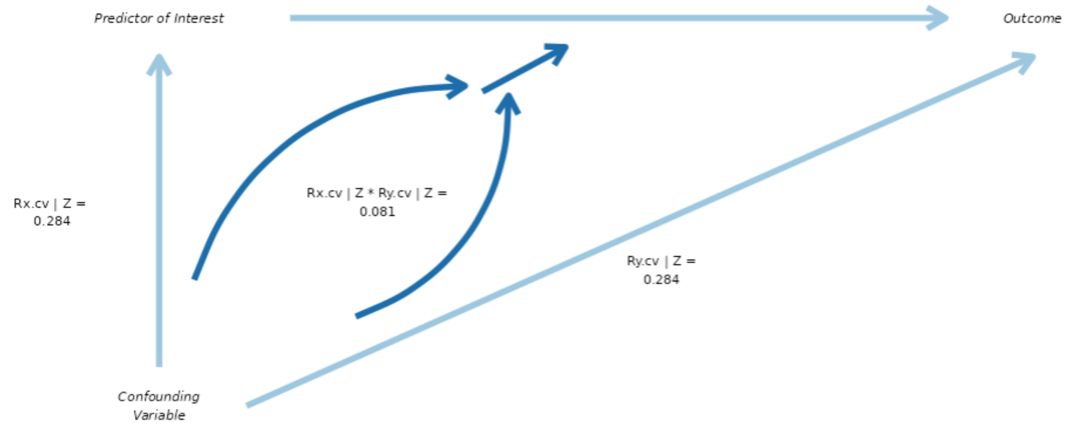
The minimum impact of an omitted variable to invalidate an inference for a null hypothesis of an effect of nu (0) is based on a correlation of 0.284 with the outcome and 0.284 with the predictor of interest (conditioning on all observed covariates in the model; signs are interchangeable). This is based on a threshold effect of a threshold effect of 0.0 for statistical significance (alpha = 0.05).

Correspondingly, the impact of an omitted variable (as defined in Frank [2000]) must be 0.284 X 0.284 = 0.081 to invalidate an inference for a null hypothesis of an effect of nu (0).

For calculation of unconditional ITCV using pkonfound(), additionally include the R^2, sd_x, and sd_y as input, and request raw output.

Graphic Output

To invalidate an inference



Would you like to generate source code?

Generate R Code

```
#install.packages('konfound')
library(konfound)  # konfound R package version: 1.0.3
pkonfound(0.548, 0.038, 24165, 3, index = 'IT')
```

COPY R CODE

Generate Stata Code

```
ssc install konfound
ssc install indeplist
ssc install moss
ssc install matsort
pkonfound 0.548 0.038 24165 3, model_type(0) indx(IT)
```

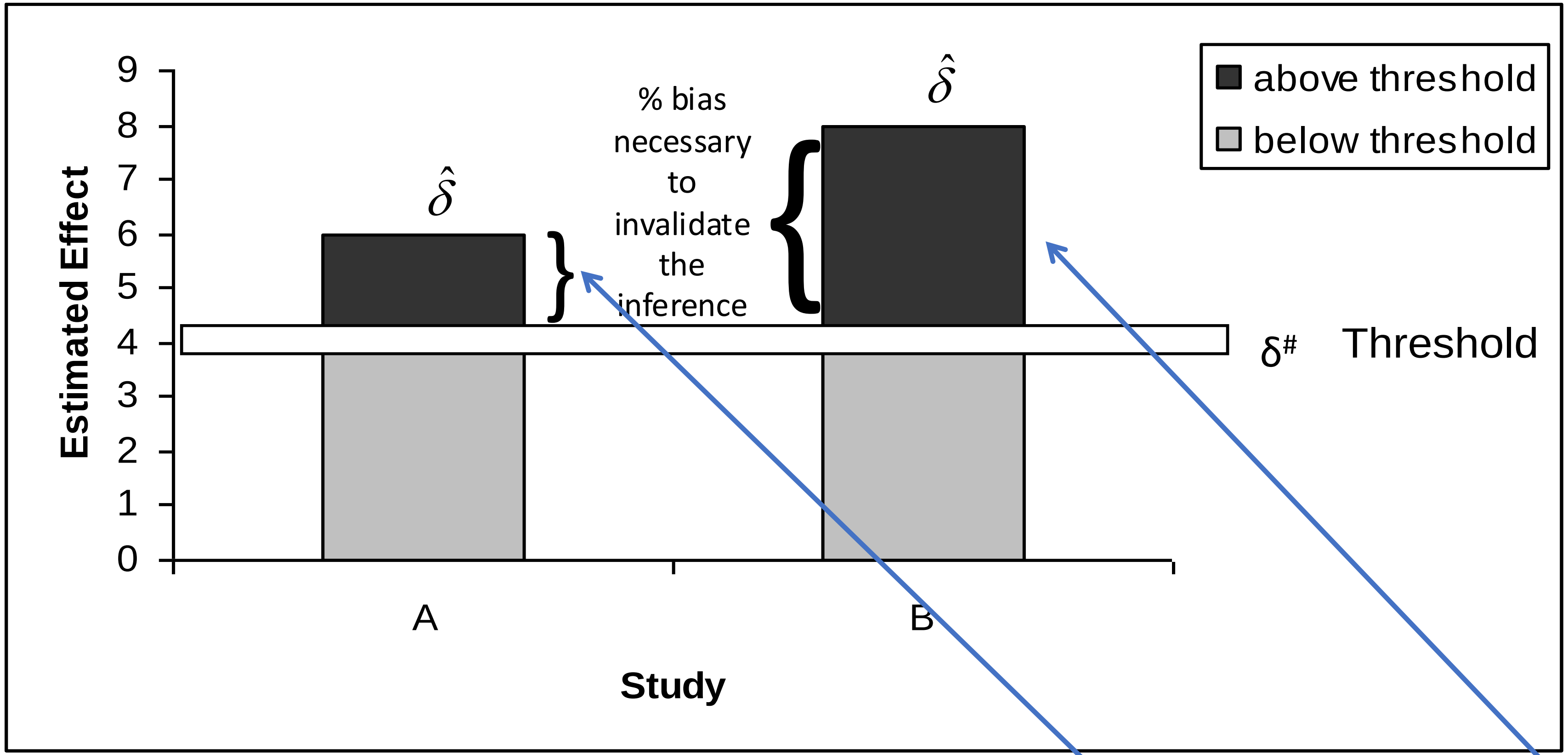
COPY STATA CODE

2. Replacement of Cases Framework

How much bias must there be to invalidate an inference?

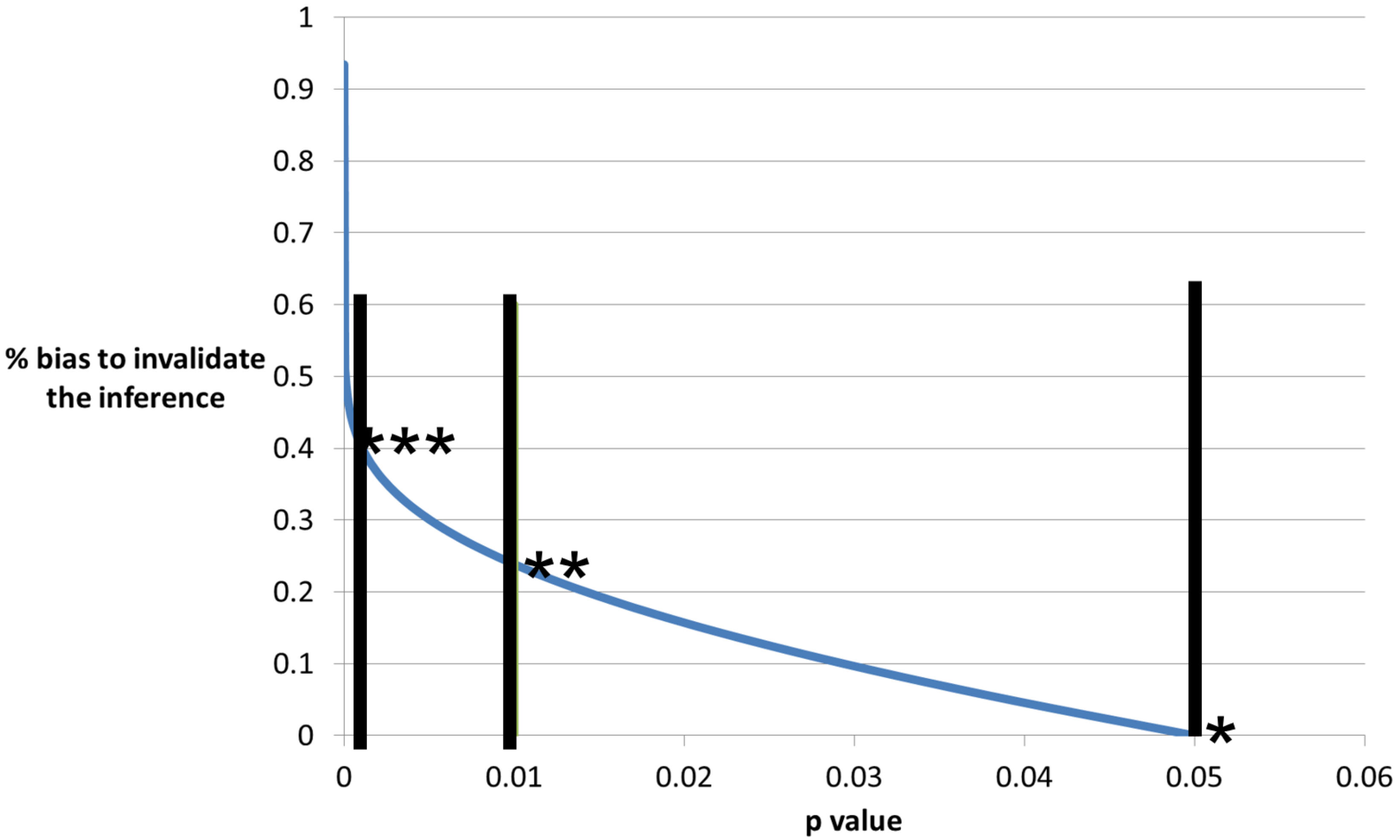
- Concerns about Internal Validity
  - What percentage of data points would you have to replace with counterfactual data points (with zero effect) to invalidate the inference?
- Concerns about External Validity
  - What percentage of data points would you have to replace with cases from an unsampled population (with zero effect) to invalidate the inference?

Figure 1
Estimated Treatment Effects in Hypothetical Studies A and B Relative to a Threshold for Inference $\delta^{\#}$



$$\% \ bias(\hat{\delta}) \text{ to invalidate} = \frac{(\hat{\delta} - \delta^{\#})}{\hat{\delta}} = 1 - \frac{\delta^{\#}}{\hat{\delta}} = 1 - \frac{4}{6} = \frac{1}{3} = 33\% \qquad 1 - \frac{4}{8} = \frac{1}{2} = 50\%$$

% Bias to Invalidate versus p-value: a better language?



(this plot is for df>500, but the curve is almost identical for smaller df; the sample size only affects the $t_{critical}$ used to calculate $r^{\#}$)

Framework for Interpreting % Bias to Invalidate an Inference: Rubin's Causal Model and the Counterfactual

1) I have a headache
2) I take an aspirin (treatment)
3) My headache goes away (outcome)

Q: Is it because I took the aspirin?

A: We'll never know – it is counterfactual – for the individual

- This is the Fundamental Problem of Causal Inference

Approximating the Counterfactual with Observed Data

	A	B	C	D	E
1			potential outcome		
2			Treatment	Control	
3	Unit	_	Y^t	Y^c	Effect
4	1	t	9	8	1
5	2	t	10	9	1
6	3	t	11	10	1
7	4	c	?	3	
8	5	c	?	4	
9	6	c	?	5	
10	Mean		10.00	9	6
11					
12	counterfactual				
13	Observed				

But how well does the observed data approximate the counterfactual? Difference between counterfactual values and observed values for the control implies the **true treatment effect of 1**

is overestimated as 6 using observed control cases with mean of 4

Fundamental problem of causal inference is that we cannot simultaneously observe Y_i^t and Y_i^c

Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81:945_70. *(25-40)*

Using the Counterfactual to Interpret % Bias to Invalidate the Inference: Replacement with Average Values

The inference would be invalid if you replaced 33% (or 2 data points) with counterfactuals for which there was no treatment effect.

	A	B	C	D	E
1			potential outcome		
2			Treatment	Control	
3	Unit		Y^t	Y^c	Effect
4	1	t	9	7	0
5	2	t	10	7	0
6	3	t	11	7	0
7	4	c	7	3	######
8	5	c	7	4	######
9	6	c	7	5	######
10	Mean		9	5	4
11					
12	counterfactual				
13	Observed				

How many data points would you have to replace with zero effect counterfactuals to change the inference? Assume threshold is 4

($\delta^{\#}$ =4): $1-\frac{\delta^{\#}}{\hat{\delta}}=1-\frac{4}{6}=0.33$

$$New\ estimate = (1 - \%replaced) \cdot \hat{\delta} + \%replaced \cdot no\ effect = Threshold\ (\delta^{\#})$$

Essentially, we are asking how bad the approximation needs to be, or how different the observed and the counterfactual needs to be to alter the inference?

Which Cases to Replace?

- Expectation: if you randomly replaced 1/3 of the data points, and repeated 1,000 times, on average the new estimate would be 4
- Assumes constant treatment effect
- Conditioning on covariates and interactions in model
- Assumes data points carry equal weight
- Extensions include selective replacement, spillover, weighted observations, logistic, “causal” designs (e.g., RD)

Let’s try this out in Rshiny.

Specification

Step 1

Select type of outcome:

- ○ Dichotomous
- ◉ Continuous

Step 2

Select source of data:

- ◉ Estimates from a linear model

Step 3

Select type of analysis:

- ○ **ITCV:** Impact Threshold for a Confounding Variable *(Basic Analysis)*
- ◉ **RIR:** Generalized Robustness of Inference to Replacement *(Basic Analysis)*
- ○ Preserve Standard Error *(Advanced Analysis)*
- ○ Coefficient of Proportionality *(Advanced Analysis; in beta)*

Step 4

Enter these values:

Estimated Effect

2

Standard Error

0,4

Number of Observations

100

Number of Covariates

3

Note that decimals must be denoted with a period, e.g., 2.1

RUN

Quick Example

Economic Connectedness and Upward Mobility

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature > articles > article

Article | Open Access | Published: 01 August 2022

Social capital I: measurement and associations with economic mobility

Raj Chetty, Matthew O. Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B. Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, Drew Johnston, Martin Koenen, Eduardo Laguna-Muggenburg, Florian Mudekereza, Tom Rutter, Nicolaj Thor, Wilbur Townsend, Ruby Zhang, Mike Bailey, Pablo Barberá, Monica Bhole & Nils Wernerfelt

Nature (2022) | Cite this article

Abstract

Social capital—the strength of an individual's social network and community—has been identified as a potential determinant of outcomes ranging from education to health[1,2,3,4,5,6,7,8]. However, efforts to understand what types of social capital matter for these outcomes have been hindered by a lack of social network data. Here, in the first of a pair of papers[9], we use data on 21 billion friendships from Facebook to study social capital. We measure and analyse three types of social capital by ZIP (postal) code in the United States: (1) connectedness between different types of people, such as those with low versus high socioeconomic status (SES); (2) social cohesion, such as the extent of cliques in friendship networks; and (3) civic engagement, such as rates of volunteering. These measures vary substantially across areas, but are not highly correlated with each other. We demonstrate the importance of distinguishing these forms of social capital by analysing their associations with economic mobility across areas. The share of high-SES friends among individuals with low SES—which we term economic connectedness—is among the strongest predictors of upward income mobility identified to date[10,11]. Other social capital measures are not strongly associated with economic mobility. If children with low-SES parents were to grow up in counties with economic connectedness comparable to that of the average child with high-SES parents, their incomes in adulthood would increase by 20% on average. Differences in economic connectedness can explain well-known relationships between upward income mobility and racial segregation, poverty rates, and inequality[12,13,14]. To support further research and policy interventions, we publicly release privacy-protected statistics on social capital by ZIP code at https://www.socialcapital.org.

Quick Example: with Konfound-it: Economic Connectedness and Upward Mobility

The share of high-SES friends among individuals with low SES—which we term economic connectedness—is among the strongest predictors of upward income mobility identified to date[10,11]

Step 4

Enter these values (Note that decimals must be denoted with a period, e.g., 2.1):

Estimated Effect

0.548

Standard Error

0.038

Number of Observations

24165

Number of Covariates

1

RUN

Table 2 | Associations between upward income m

EC versus median income and poverty rates	
Dependent variable	**Upward income mobility**
	(6)
Median income	0.209***
	(0.029)
Poverty rate	–
	–
Economic connectedness	0.548***
	(0.038)
Observations	24,165
R^2	0.496

Results

Text Output

Robustness of Inference to Replacement (RIR):
RIR = 20881

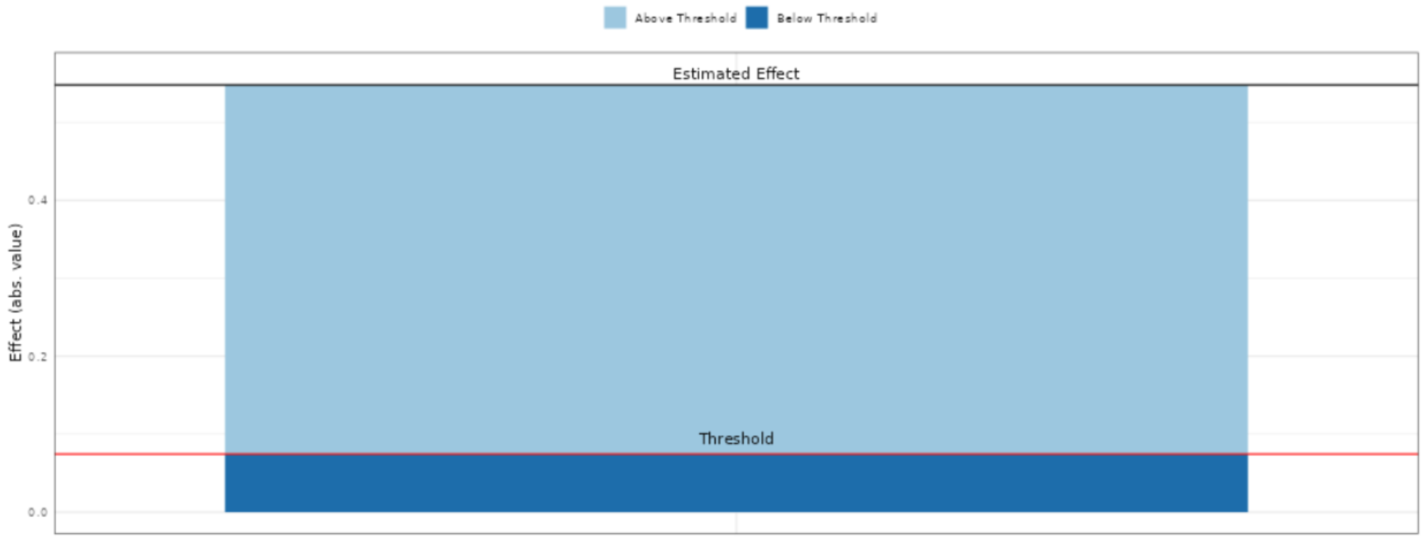
To invalidate the inference of an effect using the threshold of 0.074 for statistical significance (with null hypothesis = 0 and alpha = 0.05), 86.408% of the (2) estimate would have to be due to bias. This implies that to invalidate the inference one would expect to have to replace 20881 (86.408%) observations with data points for which the effect is 0 (RIR = 20881).

Use `to_return = "raw_output"` to see more specific results.

$$RIR = 20881$$

See Frank et al. (2013) for a description of the method.

Graphic Output



Would you like to generate source code?

Generate R Code

```
#install.packages('konfound')
library(konfound)  # konfound R package version: 1.0.3
pkonfound(0.548, 0.038, 24165, 3, index = 'RIR')
```

COPY R CODE

Generate Stata Code

```
ssc install konfound
ssc install indeplist
ssc install moss
ssc install matsort
pkonfound 0.548 0.038 24165 3, model_type(0) indx(RIR)
```

COPY STATA CODE

Evaluation of % Bias to Nullify Inference

- **Internal Benchmark:** Compare bias necessary to nullify inference with bias accounted for by *background characteristics*
  - 1% of estimated effect accounted for by background characteristics (including mother's education), once controlling for pretests (estimated effect from -9.1 to -9.01)
  - Estimate would have to change another 85% to nullify the inference.
- Interpret as a **probability from a Bayesian perspective**
  - *Frank, K. A. and *Min, K. 2007. Indices of Robustness for Sample Representation. *Sociological Methodology*. Vol 37, 349-392. * co first authors.
  - Li, Tenglong, Frank, K.A., (forthcoming). On the probability an inference is robust for internal validity. *Sociological Methods and Research*.
- **External Benchmark**: Compare with % bias necessary to nullify inference in *other studies*:
  - Use correlation metric: Adjusts for differences in scale
  - See new konfound-it web site

Summary: What Would it Take to Change your Inference?

It's causal inference: you might be wrong! But we could talk about "what would it take to change your inference."

- Impact of a confound: to invalidate your inference an omitted variable would have to be correlated at __ with your predictor and outcome.
- Case replacement: to invalidate your inference, you would have to replace __% of your data points with null effect data points

→ Our approaches could be applied to linear regression, logistic regression, mediation, multilevel models, spillover effects, survival analysis (work in progress). R and Stata packages available.

→ Other relevant techniques: preserve standard error, coefficient of proportionality

→ You could use such statements to better communicate your statistical findings, e.g., when reviewer raised some concern about unmeasured confounders.

KonFound-it!

Home Tools ▼ Support ▼ News About

KonFound-It!

Quantify the Robustness of Causal Inferences

The assumptions underlying statistical analysis are rarely fully met. Pragmatists face the challenge of knowing when evidence is strong enough to justify action, and that's where sensitivity analysis helps by testing the robustness of inferences against potential biases. For instance, sensitivity indices can quantify how much of the observed effect would need to be bias-driven to alter conclusions. We are building on these ideas by developing the methods and associated resources listed on this page.

Questions? Issues? Suggestions? Reach out through the KonFound-It! Google Group.

Start KonFounding

Try out KonFound-It! to calculate sensitivity analyses through an interactive web app.

KonFound-It! News

Read the latest updates from the KonFound-It! team.

Thank you for your time!

qinyun.lin@gu.se

Back up slides start here

Sensitivity Analysis:
What Must be the Impact of an **Unmeasured Confounding variable** invalidate the Inference?

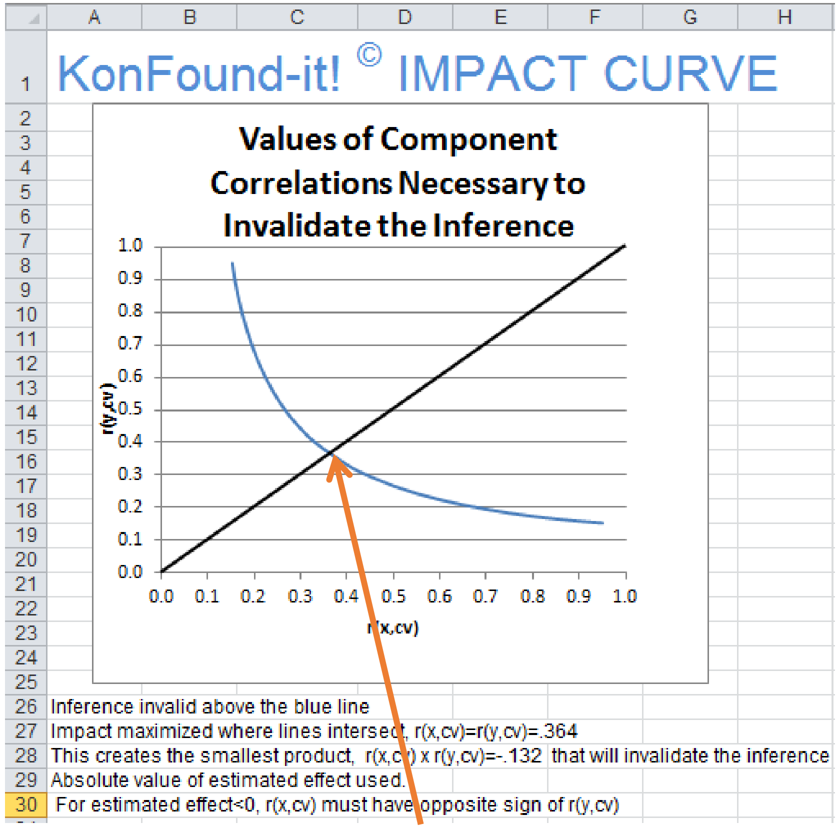
Step 1: Establish Correlation Between predictor of interest and outcome

Step 2: Define a Threshold for Inference

Step 3: Calculate the Impact Necessary to Invalidate the Inference

Step 4: Multivariate Extension, with other Covariates

Key of confounders: must have both arms



Smallest impact to invalidate inference: $r_{x,cv}=r_{y.cv}=.364$

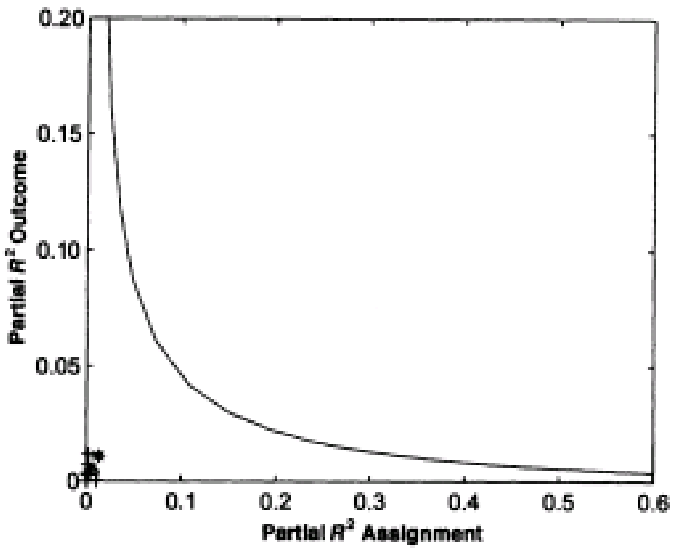


FIGURE 1. LALONDE EXPERIMENTAL SAMPLE

Imbens, G. W. (2003). Sensitivity to exogeneity assumptions in program evaluation. *American Economic Review*, *93*(2), 126-132.

Carnegie N.B., Harada M., Hill J.L. Assessing Sensitivity to Unmeasured Confounding Using a Simulated Potential Confounder. (2016) *Journal of Research on Educational Effectiveness*, 9(3) , pp. 395-420.

Evaluation of % Bias Necessary to Invalidate Inference

- Pragmatic, contentious:
  - 50% cut off– for every case you remove, I get to keep one

- Compare bias necessary to invalidate inference with bias accounted for by background characteristic
  - 1% of estimated effect accounted for by background characteristics (including mother's education), once controlling for pretests.
  - Other sources of bias would have to be 85 times more important than background characteristics

- Compare with % bias necessary to invalidate inference in other studies. Use correlation metric: Adjusts for differences in scale

Beyond *, **, and ***

- P values
  - sampling distribution framework
  - Must interpret relative to standard errors
  - Information lost for modest and high levels of robustness
- % bias to invalidate
  - counterfactual framework
  - Interpret in terms of case replacement
  - Information along a continuous distribution

Quick example on 2 by 2 table from RCT

	Exacerbated or Unchanged	Improved (Moderate or Significant)	Total
Conventional Treatment	14	17	31
Hydroxychloroquine	6 + 1 = 7	25 – 1 = 24	31
Total	20	42	62

1. Replace 3 data points from *“Improved HCQ”* group with cases for whom HCQ has no effect (zero effect data points)
2. Based on the control group, $\frac{14}{31}$ ≈ 54.84% cases experienced “Exacerbated or Unchanged”. → 1 data points out of the 2 replacement data points goes to *“Exacerbated or Unchanged HCQ”* group
3. NOT significant anymore (p changes from 0.03 to 0.06)

RIR = 2, Fragility = 1

- Frank, K. A.#, **Lin, Q.#**, Maroulis S. J.#, Strassman, A.#, Xu R., Rosenberg J., Hayter, C., Mahmoud, R., Kolak, M., & Dietz, T. (2021). Hypothetical case replacement can be used to quantify the robustness of trial results. *Journal of Clinical Epidemiology*, *134*, 150–159. https://doi.org/10.1016/j.jclinepi.2021.01.025