Sensitivity Analysis for Unmeasured Confounding and Powerful Test Identification

How can we draw credible causal conclusions from observational data?

Elaine Chiu (University of Wisconsin–Madison)

March 18, 2026


Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

About Me

Education Background

- Economics & Political Science $\rightarrow$ Statistics

Hobby

- Music lover — from classical to modern (flute; favorite band: Accusefive)

Skills

- **Methods:** Causal inference, Bayesian methods
- **Data analysis:** education intervention & achievement; tobacco exposure & lung function; social distancing & COVID-19 deaths
- **Tools:** R (`sensitivityIxJ`), R, C++, Python
- **Communication:** Consulting, mentoring


Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

Motivation: Randomized Trials

- Random assignment of treatment
- Comparable treatment groups
- Outcome differences $\rightarrow$ causal effect

Motivation: Observational Studies

- Treatment not randomly assigned
- Treatment groups may not be comparable
- Outcome differences $\neq$ causal effect

Motivation: Hybrid Controls in Industry

Hybrid control: Augments the randomized control with observational data

- Difficulty in control-arm recruitment
- Rare populations (limited sample size)

Paper	**Affiliations**
Investigational Use of Real-World Data as a Hybrid Control in Pancreatic Ductal Adenocarcinoma From the Randomized Phase Ib/II MORPHEUS Trial	Genentech; Gilead; GSK; Pfizer
Beyond Randomized Clinical Trials: Use of External Controls	Novartis
Uncontrolled Extensions of Clinical Trials and the Use of External Controls—Scoping Opportunities and Methods	J&J; IQVIA; Merck

Have we considered unmeasured confounding?

Motivation of Sensitivity Analysis in Hypothesis Testing

Type I Error Rate Control

- Type I error rate control is essential in confirmatory studies[1]
- Requires sensitivity analysis to address unmeasured confounding

Sensitivity Analysis Beyond Binary Treatments

- Most sensitivity analyses focus on **binary treatments**
- Extensions to:
  - **Categorical treatments**
    none, drink, smoke, drink & smoke vs. impaired glucose metabolism
  - **Continuous treatments**
    air pollution intensity vs. mortality risk

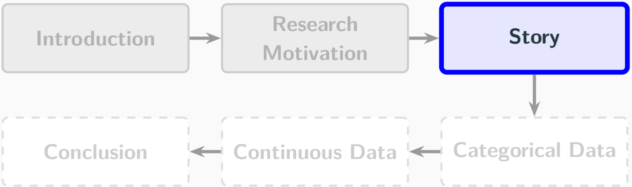
[1] FDA (2017), *Multiple Endpoints in Clinical Trials*.

A Note on Notation

$$Y = f(X) + \epsilon$$

- **Gray boxes** denote formal mathematical expressions.
- Feel free to **skip** them if they block intuition.

- Boxes represent variables.
- $A \to B$ indicates a **causal effect**.


Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

Back in 1958: Does Smoking Cause Lung Cancer?

E 9

SCIENCE IN REVIEW

Statistics Linking Cigarette Smoking With Lung Cancer Are Seriously Challenged

By WILLIAM L. LAURENCE

A serious challenge to the claim, based solely on statistical data, that there is a direct causal relationship between cigarette smoking and lung cancer is presented in the current issue of the Journal of the American Statistical Association. The challenger is Dr. Joseph Berkson, head of the Division of Biometery and Medical Statistics at the Mayo Clinic, Rochester, Minn., an internationally known medical statistician.

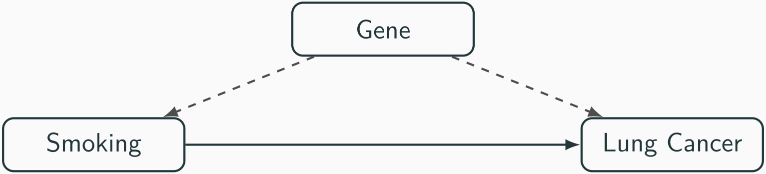
The assertion that there is a direct cause-and-effect relationship between cigarette smoking and the great increase in recent years in the incidence of lung cancer is based

in all the various categories, Dr. Berkson states, is that "persons who are non-smokers, or relatively light smokers, are of a constitutional type that is biologically disposed to self-protective habits, and that this is correlated generally with constitutional resistance to mortal forces from disease.

"If 85 to 95 per cent of a population are smokers," he asserts, "then the small minority who are not smokers would appear, on the face of it, to be of some special type of constitution. It is not implausible that they should be, on the average, relatively longevous, and this implies that death rates

Smoking and Lung Cancer: Arguments & Causal Graph

- **Evidence:** higher risk among smokers
- **Objection (Fisher):** an unmeasured confounder, a hidden gene



Smoking and Lung Cancer: Evidence

Observed association between smoking and lung cancer:

$$\text{RR}_{\text{cancer,smoke}} = \frac{\mathbb{P}(\text{cancer} \mid \text{smoker})}{\mathbb{P}(\text{cancer} \mid \text{non-smoker})} \approx 9$$

Smoking and Lung Cancer: Calculation and Implication

$$\mathrm{RR}_{\text{cancer, smoke}} = \frac{\mathbb{P}(\text{cancer} \mid \text{smoker})}{\mathbb{P}(\text{cancer} \mid \text{non-smoker})}$$

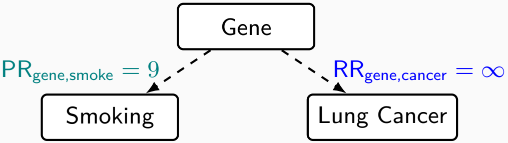
$$\mathrm{PR}_{\text{gene, smoke}} = \frac{\mathbb{P}(\text{gene} \mid \text{smoker}) / \mathbb{P}(\text{no gene} \mid \text{smoker})}{\mathbb{P}(\text{gene} \mid \text{non-smoker}) / \mathbb{P}(\text{no gene} \mid \text{non-smoker})}$$

$$\mathrm{RR}_{\text{gene,cancer}} = \frac{\mathbb{P}(\text{cancer} \mid \text{gene})}{\mathbb{P}(\text{cancer} \mid \text{no gene})}$$

Observed Evidence

$\mathrm{RR}_{\text{cancer, smoke}} = 9$ $=$

If smoking has no causal effect:



Implication

If a $\mathrm{PR}_{\text{gene,smoke}}$ of 9 is not feasible, the causal link remains credible.

Smoking and Lung Cancer: Takeaway

Association might reveal causation if we discuss the magnitude of unmeasured confounding.

General Definition of Sensitivity Analysis

If we allow unmeasured confounding to affect treatment assignment to some extent, would the causal conclusion change?


Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

Sensitivity Analysis in Hypothesis Testing with Categorical Data

"Exact, Nonparametric Sensitivity Analysis for Observational Studies of Contingency Tables,"
under major revision at
Journal of the American Statistical Association (Theory and Methods).

Lifestyle and New-Onset Impaired Glucose Metabolism

Table 1: Lifestyle Habits and NO-IGM [2]

AIP Patient Status	NO-IGM	No NO-IGM
Neither smoking nor drinking	31	112
Smoking only	54	48
Drinking only	4	4
Both smoking and drinking	35	17

Why can't we just do the usual hypothesis testing?

→ Unmeasured confounder: poor health awareness

[2] Xi, W., et al. (2025), *Smoking, alcohol consumption, and new-onset impaired glucose metabolism in male patients with type 1 autoimmune pancreatitis: a retrospective cohort study*. Therapeutic Advances in Chronic Disease.

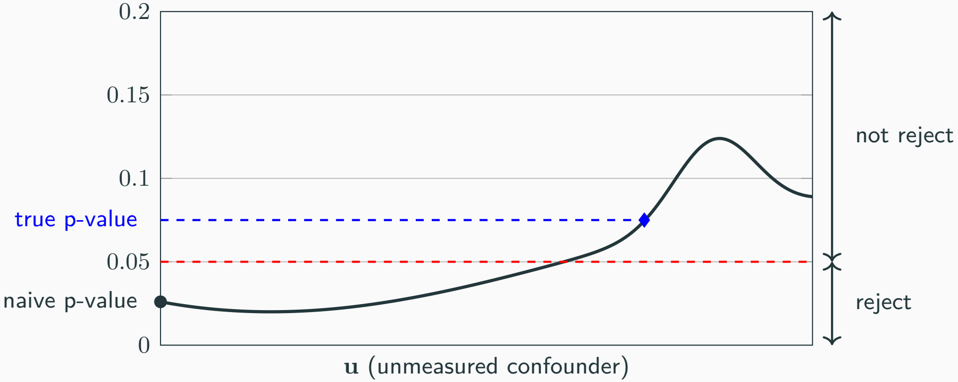
General Review: Hypothesis Testing

	Reject H_0	**Fail to reject H_0**
H_0 **is true**	**Type I error**	Correct decision
H_0 **is false**	Correct decision	**Type II error**

- **P-value:** Probability of data as extreme as observed, given H_0
- **Decision Rule:** Reject H_0 if $p < \alpha$

Unmeasured Confounding and Type I Error Rate Inflation

- **Standard Rule:** Reject H_0 if $p < \alpha$
- **Hidden Bias:** p is underestimated → Type I error inflated
- **Sensitivity Rule:** Reject H_0 only if $p_{\text{upper}} < \alpha$

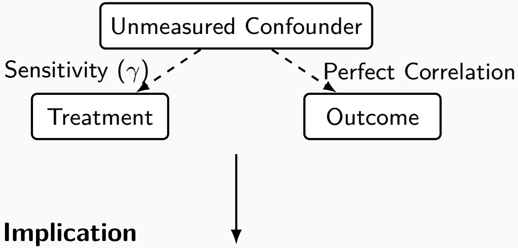


Categorical Data Sensitivity Analysis in Hypothesis Testing

Observed Evidence

Observed Test Value =

If there is no causal effect (H_0 is true):



Implication

Causal conclusion is robust if the p_{upper} is small at large γ

Unmeasured Confounder → Treatment Assignment (follow the cue)

s : subject index, $s \in \{1, \ldots, N\}$
i : treatment index, $i \in \{1, \ldots, I\}$
Z_s : treatment received by subject s
u_s : unmeasured confounder, $u_s \in [0, 1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Model setup**

$$\delta_i = \begin{cases} 1, & \text{favored by } u \\ 0, & \text{otherwise} \end{cases} \qquad e^{\gamma} = \text{Odds Ratio bound}$$

- **Implication,** $\delta_i = 0, \delta_{i'} = 1$**:**

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \leq e^{\gamma}$$

- **Example**: $\delta_{\text{neither}} = 0, \delta_{\text{smoking}} = \delta_{\text{drinking}} = 1, \gamma = \ln(3) \implies$ unmeasured confounder can triple the unhealthy habit odds.

Unmeasured Confounder $\to$ Treatment Assignment (follow the cue)

s	: subject index, $s \in \{1, \dots, N\}$
i	: treatment index, $i \in \{1, \dots, I\}$
Z_s	: treatment received by subject s
u_s	: unmeasured confounder, $u_s \in [0, 1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Model setup**

$$\delta_i = \begin{cases} 1, & \text{favored by } u \\ 0, & \text{otherwise} \end{cases} \qquad e^{\gamma} = \text{Odds Ratio bound}$$

- **Implication,** $\delta_i = 0, \delta_{i'} = 1$**:**

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \leq e^{\gamma}$$

- **Example**: $\delta_{\text{neither}} = 0, \delta_{\text{smoking}} = \delta_{\text{drinking}} = 1, \gamma = \ln(3) \implies$ unmeasured confounder can triple the unhealthy habit odds.

Unmeasured Confounder $\to$ Treatment Assignment (follow the cue)

s : subject index, $s \in \{1, \dots, N\}$
i : treatment index, $i \in \{1, \dots, I\}$
Z_s : treatment received by subject s
u_s : unmeasured confounder, $u_s \in [0, 1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Model setup**

$$\delta_i = \begin{cases} 1, & \text{favored by } u \\ 0, & \text{otherwise} \end{cases} \qquad e^\gamma = \text{Odds Ratio bound}$$

- **Implication,** $\delta_i = 0, \delta_{i'} = 1$:

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \le e^\gamma$$

- **Example**: $\delta_{\text{neither}} = 0, \delta_{\text{smoking}} = \delta_{\text{drinking}} = 1, \gamma = \ln(3) \implies$ unmeasured confounder can triple the unhealthy habit odds.

Unmeasured Confounder → Treatment Assignment (follow the cue)

s : subject index, $s \in \{1, \dots, N\}$
i : treatment index, $i \in \{1, \dots, I\}$
Z_s : treatment received by subject s
u_s : unmeasured confounder, $u_s \in [0, 1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Model setup**

$$\delta_i = \begin{cases} 1, & \text{favored by } u \\ 0, & \text{otherwise} \end{cases} \qquad e^{\gamma} = \text{Odds Ratio bound}$$

- **Implication, $\delta_i = 0, \delta_{i'} = 1$:**

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \leq e^{\gamma}$$

- **Example**: $\delta_{\text{neither}} = 0, \delta_{\text{smoking}} = \delta_{\text{drinking}} = 1, \gamma = \ln(3) \implies$ unmeasured confounder can triple the unhealthy habit odds.

Unmeasured Confounder → Treatment Assignment (follow the cue)

s	: subject index, $s \in \{1, \ldots, N\}$
i	: treatment index, $i \in \{1, \ldots, I\}$
Z_s	: treatment received by subject s
u_s	: unmeasured confounder, $u_s \in [0, 1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Model setup**

$$\delta_i = \begin{cases} 1, & \text{favored by } u \\ 0, & \text{otherwise} \end{cases} \qquad e^{\gamma} = \text{Odds Ratio bound}$$

- **Implication,** $\delta_i = 0, \delta_{i'} = 1$**:**

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \leq e^{\gamma}$$

- **Example**: $\delta_{\text{neither}} = 0, \delta_{\text{smoking}} = \delta_{\text{drinking}} = 1, \gamma = \ln(3) \implies$ unmeasured confounder can triple the unhealthy habit odds.

The Computation Challenge for Sensitivity Analysis (follow the cue)

s: subject index, $s \in \{1, \ldots, N\}$

$$\mathbb{P}(T \geq t) = \frac{\sum_{\mathbf{z}\in\mathcal{Z}} \mathbb{I}\big(T \geq t\big) \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(z_s = i)\, u_s\right)}{\sum_{b\in\mathcal{Z}} \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(b_s = i)\, u_s\right)}$$

- **Unmeasured confounder vector:** $\mathbf{u} = (u_1, \ldots, u_N) \in [0,1]^N$
- **Optimization problem:** for each γ and δ_i's, compute

$$\max_{\mathbf{u}\in[0,1]^N} \mathbb{P}(T \geq t), \qquad \mathbf{u}^+ = \arg\max_{\mathbf{u}\in[0,1]^N} \mathbb{P}(T \geq t).$$

- **Challenge:**
  - P-value is non-differentiable.
  - $[0,1]^N$ is a large space.

The Computation Challenge for Sensitivity Analysis (follow the cue)

s: subject index, $s \in \{1, \ldots, N\}$

$$\mathbb{P}(T \geq t) = \frac{\sum_{\mathbf{z} \in \mathcal{Z}} \mathbb{I}\big(T \geq t\big) \, \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(z_s = i) \, u_s\right)}{\sum_{b \in \mathcal{Z}} \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(b_s = i) \, u_s\right)}$$

- **Unmeasured confounder vector:** $\mathbf{u} = (u_1, \ldots, u_N) \in [0, 1]^N$
- **Optimization problem:** for each γ and δ_i's, compute

$$\max_{\mathbf{u} \in [0,1]^N} \mathbb{P}(T \geq t), \qquad \mathbf{u}^+ = \arg\max_{\mathbf{u} \in [0,1]^N} \mathbb{P}(T \geq t).$$

- **Challenge:**
  - P-value is non-differentiable.
  - $[0, 1]^N$ is a large space.

The Computation Challenge for Sensitivity Analysis (follow the cue)

s: subject index, $s \in \{1, \dots, N\}$

$$\mathbb{P}(T \geq t) = \frac{\sum_{\mathbf{z} \in \mathcal{Z}} \mathbb{I}\big(T \geq t\big) \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(z_s = i) \, u_s\right)}{\sum_{b \in \mathcal{Z}} \exp\left(\gamma \sum_{s=1}^{N} \sum_{i=1}^{I} \delta_i \, \mathbb{I}(b_s = i) \, u_s\right)}$$

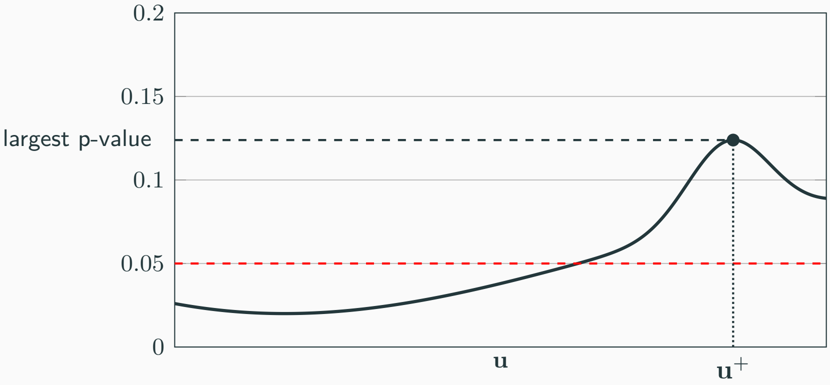
- **Unmeasured confounder vector:** $\mathbf{u} = (u_1, \dots, u_N) \in [0, 1]^N$
- **Optimization problem:** for each γ and δ_i's, compute

$$\max_{\mathbf{u} \in [0,1]^N} \mathbb{P}(T \geq t), \qquad \mathbf{u}^+ = \arg\max_{\mathbf{u} \in [0,1]^N} \mathbb{P}(T \geq t).$$

- **Challenge:**
  - P-value is non-differentiable.
  - $[0, 1]^N$ is a large space.

Our Contribution: Reducing the Search Space for $\mathbf{u}^+$

- $\mathbf{u}^+$ lies in a much smaller space than $[0,1]^N$.
- Candidate structure:
  - Binary outcome $\rightarrow$ 1 candidate.
  - Multi-category outcome $\rightarrow N+1$ candidates.



Lifestyle and New-Onset Impaired Glucose Metabolism

Table 1: Lifestyle Habits and NO-IGM [2]

AIP Patient Status	NO-IGM	No NO-IGM
Neither smoking nor drinking	31	112
Smoking only	54	48
Drinking only	4	4
Both smoking and drinking	35	17

- Unmeasured confounder: poor health awareness
- $(\delta_1, \delta_2, \delta_3, \delta_4) = (0, 1, 1, 1)$
  - Odds ratio $= 1 \Rightarrow p \approx 0$
  - Odds ratio $= 3 \Rightarrow p = 0.002$
  - Odds ratio $= 4.5 \Rightarrow p = 0.10$

[2] Xi, W., et al. (2025), *Smoking, alcohol consumption, and new-onset impaired glucose metabolism in male patients with type 1 autoimmune pancreatitis: a retrospective cohort study*. Therapeutic Advances in Chronic Disease.

Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

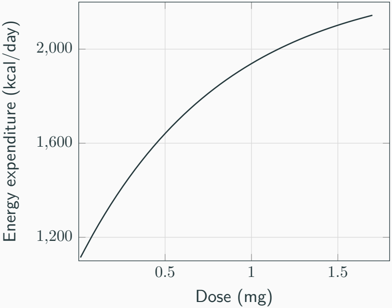
Powerful Test with Continuous Treatment

"Towards Robust Matched Observational Studies with General Treatment Types,"
under review at
Journal of the Royal Statistical Society: Series B (Statistical Methodology).

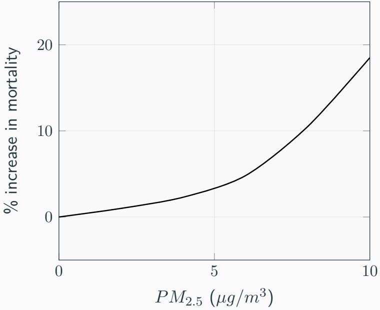
Powerful Test Identification Challenge

- **Feature**: Health studies exhibit different dose-response curves.
- **Challenge**: Comparing test power requires intensive simulations.

Thyroid dose vs. energy expenditure[3]



$PM_{2.5}$ vs. mortality risk[4]



[3] Aronson, J. K. (2007), *Concentration-effect and dose-response relations in clinical pharmacology*.

[4] Shi, L., et al. (2016), *Low-concentration $PM_{2.5}$ and mortality in the Medicare population*.

Contribution: Design Sensitivity for Test Power

Curve	**Test**	**Design sensitivity**	**Power ($\alpha = 0.01$)**	**Power ($\alpha = 0.05$)**
Curve 1	Test 1	2.55	0.70	0.85
	Test 2	2.10	0.64	0.80
Curve 2	Test 1	3.10	0.65	0.75
	Test 2	4.20	0.78	0.90


Introduction
Research Motivation
Story
Categorical Data
Continuous Data
Conclusion

I Would Love to Join You at Genentech!

- A curious, proactive, and detail-oriented person.
- Working on new applications and collaborating with others.
- Devoting my work to improving people's health.

Technical Details

- Categorical Data Visualization–Our Main Object
- Ordinal Test Result
- Binary Outcome Test Result
- Continuous Treatment Sensitivity Analysis
- A Family of Rank-Based Test

Categorical Data Visualization – Our Main Object

Figure 1: A Contingency Table

Outcome (r)

Treatment (Z)

	1	2	$\cdots$	J	
1	N_{11}	N_{12}	$\cdots$	N_{1J}	$N_{1.}$
2	N_{21}	N_{22}	$\cdots$	N_{2J}	$N_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
I	N_{I1}	N_{I2}	$\cdots$	N_{IJ}	$N_{I.}$
	$N_{.1}$	$N_{.2}$	$\cdots$	$N_{.J}$	N

Theorem 2

- A test statistic $T(\mathbf{Z}, \mathbf{r})$ is an ordinal test if

$$T(\mathbf{Z}, \mathbf{r}) = \sum_{i=1}^{I} \sum_{j=1}^{J} w_i v_j N_{ij}, \text{ where } w_1 \leq \cdots \leq w_I \text{ and } v_1 \leq \cdots \leq v_J$$

- Suppose $r_1 \leq r_2 \leq \cdots \leq r_N$. $\mathbf{u}^+$ of an ordinal test T is

$$\mathbf{u}^+ = \operatorname{argmax}_{\mathbf{u} \in \mathcal{U}_{\mathrm{O}}} \alpha(T, \mathbf{r}, \mathbf{u}), \quad \mathcal{U}_{\mathrm{O}} = \left\{ \mathbf{u} \in \{0,1\}^N \mid u_1 \leq \ldots \leq u_N \right\}.$$

- Only needs to search $N + 1$ binary $\mathbf{u}$'s with non-decreasing entries.

Theorem 3

- Suppose the outcome is binary (i.e., $J = 2$). A test statistic $T(\mathbf{Z}, \mathbf{r})$ is a sign-score test if

$$T(\mathbf{Z}, \mathbf{r}) = \sum_{i=1}^{I} w_i N_{i2}, \text{ where } w_1 \leq \cdots \leq w_I$$

- Suppose $r_1 \leq r_2 \leq \cdots \leq r_N$. For a sign-score test,

$$\mathbf{u}^+ = (\underbrace{0, \ldots, 0}_{N_{\cdot 1}}, \underbrace{1, \ldots, 1}_{N_{\cdot 2}}).$$

- The upper bound on the p-value is largest when the outcome $\mathbf{r}$ perfectly predicts $\mathbf{u}^+$ ($N_{\cdot 1}$ is the number of the lower observed outcome subjects.)

Contribution I: Type I Error Control via Sensitivity Analysis

s	:	subject index
Z_s	:	treatment dose
u_s	:	unmeasured confounder, $u_s \in [0,1]$

- **Sensitivity model**

$$\mathbb{P}(Z_s \in dz) = \frac{\int_{z \in dz} \exp(\gamma z u_s)}{\int_{t \in \mathcal{Z}} \exp(\gamma t u_s)}$$

Sensitivity parameter chosen by the analyst: $\gamma > 0$

- **Upper bound on the p-value**

$$\mathbf{u}^+ = \underset{\mathbf{u} \in [0,1]^N}{\operatorname{argmax}} \, \mathbb{P}(T \geq t)$$

$\mathbf{u}^+$ can be uniquely identified for each γ.

A Family of Rank-Based Test Statistics

General test statistic

$$T_\psi(\mathbf{Z}, \mathbf{Y}) = \sum_{i=1}^{I} \psi(r^z_{I,i}, r^y_{I,i})\, \mathbb{I}\{D_i > 0\}, \qquad D_i = (Z_{i1} - Z_{i2})(Y_{i1} - Y_{i2})$$

Ranks of pair differences

$$r^z_{I,i} = I^{-1} \sum_{i'=1}^{I} \mathbb{I}\{|Z_{i1} - Z_{i2}| \geq |Z_{i'1} - Z_{i'2}|\}$$

$$r^y_{I,i} = I^{-1} \sum_{i'=1}^{I} \mathbb{I}\{|Y_{i1} - Y_{i2}| \geq |Y_{i'1} - Y_{i'2}|\}$$

Examples of ψ

- $\psi = 1$: Sign
- $\psi = r^z_{I,i}$: McNemar
- $\psi = r^y_{I,i}$: Wilcoxon
- $\psi = r^z_{I,i} r^y_{I,i}$: Weighted Wilcoxon

Q & A

Common Questions

- Hybrid Control with Sensitivity Analysis
- Why Do Multiple Sensitivity Analyses Exist?
- Can We Not Just Guess?
- Why $u_s \in [0, 1]$
- Examples of Rank-Based Tests for Continuous Data
- Dose-Response Curve & Powerful Tests

Hybrid Control with Sensitivity Analysis

Idea

- Combine randomized controls with observational controls.
- Randomization protects study arms from hidden bias.
- Observational controls may be affected by an unmeasured confounder.

ID	Arm / Source	u
1	Study Treatment	0
2	Study Treatment	0
3	Study Control	0
4	Study Control	0
5	Observational Control	$u_5 \in [0,1]$
6	Observational Control	$u_6 \in [0,1]$

$$\max_{u_i \in [0,1]} p\text{-value}(u)$$

Sensitivity analysis

$$u_i = \begin{cases} 0 & \text{if subject from randomized study} \\ \in [0,1] & \text{if subject from observational control} \end{cases}$$

Why Do Multiple Sensitivity Analyses Exist?

Quantity of Interest	Example Papers
Regression coefficient	Cinelli & Hazlett (2020)[1]
ATT / ATE	Huang & Pimentel (2025)[2]
P-value in hypothesis testing	Rosenbaum & Krieger (1990)[3]

[1] Cinelli, C. & Hazlett, C. (2020). *Making Sense of Sensitivity: Extending Omitted Variable Bias*.

[2] Huang, M. & Pimentel, S. D. (2025). *Variance-based sensitivity analysis for weighting estimators results in more informative bounds*.

[3] Rosenbaum, P. R. & Krieger, A. M. (1990). *Sensitivity of Two-Sample Permutation Inferences in Observational Studies*.

However, I think my formulation is the most helpful one if the target is p-value/hypothesis testing

Can we not just guess?

- We can use some other variables in a causal graph
- Negative control + sensitivity Analysis
- Negative control outcome: an outcome the intervention should not affect, if it is affected, there is a bias

Why $u_s \in [0, 1]$

s	: subject index, $s \in \{1, \ldots, N\}$
i	: treatment index, $i \in \{1, \ldots, I\}$
Z_s	: treatment received by subject s
u_s	: unmeasured confounder, $u_s \in [0, 1]$

- u_s is better thought as a "missing propensity score", not a variable.
- We set it up to conveniently quantify the treatment odds bound.
- **Sensitivity model**

$$\mathbb{P}(Z_s = i) = \frac{\exp(\gamma\, \delta_i\, u_s)}{\sum_{k=1}^{I} \exp(\gamma\, \delta_k\, u_s)}$$

- **Implication,** $\delta_i = 0, \delta_{i'} = 1$**:**

$$\frac{\mathbb{P}(Z_s = i')/\mathbb{P}(Z_s = i)}{\mathbb{P}(Z_{s'} = i')/\mathbb{P}(Z_{s'} = i)} = \exp\{\gamma(u_s - u_{s'})\} \leq e^{\gamma}$$

- The upper bound happens when $u_s = 1$, $u_{s'} = 0$:

$$\exp(\gamma(u_s - u_{s'})) = \exp(\gamma(1 - 0)) = e^{\gamma}$$

Examples of Rank-Based Tests for Continuous Data

Pair	Z_{i1}	Z_{i2}	$Z_{i1} - Z_{i2}$	Y_{i1}	Y_{i2}	$Y_{i1} - Y_{i2}$	$D_i = (Z_{i1} - Z_{i2})(Y_{i1} - Y_{i2})$
1	8	5	3	12	9	3	$+$
2	4	7	-3	10	8	2	$-$
3	9	6	3	11	13	-2	$-$

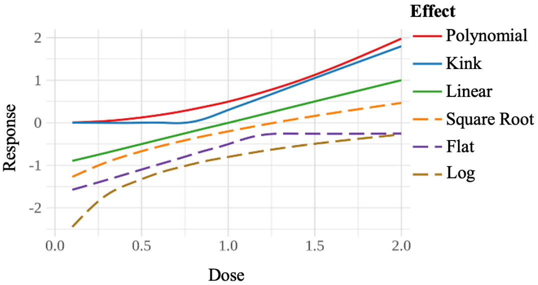
Wilcoxon signed-rank test

$$T_{\text{Wilcoxon}} = \sum_{i=1}^{I} \text{rank}(|Y_{i1} - Y_{i2}|) \cdot \text{sign}(D_i)$$

Dose–response Wilcoxon-type test

$$T = \sum_{i=1}^{I} \text{rank}(|Z_{i1} - Z_{i2}| \times |Y_{i1} - Y_{i2}|) \cdot \text{sign}(D_i))$$

Dose-Response Curve & Powerful Tests



Wilcox	D-Wilcox	U	D-U
2.32	2.44	3.32	4.16
2.26	2.43	3.17	4.28
2.68	2.68	4.16	4.43
2.45	2.38	3.67	3.50
2.60	2.45	4.02	3.37
2.83	2.63	4.65	3.84

Figure 2: Design Sensitivity & Dose Response Curves