# Apportioning Causal Responsibility of Two Risk Factors for an Adverse Outcome via Counterfactual Attribution

**Shanshan Luo**

Beijing Technology and Business University

Joint work with Yafang Deng, Qingyuan Zhao, and Zhi Geng

狗熊会 Online Seminar

北京工商大學
BEIJING TECHNOLOGY AND BUSINESS UNIVERSITY

arXiv:2606.18459

# Collaborators

## Yafang Deng

Second-year Master Student

School of Mathematics and Statistics
Beijing Technology and Business University

## Qingyuan Zhao

Professor

Statistical Laboratory
University of Cambridge

## Zhi Geng

Professor

School of Mathematical Sciences
Peking University
Beijing Technology and Business University

# Outline

PART I OF IV

Part I begins: motivation, existing approaches, and the attribution gap.

1 **Motivation and problem**

Causes of effects, why separate but-for scores do not form an allocation.

2 **Counterfactual framework**

Potential outcomes, latent causal types, and average causal responsibility.

3 **Identification results**

Active-only and balanced rules, point identification, and sharp bounds.

4 **Numerical illustrations**

Two smoking–asbestos scenarios, rule comparisons, and final takeaways.

# Causal Attribution Focuses on Counterfactual Questions

## Level 3: Counterfactuals

*Today's focus*

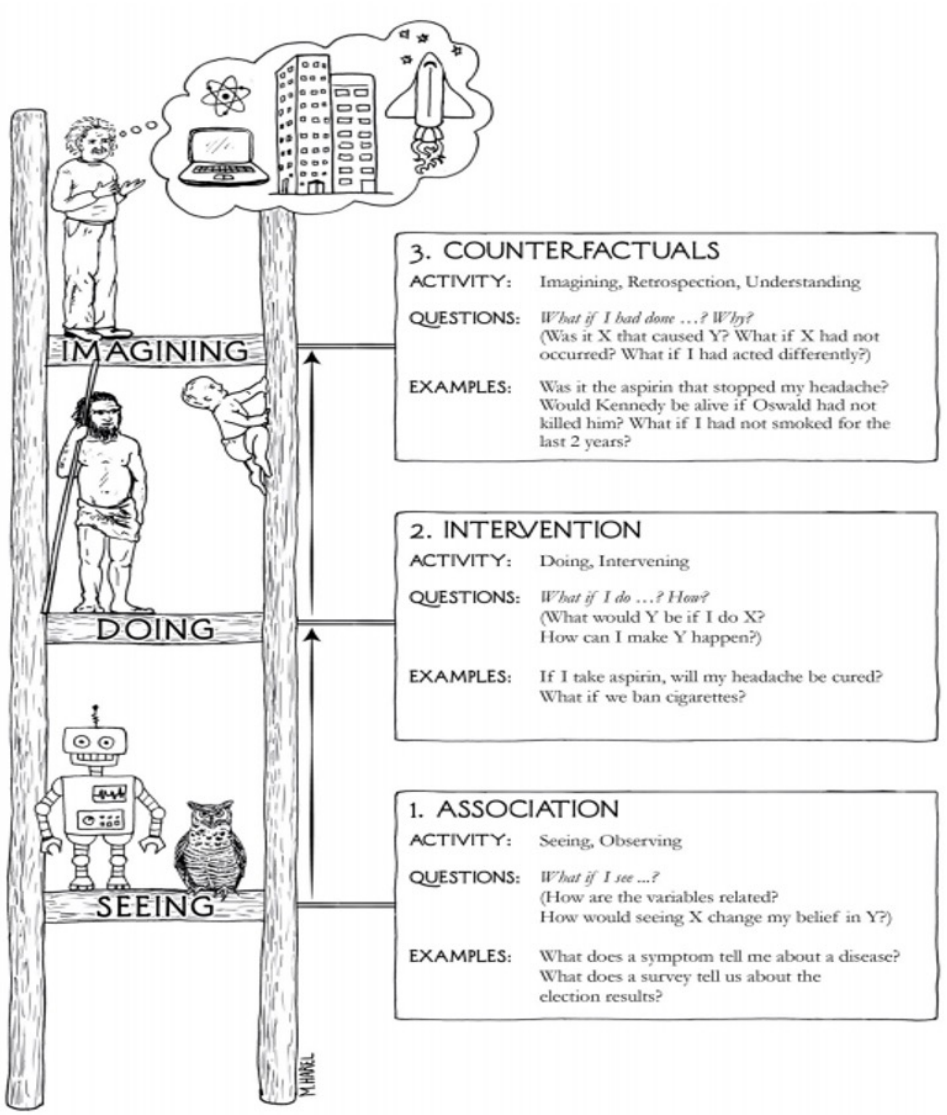
- Legal liability allocation
- Medical compensation

## Level 2: Intervention

- Treatment effect evaluation
- Public health policy design

## Level 1: Association

- Risk prediction
- Correlation




We move from predicting effects to explaining an outcome that already occurred.

# Effects of Causes vs. Causes of Effects

## Level 2: Intervention, Effects of Causes

- Prospective and forward-looking: predict what would happen under an intervention.
- Question: If we ban smoking, how much will lung cancer rates decrease?
- Mathematics: average treatment effect, conditional average treatment effect.
- Methods: propensity scores, doubly robust estimation, instrumental variables.

## Level 3: Counterfactuals, Causes of Effects

- Retrospective and backward-looking: infer why an observed outcome occurred.
- Question: Given Bob's smoking, would he still get lung cancer if he didn't smoke?
- Mathematics: condition on the observed outcome and counterfactual mechanisms.
- Application: legal liability, insurance and medical compensation.

# The attribution problem: why did Bob get lung cancer?

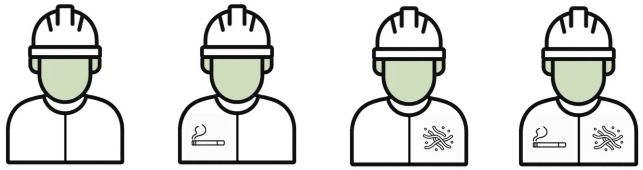
## Bob's claim

observed case:
smoking + asbestos
+ lung cancer

**Total loss: £10,000**

Bob seeks compensation.
The court asks how to split
£10,000.

## Lung cancer rates



$\xi_{00}$ = 0.12%  $\xi_{10}$ = 0.95%  $\xi_{01}$ = 0.67%  $\xi_{11}$ = 4.51%

👉 **Smoking's risk difference:** $\xi_{10} - \xi_{00} = 0.83\%$

👉 **Asbestos's risk difference :** $\xi_{01} - \xi_{00} = 0.55\%$

**The court asks: how should responsibility be apportioned between smoking and asbestos? Who should pay what share?**

**Smoking company**

Did smoking contribute to this cancer?

**Asbestos / employer**

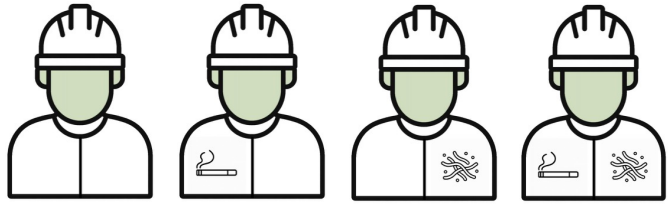
Did workplace asbestos contribute?

**Background risk**

Would it have happened anyway?

# Naïve method

**Lung cancer rates**



$\xi_{00}$ = 0.12% $\xi_{10}$ = 0.95% $\xi_{01}$ = 0.67% $\xi_{11}$ = 4.51%

👉 **Smoking's risk difference:** $\xi_{10}-\xi_{00} = 0.83\%$

👉 **Asbestos's risk difference :** $\xi_{01}-\xi_{00} = 0.55\%$

$$\text{Smoking's responsibility} = \frac{\xi_{10}-\xi_{00}}{\xi_{10}-\xi_{00}+\xi_{01}-\xi_{00}} = 60.14\%$$

$$\text{Asbestos's responsibility} = \frac{\xi_{01}-\xi_{00}}{\xi_{10}-\xi_{00}+\xi_{01}-\xi_{00}} = 39.86\%$$

**It converts separate risk differences into shares by construction. The resulting 100% allocation has no explicit counterfactual justification.**

# Single binary cause – probability of necessity

$$\mathrm{PN}_Z = \mathrm{pr}(Y_{Z=0} = 0 \mid Z = 1, Y = 1)$$
$$\mathrm{PN}_M = \mathrm{pr}(Y_{M=0} = 0 \mid M = 1, Y = 1)$$

## Interpretation

$\mathrm{PN}_Z$ measures the probability that the observed exposure was necessary for the occurred outcome.

$\mathrm{PN}_Z = 1$ means the outcome would not have occurred without the exposure.

## Key point

Randomization identifies interventional quantities such as $E(Y_{Z=1} - Y_{Z=0})$, but it does not by itself identify this retrospective probability of necessity.

**It answers a separate but-for question for each exposure. The two probabilities of necessity do not necessarily sum to one.**


Pearl, J. (2000). Causality: Models, Reasoning, and Inference. Cambridge University Press.
Tian, J., & Pearl, J. (2000). Probabilities of causation: Bounds and identification. Annals of Mathematics and Artificial Intelligence, 28(1–4), 287–313.
Dawid, A. P. & Musio, M. (2022). Effects of causes and causes of effects. Annual Review of Statistics and Its Application, 9:261–287.

# Multiple binary causes: posterior causal effects

## Smoking contribution

$$\text{PostTCE}_Z = E(Y_{Z=1} - Y_{Z=0} \mid Z = 1, M = 1, Y = 1)$$

This removes smoking while keeping asbestos fixed.

## Asbestos contribution

$$\text{PostTCE}_M = E(Y_{M=1} - Y_{M=0} \mid Z = 1, M = 1, Y = 1)$$

This removes asbestos while keeping smoking fixed.

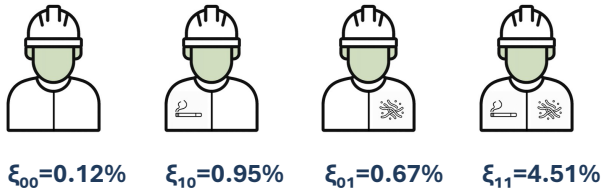
$$\text{PostTCE}_Z = E(Y_{Z=1} - Y_{Z=0} \mid Z = 1, M = 1, Y = 1) = \text{pr}(Y_{Z=0} = 0 \mid Z = 1, M = 1, Y = 1) = \frac{\xi_{11} - \xi_{01}}{\xi_{11}}$$

$$\text{PostTCE}_M = E(Y_{M=1} - Y_{M=0} \mid Z = 1, M = 1, Y = 1) = \text{pr}(Y_{M=0} = 0 \mid Z = 1, M = 1, Y = 1) = \frac{\xi_{11} - \xi_{10}}{\xi_{11}}$$

These quantities are posterior total causal effects. For a single binary cause, under randomization and monotonicity, PostTCE reduces to PN.

Lu, Z., Geng, Z., Li, W., Zhu, S., and Jia, J. (2023). Evaluating causes of effects by posterior effects of causes. Biometrika, 110(2), 449–465.
Li, W., Lu, Z., Jia, J., Xie, M., and Geng, Z. (2024). Retrospective causal inference with multiple effect variables. Biometrika, 111(2), 573–589.

# Separate posterior effects still do not allocate responsibility




Smoking's risk difference: $\xi_{10}-\xi_{00} = 0.83\%$

Asbestos's risk difference : $\xi_{01}-\xi_{00} = 0.55\%$

## Smoking / tobacco company

$$\text{PostTCE}_Z = \frac{\xi_{11} - \xi_{01}}{\xi_{11}} = 85.14\%$$

No smoking, asbestos still present: would cancer still have occurred?

## Asbestos / employer

$$\text{PostTCE}_M = \frac{\xi_{11} - \xi_{10}}{\xi_{11}} = 78.94\%$$

No asbestos, smoking still present: would cancer still have occurred?

**They measure the expected outcome change from removing each cause while holding the other cause fixed. They quantify the two causes separately.**

# Outline

PART II OF IV

Part II begins: potential outcomes, latent causal types, and average causal responsibility.

1 **Motivation and problem**

Causes of effects, why separate but-for scores do not form an allocation.

2 **Counterfactual framework**

Potential outcomes, latent causal types, and average causal responsibility.

3 **Identification results**

Active-only and balanced rules, point identification, and sharp bounds.

4 **Numerical illustrations**

Two smoking–asbestos scenarios, rule comparisons, and final takeaways.

# Formal Notation: Variables and Potential Outcomes

**Risk factors**

$$Z, M \in \{0,1\}$$

$Z$ denotes smoking status; M denotes workplace asbestos exposure.

**Outcome**

$$Y \in \{0,1\}$$

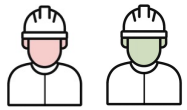
$Y$ denotes the lung cancer outcome. $Y$=1 means cancer occurred; $Y$=0 means it did not occur.

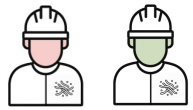
**Potential outcome**

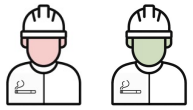
$$Y_{zm} \in \{0,1\}$$

$Y_{zm}$ is the outcome that would occur under $Z = z$ and $M = m$ for the same worker.

**Four counterfactual worlds**

$Y_{00}$ 

$Y_{01}$ 

$Y_{10}$ 

$Y_{11}$ 

# Observed risks under four exposure regimes

**No confounding:** $(Z, M) \perp (Y_{00}, Y_{01}, Y_{10}, Y_{11}) \Rightarrow \quad \xi_{zm} = \mathrm{pr}(Y = 1 \mid Z = z, M = m) = \mathrm{pr}(Y_{zm} = 1)$

**$Y_{00}$**

**Neither**

$\xi_{00} = \mathrm{pr}(Y_{00} = 1)$

**$Y_{01}$**

**Asbestos only**

$\xi_{01} = \mathrm{pr}(Y_{01} = 1)$

**$Y_{10}$**

**Smoking only**

$\xi_{01} = \mathrm{pr}(Y_{01} = 1)$

**$Y_{11}$**

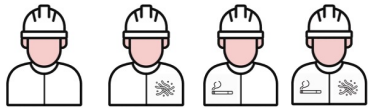
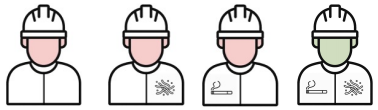
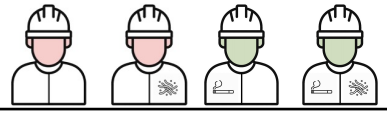
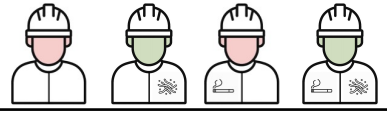
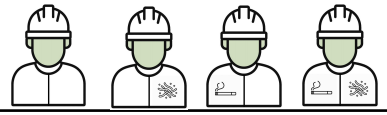
**Both exposures**

$\xi_{11} = \mathrm{pr}(Y_{11} = 1)$

# Formal Notation: Latent Causal Type

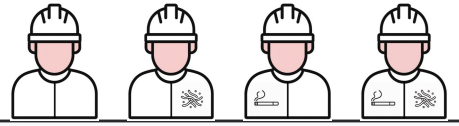
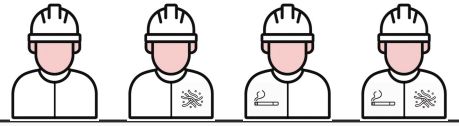
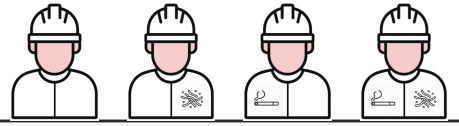
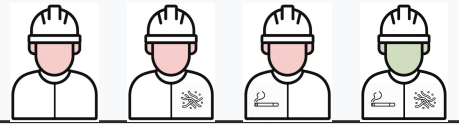
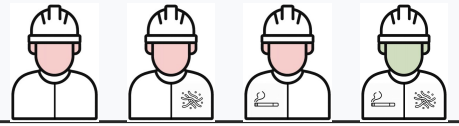
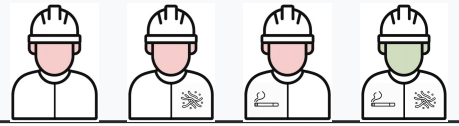
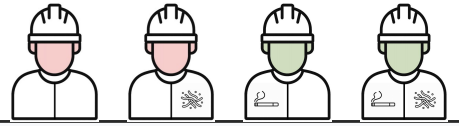
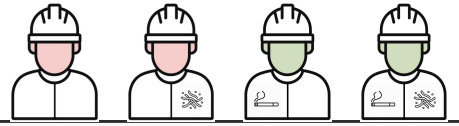
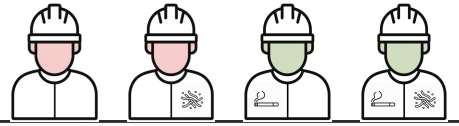
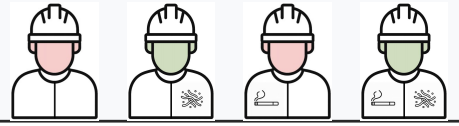
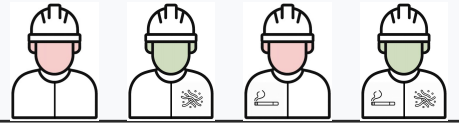
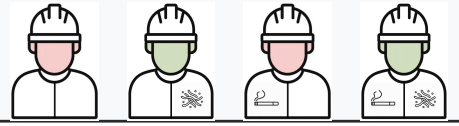
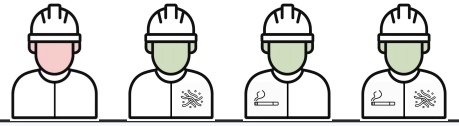
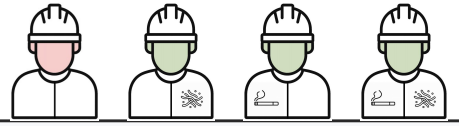
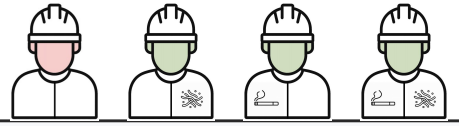
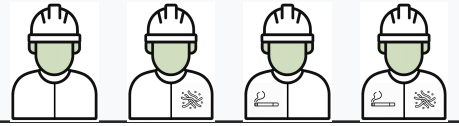
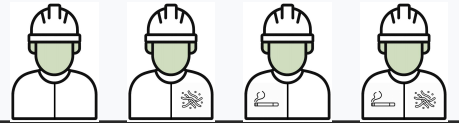
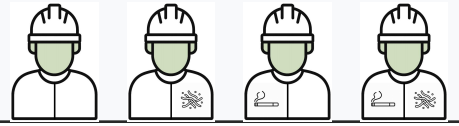
**Monotonicity** $Y_{00} \leq Y_{01} \leq Y_{11}, \quad Y_{00} \leq Y_{10} \leq Y_{11}$

**Neither smoking nor asbestos can be protective for any individual.**

| Story | Counterfactual pattern | Potential Outcomes $(Y_{00}, Y_{01}, Y_{10}, Y_{11})$ |
|---|---|---|
| Immune |  | $(Y_{00}= 0, Y_{01} = 0, Y_{10} = 0, Y_{11} = 0)$ |
| Synergy |  | $(Y_{00} = 0, Y_{01} = 0, Y_{10} = 0, Y_{11} = 1)$ |
| Smoking only |  | $(Y_{00} = 0, Y_{01} = 0, Y_{10} = 1, Y_{11} = 1)$ |
| Asbestos only |  | $(Y_{00} = 0, Y_{01} = 1, Y_{10} = 0, Y_{11} = 1)$ |
| Parallel |  | $(Y_{00} = 0, Y_{01} = 1, Y_{10} = 1, Y_{11} = 1)$ |
| Doomed |  | $(Y_{00} = 1, Y_{01} = 1, Y_{10} = 1, Y_{11} = 1)$ |

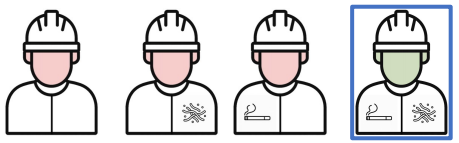
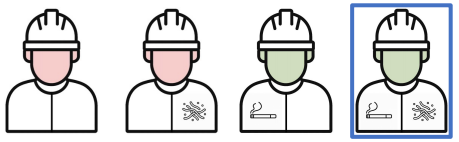
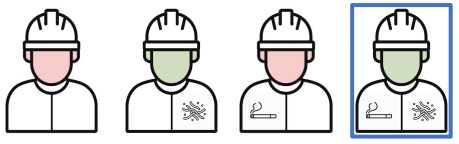
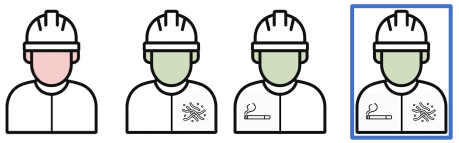
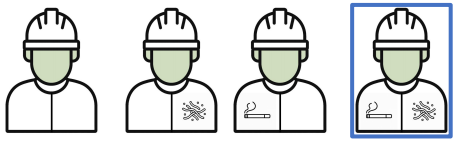
# Evidence and latent causal type

Evidence eliminates impossible types, but it does not reveal the unique mechanism.

| Story | $\mathcal{E}_{11}$: Z=1, M=1, Y=1 | $\mathcal{E}_{10}$: Z=1, M=0, Y=1 | $\mathcal{E}_{01}$: Z=0, M=1, Y=1 | Type |
|---|---|---|---|---|
| **Immune** |  × |  × |  × | 0000 |
| **Synergy** |  ✓ |  × |  × | 0001 |
| **Smoking only** |  ✓ |  ✓ |  × | 0011 |
| **Asbestos only** |  ✓ |  × |  ✓ | 0101 |
| **Parallel** |  ✓ |  ✓ |  ✓ | 0111 |
| **Doomed** |  ✓ |  ✓ |  ✓ | 1111 |

**Key message: several causal stories may be compatible with the same observed cancer case.**

# Normative input: type-specific responsibilities

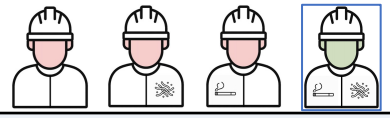
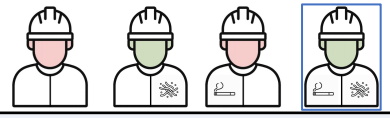
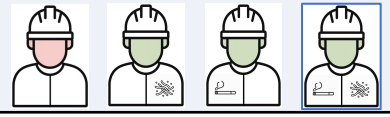
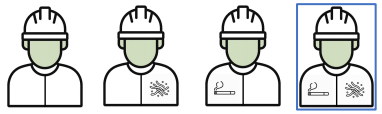
| Story | Counterfactual pattern (worlds with cancer) | Smoking's Responsibility | Asbestos's Responsibility | Background's Responsibility |
|---|---|---|---|---|
| **Synergy** |  | $R_Z(0001 \mid \mathcal{E}_{11})$ | $R_M(0001 \mid \mathcal{E}_{11})$ | $R_0(0001 \mid \mathcal{E}_{11})$ |
| **Smoking only** |  | $R_Z(0011 \mid \mathcal{E}_{11})$ | $R_M(0011 \mid \mathcal{E}_{11})$ | $R_0(0011 \mid \mathcal{E}_{11})$ |
| **Asbestos only** |  | $R_Z(0101 \mid \mathcal{E}_{11})$ | $R_M(0101 \mid \mathcal{E}_{11})$ | $R_0(0101 \mid \mathcal{E}_{11})$ |
| **Parallel** |  | $R_Z(0111 \mid \mathcal{E}_{11})$ | $R_M(0111 \mid \mathcal{E}_{11})$ | $R_0(01111 \mid \mathcal{E}_{11})$ |
| **Background** |  | $R_Z(1111 \mid \mathcal{E}_{11})$ | $R_M(1111 \mid \mathcal{E}_{11})$ | $R_0(1111 \mid \mathcal{E}_{11})$ |

**For each latent type, the rule specifies how responsibility is divided among smoking, asbestos, and background.**

# Partially identified posterior type weights $\text{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm})$

$$\nu = \text{pr}(Y_{01} = 1, Y_{10} = 1)$$
$$\nu^{\ell} = \max\{\xi_{01} + \xi_{10} - \xi_{11}, \xi_{00}\},\ \nu^{u} = \min\{\xi_{01}, \xi_{10}\}$$

**The same $\nu$ controls all rows, so the endpoints cannot be chosen independently.**

| Story | Counterfactual pattern (worlds with cancer) | Statistics Weight (randomization & monotonicity | Sharp Bounds (randomization & monotonicity) |
|---|---|---|---|
| **Synergy** $\text{pr}(0001 \mid \mathcal{E}_{11})$ |  | $\frac{\xi_{11} - \xi_{10} - \xi_{01} + \nu}{\xi_{11}}$ | $\left[\frac{\xi_{11} - \xi_{10} - \xi_{01} + \nu^{l}}{\xi_{11}}, \frac{\xi_{11} - \xi_{10} - \xi_{01} + \nu^{u}}{\xi_{11}}\right]$ |
| **Smoking only** $\text{pr}(0011 \mid \mathcal{E}_{11})$ |  | $\frac{\xi_{01} - \nu}{\xi_{11}}$ | $\left[\frac{\xi_{01} - \nu^{u}}{\xi_{11}}, \frac{\xi_{01} - \nu^{l}}{\xi_{11}}\right]$ |
| **Asbestos only** $\text{pr}(0101 \mid \mathcal{E}_{11})$ |  | $\frac{\xi_{10} - \nu}{\xi_{11}}$ | $\left[\frac{\xi_{10} - \nu^{u}}{\xi_{11}}, \frac{\xi_{10} - \nu^{l}}{\xi_{11}}\right]$ |
| **Parallel** $\text{pr}(0111 \mid \mathcal{E}_{11})$ |  | $\frac{\nu - \xi_{00}}{\xi_{11}}$ | $\left[\frac{\nu^{l} - \xi_{00}}{\xi_{11}}, \frac{\nu^{u} - \xi_{00}}{\xi_{11}}\right]$ |
| **Background** $\text{pr}(1111 \mid \mathcal{E}_{11})$ |  | $\frac{\xi_{00}}{\xi_{11}}$ | $\frac{\xi_{00}}{\xi_{11}}$ |

# Average Causal Responsibility for Evidence $\mathcal{E}_{11}$

**Observed evidence $\mathcal{E}_{11}$**

Bob smoked,
was exposed to asbestos,
and developed lung cancer.

$$R_j(\mathcal{E}_{11}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_j(\text{possible story} \mid \mathcal{E}_{11})\text{pr}(\text{possible story} \mid \mathcal{E}_{11})$$

**Legal input**

$R_j(\text{possible story} \mid \mathcal{E}_{11})$ is the responsibility assigned to factor $j$ inside one counterfactual story.

**Statistical input**

$\text{pr}(\text{possible story} \mid \mathcal{E}_{11})$ is the posterior probability that Bob belongs to that story.

**In words: responsibility = legal share × statistical weight, summed over possible stories.**

# Formal Definition: Average Causal Responsibility

A responsibility rule is averaged over the posterior distribution of latent types.

For $j \in \{\text{smoking, asbestos, background}\}$,

$$R_j(\mathcal{E}_{zm}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_j(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm}) \text{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm})$$

**Legal / normative part**

$$R_j(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm})$$

Within each latent causal type t, the rule says how responsibility should be divided among smoking, asbestos, and background.

**Statistical part**

$$\text{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm})$$

The posterior probability that Bob belongs to latent type t, conditional on the observed cancer case.

**Identification question: can this posterior average be written using only the observed risks $\xi_{00}$, $\xi_{10}$, $\xi_{01}$, $\xi_{11}$?**

# Outline

PART III OF IV

Part III begins: responsibility rules, point identification, and sharp bounds.

1 **Motivation and problem**

Causes of effects, why separate but-for scores do not form an allocation.

2 **Counterfactual framework**

Potential outcomes, latent causal types, and average causal responsibility.

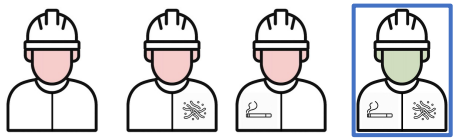
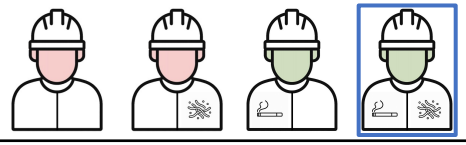
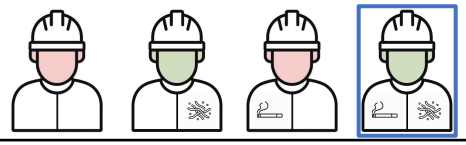
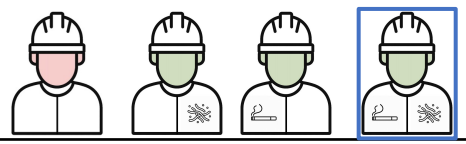
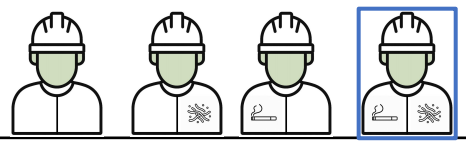
3 **Identification results**

Active-only and balanced rules, point identification, and sharp bounds.

4 **Numerical illustrations**

Two smoking–asbestos scenarios, rule comparisons, and final takeaways.

# Assumption: active-only rule for $\mathcal{E}_{11}$

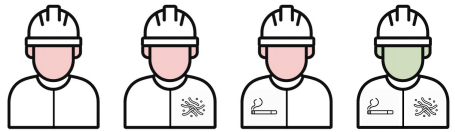
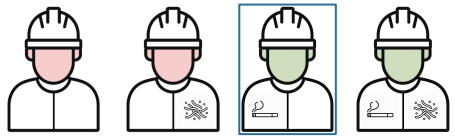
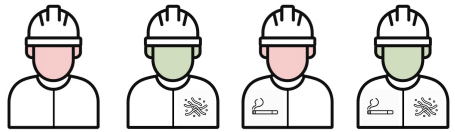
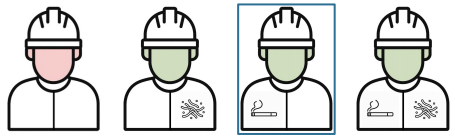
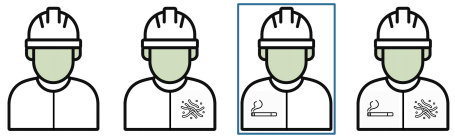
| Story | Counterfactual pattern (worlds with cancer) | Smoking's Responsibility | Asbestos's Responsibility | Background's Responsibility |
|---|---|---|---|---|
| **Synergy** |  | $R_Z(0001 \mid \mathcal{E}_{11})$=50% | $R_M(0001 \mid \mathcal{E}_{11})$=50% | $R_0(0001 \mid \mathcal{E}_{11})$=0% |
| **Smoking only** |  | $R_Z(0011 \mid \mathcal{E}_{11})$=100% | $R_M(0011 \mid \mathcal{E}_{11})$=0% | $R_0(0011 \mid \mathcal{E}_{11})$=0% |
| **Asbestos only** |  | $R_Z(0101 \mid \mathcal{E}_{11})$=0% | $R_M(0101 \mid \mathcal{E}_{11})$=100% | $R_0(0101 \mid \mathcal{E}_{11})$=0% |
| **Parallel** |  | $R_Z(0111 \mid \mathcal{E}_{11})$=50% | $R_M(0111 \mid \mathcal{E}_{11})$=50% | $R_0(0111 \mid \mathcal{E}_{11})$=0% |
| **Background** |  | $R_Z(1111 \mid \mathcal{E}_{11})$=0% | $R_M(1111 \mid \mathcal{E}_{11})$=0% | $R_0(1111 \mid \mathcal{E}_{11})$=100% |

**A sole active factor receives 100%; two active factors split 50–50%; causally inactive factors receive 0%.**

# Assumption: active-only rule for $\varepsilon_{10}$

?

Only one mechanism is true for Bob, but we do not observe it.

| Story | Counterfactual pattern $\varepsilon_{10}$ (worlds with cancer) | Responsibility |
|---|---|---|
| Synergy |  | NA |
| Smoking only |  | Smoking: 100%<br>Asbestos: 0%<br>Background: 0% |
| Asbestos only |  | NA |
| Parallel |  | Smoking: 100%<br>Asbestos: 0%<br>Background: 0% |
| Background |  | Smoking: 0%<br>Asbestos: 0%<br>Background: 100% |

Only one mechanism is true for the worker, but it is not observed.
We therefore average the rule over the compatible posterior type probabilities.

# Theorem 1: point identification under the active-only rule

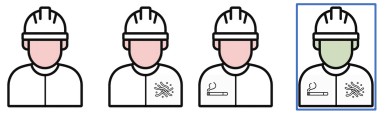
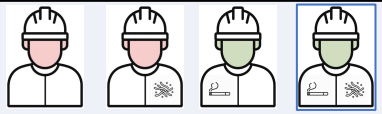
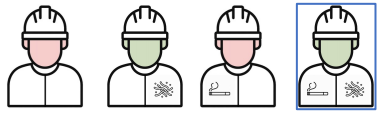
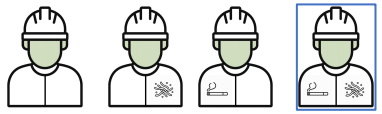
$$R_j(\mathcal{E}_{zm}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_j(y_{00}y_{01}y_{10}y_{11}\,|\mathcal{E}_{zm})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{zm})$$

**Assumptions: no confounding and monotonicity.**
**Rule: one active factor receives 100%; two active factors split 50–50; inactive factors receive 0%.**

| | | | |
|---|---|---|---|
| **Dual exposure $\mathcal{E}_{11}$** | $R_0(\mathcal{E}_{11}) = \frac{\xi_{00}}{\xi_{11}}$ | $R_Z(\mathcal{E}_{11}) = \frac{\xi_{11} + \xi_{10} - \xi_{01} - \xi_{00}}{2\xi_{11}}$ | $R_M(\mathcal{E}_{11}) = \frac{\xi_{11} - \xi_{10} + \xi_{01} - \xi_{00}}{2\xi_{11}}$ |
| **Smoking-only $\mathcal{E}_{10}$** | $R_0(\mathcal{E}_{10}) = \frac{\xi_{00}}{\xi_{10}}$ | $R_Z(\mathcal{E}_{10}) = \frac{\xi_{10} - \xi_{00}}{\xi_{10}}$ | $R_M(\mathcal{E}_{10}) = 0$ |
| **Asbestos-only $\mathcal{E}_{01}$** | $R_0(\mathcal{E}_{01}) = \frac{\xi_{00}}{\xi_{01}}$ | $R_Z(\mathcal{E}_{01}) = 0$ | $R_M(\mathcal{E}_{01}) = \frac{\xi_{01} - \xi_{00}}{\xi_{01}}$ |

**The latent type distribution is only partially identified, yet average responsibility is point-identified because v cancels.**

# Why the unidentified parameter cancels

| Story | Counterfactual pattern (worlds with cancer) | Fair Responsibility Rule | Statistics Weight (randomization & monotonicity) |
|---|---|---|---|
| **Synergy** |  | **Smoking: 50%  Asbestos: 50%** | $\frac{\xi_{11} - \xi_{10} - \xi_{01} + \nu}{\xi_{11}}$ |
| **Smoking only** |  | **Smoking: 100%  Asbestos: 0%** | $\frac{\xi_{01} - \nu}{\xi_{11}}$ |
| **Asbestos only** |  | **Smoking: 0%  Asbestos: 100%** | $\frac{\xi_{10} - \nu}{\xi_{11}}$ |
| **Parallel** |  | **Smoking: 50%  Asbestos: 50%** | $\frac{\nu - \xi_{00}}{\xi_{11}}$ |
| **Background** |  | **Smoking: 0%  Asbestos: 0%** | $\frac{\xi_{00}}{\xi_{11}}$ |

Smoking's responsibility: $\frac{\xi_{11}+\xi_{10}-\xi_{01}-\xi_{00}}{2\xi_{11}}$  Asbestos's responsibility: $\frac{\xi_{11}+\xi_{01}-\xi_{10}-\xi_{00}}{2\xi_{11}}$

# Single-exposure responsibility reduces to PN

### Smoking-only evidence $\mathcal{E}_{10}$

Observed evidence: $\mathcal{E}_{10} = (Z{=}1,\ M{=}0,\ Y{=}1)$

$$R_0(\mathcal{E}_{10}) = \frac{\xi_{00}}{\xi_{10}}$$

$$R_Z(\mathcal{E}_{10}) = \frac{\xi_{10} - \xi_{00}}{\xi_{10}}$$

$$R_M(\mathcal{E}_{10}) = 0$$

### Asbestos-only evidence $\mathcal{E}_{01}$

Observed evidence $\mathcal{E}_{01}$ : $(Z{=}0,\ M{=}1,\ Y{=}1)$

$$R_0(\mathcal{E}_{01}) = \frac{\xi_{00}}{\xi_{10}}$$

$$R_Z(\mathcal{E}_{01}) = 0$$

$$R_M(\mathcal{E}_{01}) = \frac{\xi_{10} - \xi_{00}}{\xi_{10}}$$

**When only one exposure is present, the active-only rule coincides with the usual probability of necessity for that exposure.**

# Numerical result: active-only rule

A £10,000 loss is translated into responsibility shares for the occurred cancer.

**Observed risks:** $\xi_{00}$=0.12% $\xi_{10}$=0.95% $\xi_{01}$=0.67% $\xi_{11}$=4.51%

Smoking's responsibility: $\frac{\xi_{11}+\xi_{10}-\xi_{01}-\xi_{00}}{2\xi_{11}} = 51.77\%,$

Asbestos's responsibility: $\frac{\xi_{11}+\xi_{01}-\xi_{10}-\xi_{00}}{2\xi_{11}} = 45.57\%.$

## Bob's claim

Observed case:
smoking + asbestos
+ lung cancer

**Total loss: £10,000**

Bob seeks compensation.
The court asks how to split £10,000.

**Who should pay what share?**

**Smoking / tobacco company** **51.77%** **£5,177**

**Workplace asbestos / employer** **45.57%** **£4,557**

**Background vulnerability** **2.66%** **£266**

**Interpretation: smoking bears slightly more than half; asbestos bears almost the other half; background risk is small but not zero.**

# Why consider alternative responsibility rules?

### 1. Baseline vulnerability

Age, genetics, chronic illness, and lifestyle may deserve a nonzero share even when exposures are causally active.

### 2. Asymmetric culpability

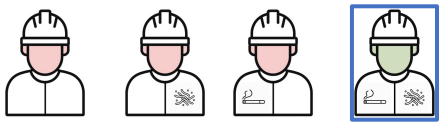
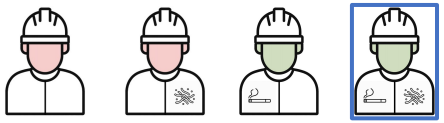
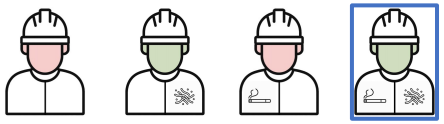
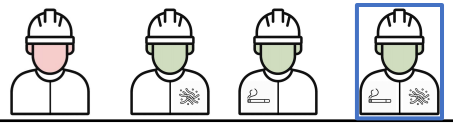
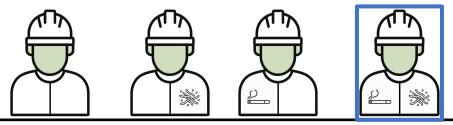
Negligent workplace exposure may be judged more blameworthy than voluntary smoking.

### 3. Domain-specific credit

Advertising and policy settings may value independent sufficiency differently from synergy.

**Separation principle: the framework separates rule choice from statistical identification. Any pre-specified type-specific allocation can be analyzed.**

# Structural balance: the exact condition for cancellation

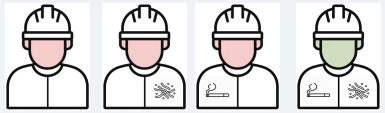
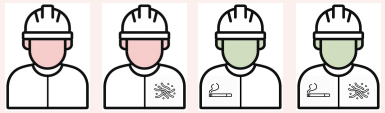
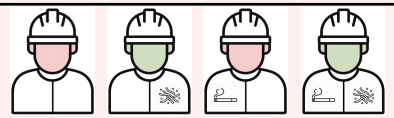
| Story | Counterfactual pattern (worlds with cancer) | Smoking's Responsibility | Asbestos's Responsibility | Background's Responsibility |
|---|---|---|---|---|
| Synergy |  | $R_Z(0001 \mid \mathcal{E}_{11})=30\%$ | $R_M(0001 \mid \mathcal{E}_{11})=30\%$ | $R_0(0001 \mid \mathcal{E}_{11})=40\%$ |
| Smoking only |  | $R_Z(0011 \mid \mathcal{E}_{11})=60\%$ | $R_M(0011 \mid \mathcal{E}_{11})=20\%$ | $R_0(0011 \mid \mathcal{E}_{11})=20\%$ |
| Asbestos only |  | $R_Z(0101 \mid \mathcal{E}_{11})=10\%$ | $R_M(0101 \mid \mathcal{E}_{11})=50\%$ | $R_0(0101 \mid \mathcal{E}_{11})=40\%$ |
| Parallel |  | $R_Z(0111 \mid \mathcal{E}_{11})=40\%$ | $R_M(0111 \mid \mathcal{E}_{11})=40\%$ | $R_0(0111 \mid \mathcal{E}_{11})=20\%$ |
| Background |  | $R_Z(1111 \mid \mathcal{E}_{11})=10\%$ | $R_M(1111 \mid \mathcal{E}_{11})=10\%$ | $R_0(1111 \mid \mathcal{E}_{11})=80\%$ |

**Smoking balance:**
30% + 40 % = 60 % + 10 % = 70 %

**Asbestos balance:**
30 % + 40 % = 20 % + 50 % = 70 %

**Background may be positive and allocations may be asymmetric across types, yet average responsibility remains point-identified.**

# Structural Balance: the exact condition for cancellation

| Story | Counterfactual pattern (worlds with cancer) | Type-specific responsibility $R_{\mathbf{z}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$ | Type-specific responsibility $R_{\mathrm{M}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$ |
|---|---|---|---|
| **Synergy** $(Y_{00}=0, Y_{01}=0, Y_{10}=0, Y_{11}=1)$ |  | $R_Z(0001 \mid \mathcal{E}_{11})$=30% | $R_M(0001 \mid \mathcal{E}_{11})$=30% |
| **Smoking only** $(Y_{00}=0, Y_{01}=0, Y_{10}=1, Y_{11}=1)$ |  | $R_Z(0011 \mid \mathcal{E}_{11})$=60% | $R_M(0011 \mid \mathcal{E}_{11})$=20% |
| **Asbestos only** $(Y_{00}=0, Y_{01}=1, Y_{10}=0, Y_{11}=1)$ |  | $R_Z(0101 \mid \mathcal{E}_{11})$=10% | $R_M(0101 \mid \mathcal{E}_{11})$=50% |
| **Parallel** $(Y_{00}=0, Y_{01}=1, Y_{10}=1, Y_{11}=1)$ |  | $R_Z(0111 \mid \mathcal{E}_{11})$=40% | $R_M(0111 \mid \mathcal{E}_{11})$=40% |

**For each factor j, the interactive total must equal the single-factor total:**

$$\underbrace{R_j(0001 \mid \mathcal{E}_{11}) + R_j(0111 \mid \mathcal{E}_{11})}_{\text{Interaction types}} = \underbrace{R_j\,(0011 \mid \mathcal{E}_{11}) + R_j(0101 \mid \mathcal{E}_{11})}_{\text{Independent types}}$$

# Theorem 2: any balanced rule is point-identified

**Assumptions**

$$R_{\mathrm{Z}}(0001 \mid \mathcal{E}_{11}) + R_{\mathrm{Z}}(0111 \mid \mathcal{E}_{11}) = R_{\mathrm{Z}}\ (0011 \mid \mathcal{E}_{11}) + R_{\mathrm{Z}}(0101 \mid \mathcal{E}_{11})$$
$$R_{\mathrm{M}}(0001 \mid \mathcal{E}_{11}) + R_{\mathrm{M}}(0111 \mid \mathcal{E}_{11}) = R_{\mathrm{M}}\ (0011 \mid \mathcal{E}_{11}) + R_{\mathrm{M}}(0101 \mid \mathcal{E}_{11})$$

$+$

*Randomization:* $(Z, M) \perp (Y_{00}, Y_{01}, Y_{10}, Y_{11})$

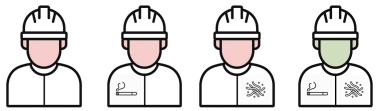
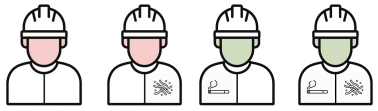
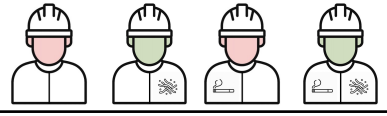
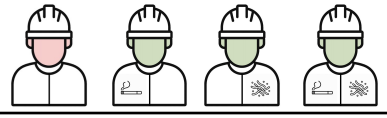
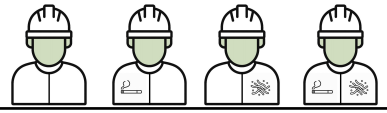
*Monotonicity:* $Y_{00} \leq Y_{01} \leq Y_{11}, Y_{00} \leq Y_{10} \leq Y_{11}$

**Identification**

For $j \in \{Z, M\}$, *we have:*

$$\begin{aligned} R_j(\mathcal{E}_{11}) &= \sum_{y_{00}y_{01}y_{10}y_{11}} R_j(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11}) \\ &= \{R_j(1111 \mid \mathcal{E}_{11}) - R_j(0111 \mid \mathcal{E}_{11})\}\frac{\xi_{00}}{\xi_{11}} + \{R_j(0101 \mid \mathcal{E}_{11}) - R_j(0001 \mid \mathcal{E}_{11})\}\frac{\xi_{01}}{\xi_{11}} \\ &\quad + \{R_j(0011 \mid \mathcal{E}_{11}) - R_j(0001 \mid \mathcal{E}_{11})\}\frac{\xi_{10}}{\xi_{11}} + R_j(0001 \mid \mathcal{E}_{11})\frac{\xi_{00}}{\xi_{11}} \end{aligned}$$

# Balance Condition Example - 1

| Counterfactual pattern (worlds with cancer) | Responsibility Rule |
|---|---|
|  | **Smoking: 50%  Asbestos: 50%** |
|  | **Smoking: 100%  Asbestos: 0%** |
|  | **Smoking: 0%  Asbestos: 100%** |
|  | **Smoking: 50%  Asbestos:50%** |
|  | **Smoking: 0%  Asbestos: 0%** |

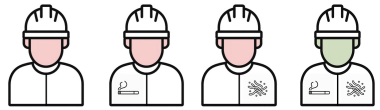
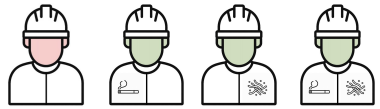
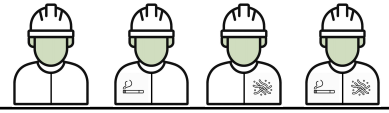
$$R_{\mathrm{Z}}(\mathcal{E}_{11}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_{\mathrm{Z}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$$

$$= \frac{\xi_{11} + \xi_{10} - \xi_{01} - \xi_{00}}{2\xi_{11}}$$

$$R_{\mathrm{M}}(\mathcal{E}_{11}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_{\mathrm{M}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$$

$$= \frac{\xi_{11} - \xi_{10} + \xi_{01} - \xi_{00}}{2\xi_{11}}$$

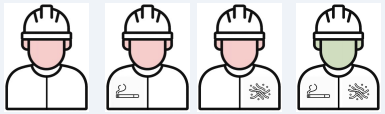
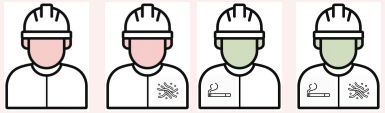
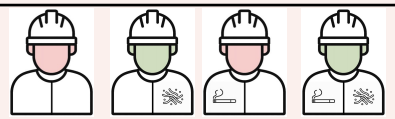
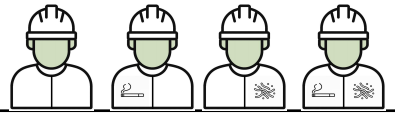
# Balance Condition Example - 2

| Counterfactual pattern (worlds with cancer) | Responsibility Rule |
|---|---|
|  | **Smoking: 30%  Asbestos: 30%** |
|  | **Smoking: 60%  Asbestos: 20%** |
|  | **Smoking: 10%  Asbestos: 50%** |
|  | **Smoking: 40%  Asbestos:40%** |
|  | **Smoking: 10%  Asbestos: 10%** |

$$R_{\mathrm{Z}}(\mathcal{E}_{11}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_{\mathrm{Z}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$$
$$= \frac{0.3\xi_{11} + 0.3\xi_{10} - 0.2\xi_{01} - 0.3\xi_{00}}{2\xi_{11}}$$

$$R_{\mathrm{M}}(\mathcal{E}_{11}) = \sum_{y_{00}y_{01}y_{10}y_{11}} R_{\mathrm{M}}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$$
$$= \frac{-0.3\xi_{11} - 0.1\xi_{10} + 0.2\xi_{01} - 0.3\xi_{00}}{2\xi_{11}}$$

# Workplace oriented allocation – unbalanced rule

| Story | Counterfactual pattern (worlds with cancer) | Type-specific responsibility $R_z(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$ | Type-specific responsibility $R_M(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})$ |
|---|---|---|---|
| **Synergy** $(Y_{00}=0, Y_{01}=0, Y_{10}=0, Y_{11}=1)$ |  | $R_z(0001 \mid \mathcal{E}_{11}) = 40\ \%$ | $R_M(0001 \mid \mathcal{E}_{11}) = 60\ \%$ |
| **Smoking only** $(Y_{00}=0, Y_{01}=0, Y_{10}=1, Y_{11}=1)$ |  | $R_z(0011 \mid \mathcal{E}_{11}) = 80\ \%$ | $R_M(0011 \mid \mathcal{E}_{11}) = 20\ \%$ |
| **Asbestos only** $(Y_{00}=0, Y_{01}=1, Y_{10}=0, Y_{11}=1)$ |  | $R_z(0101 \mid \mathcal{E}_{11}) = 0\ \%$ | $R_M(0101 \mid \mathcal{E}_{11}) = 100\ \%$ |
| **Parallel** $(Y_{00}=0, Y_{01}=1, Y_{10}=1, Y_{11}=1)$ |  | $R_z(0111 \mid \mathcal{E}_{11}) = 30\ \%$ | $R_M(0111 \mid \mathcal{E}_{11}) = 70\ \%$ |
| **Background** $(Y_{00}=1, Y_{01}=1, Y_{10}=1, Y_{11}=1)$ |  | $R_z(1111 \mid \mathcal{E}_{11}) = 0\ \%$ | $R_M(1111 \mid \mathcal{E}_{11}) = 10\ \%$ |

The rule emphasises occupational asbestos, violates structural balance.
Even a smoking-only mechanism gives asbestos 20% for the hazardous workplace environment.

# Theorem 3: sharp bounds for arbitrary type-specific rules

**Assumptions**

*Randomization:* $(Z, M) \perp (Y_{00}, Y_{01}, Y_{10}, Y_{11})$ *Monotonicity:* $Y_{00} \leq Y_{01} \leq Y_{11}, Y_{00} \leq Y_{10} \leq Y_{11}$

**Sharp Bounds**

$$\alpha_j(\mathcal{E}_{11}) + \beta_j(\mathcal{E}_{11})\nu^l \leq R_j(\mathcal{E}_{11}) \leq \alpha_j(\mathcal{E}_{11}) + \beta_j(\mathcal{E}_{11})\nu^u,$$

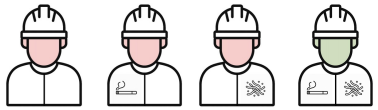
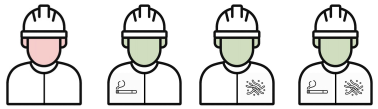
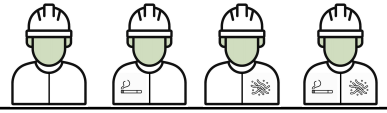
$$\nu^\ell = \max\{\xi_{01} + \xi_{10} - \xi_{11}, \xi_{00}\}, \nu^u = \min\{\xi_{01}, \xi_{10}\}$$

$$\begin{aligned}\alpha_j(\mathcal{E}_{11}) &= \{R_j(1111 \mid \mathcal{E}_{11}) - R_j(0111 \mid \mathcal{E}_{11})\}\frac{\xi_{00}}{\xi_{11}} + \{R_j(0101 \mid \mathcal{E}_{11}) - R_j(0001 \mid \mathcal{E}_{11})\}\frac{\xi_{01}}{\xi_{11}} \\ &\quad + \{R_j(0011 \mid \mathcal{E}_{11}) - R_j(0001 \mid \mathcal{E}_{11})\}\frac{\xi_{10}}{\xi_{11}} + R_j(0001 \mid \mathcal{E}_{11})\frac{\xi_{00}}{\xi_{11}}\end{aligned}$$

$$\begin{aligned}\beta_j(\mathcal{E}_{11}) &= \sum_{y_{00}y_{01}y_{10}y_{11}} R_j(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11})\mathrm{pr}(y_{00}y_{01}y_{10}y_{11} \mid \mathcal{E}_{11}) \\ &= \{R_j(0111 \mid \mathcal{E}_{11}) - R_j(0101 \mid \mathcal{E}_{11}) + R_j(0001 \mid \mathcal{E}_{11}) - R_j(0011 \mid \mathcal{E}_{11})\}\frac{1}{\xi_{11}}\end{aligned}$$

# Workplace-oriented rule: sharp bounds

$$\alpha_j(\mathcal{E}_{11}) + \beta_j(\mathcal{E}_{11})\nu^l \leq R_j(\mathcal{E}_{11}) \leq \alpha_j(\mathcal{E}_{11}) + \beta_j(\mathcal{E}_{11})\nu^u$$

| Counterfactual pattern (worlds with cancer) | Fair Responsibility Rule |
|---|---|
|  | **Smoking: 40%  Asbestos: 60%** |
|  | **Smoking: 80%  Asbestos: 20%** |
|  | **Smoking: 0%  Asbestos:100%** |
|  | **Smoking:30%  Asbestos:70%** |
|  | **Smoking: 0%  Asbestos: 10%** |

$$\nu^{\ell} = \max\{\xi_{01} + \xi_{10} - \xi_{11}, \xi_{00}\},\ \nu^u = \min\{\xi_{01}, \xi_{10}\}$$

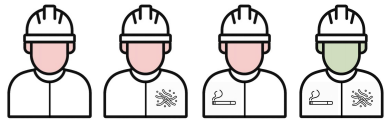
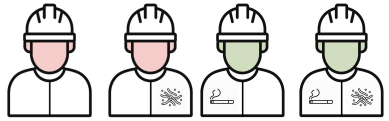
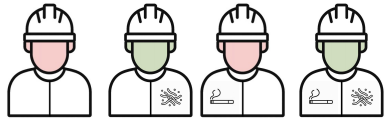
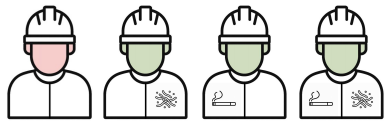
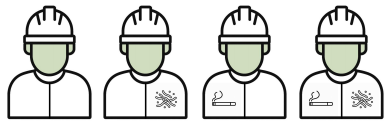
$$\alpha_Z(\mathcal{E}_{11}) = \frac{0.4\xi_{11} + 0.4\xi_{10} - 0.4\xi_{01} - 0.3\xi_{00}}{\xi_{11}}$$

$$\beta_Z(\mathcal{E}_{11}) = \frac{-0.10\xi_{00}}{\xi_{11}}$$

$$\alpha_M(\mathcal{E}_{11}) = \frac{0.6\xi_{11} - 0.4\xi_{10} + 0.4\xi_{01} - 0.6\xi_{00}}{\xi_{11}}$$

$$\beta_M(\mathcal{E}_{11}) = \frac{0.10\xi_{00}}{\xi_{11}}$$

# Three type-specific allocation rules

| Counterfactual pattern (worlds with cancer) | Active-only rule (identifiable) | Balanced rule (identifiable) | Workplace-oriented rule (sharp bounds) |
|---|---|---|---|
|  | Smoking: 50% Asbestos: 50% | Smoking: 30% Asbestos: 30% | Smoking: 40% Asbestos: 60% |
|  | Smoking: 100% Asbestos: 0% | Smoking: 60% Asbestos: 20% | Smoking: 80% Asbestos: 20% |
|  | Smoking: 0% Asbestos: 100% | Smoking: 10% Asbestos: 50% | Smoking: 0% Asbestos: 100% |
|  | Smoking: 50% Asbestos:50% | Smoking: 40% Asbestos:40% | Smoking: 30% Asbestos:70% |
|  | Smoking: 0% Asbestos: 0% | Smoking: 10% Asbestos: 10% | Smoking: 0% Asbestos: 10% |

# Outline

PART IV OF IV

Part IV begins: numerical scenarios, rule comparisons, and takeaways.

1 **Motivation and problem**

Causes of effects, why separate but-for scores do not form an allocation.

2 **Counterfactual framework**

Potential outcomes, latent causal types, and average causal responsibility.

3 **Identification results**

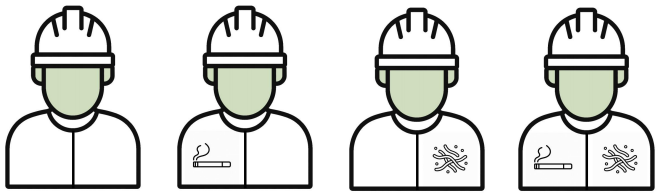
Active-only and balanced rules, point identification, and sharp bounds.

4 **Numerical illustrations**

Two smoking–asbestos scenarios, rule comparisons, and final takeaways.

# Scenario I: synergy dominates

A responsibility rule is averaged over the posterior distribution of latent types.



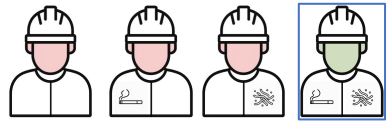
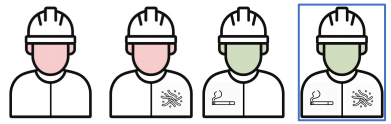
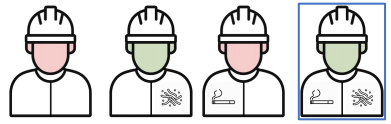
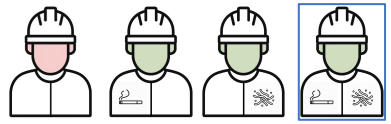
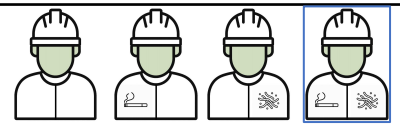
$\xi_{00} = 0.12\%$ $\xi_{10} = 0.95\%$ $\xi_{01} = 0.67\%$ $\xi_{11} = 4.51\%$

**Bob's claim**

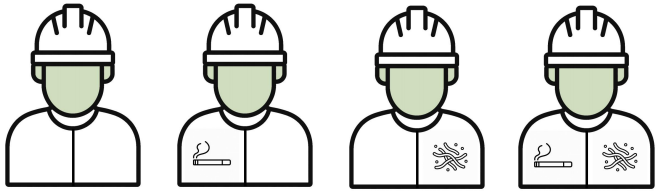
Observed case:
smoking + asbestos
+ lung cancer

**Total loss: £10,000**

**Dominant latent type: Synergy**

| Story | Counterfactual pattern (worlds with cancer) | Numerical Results |
|---|---|---|
| **Synergy** |  | [66.74%, 78.94%] |
| **Smoking only** |  | [6.21%,18.40%] |
| **Asbestos only** |  | [0.00%,12.20%] |
| **Parallel** |  | [0.00%,12.20%] |
| **Background** |  | 2.66% |

# Scenario I: synergy dominates



$\xi_{00} = 0.12\%$ $\xi_{10} = 0.95\%$ $\xi_{01} = 0.67\%$ $\xi_{11} = 4.51\%$

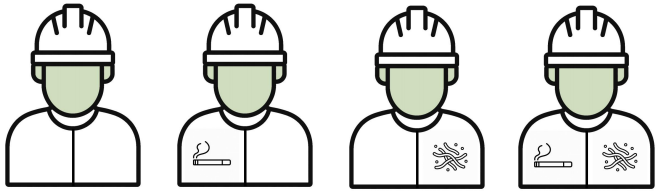
| Types | Active-only rule (identifiable) | Balanced rule (identifiable) | Workplace rule (sharp bounds) |
|---|---|---|---|
| 0001 | Smoking: 50%<br>Asbestos: 50% | Smoking: 30%<br>Asbestos: 30% | Smoking: 40%<br>Asbestos: 60% |
| 0101 | Smoking: 100%<br>Asbestos: 0% | Smoking: 60%<br>Asbestos: 20% | Smoking: 80%<br>Asbestos: 20% |
| 0011 | Smoking: 0%<br>Asbestos: 100% | Smoking: 10%<br>Asbestos: 50% | Smoking: 0%<br>Asbestos: 100% |
| 0111 | Smoking: 50%<br>Asbestos:50% | Smoking: 40%<br>Asbestos:40% | Smoking: 30%<br>Asbestos:70% |
| 1111 | Smoking: 0%<br>Asbestos: 0% | Smoking: 10%<br>Asbestos: 10% | Smoking: 0%<br>Asbestos: 10% |

| | Active-only rule (identifiable) | Balanced rule (identifiable) | Workplace-oriented rule (sharp bounds) |
|---|---|---|---|
| **Smoking** | **51.77%** | **32.55%** | **[40.20%, 41.42%]** |
| **Asbestos** | **45.57%** | **30.07%** | **[56.19%, 57.41%]** |
| **Background** | **2.66%** | **37.38%** | **2.39%** |

# Scenario II: strong independent asbestos risk

A responsibility rule is averaged over the posterior distribution of latent types.

**Dominant latent type: Asbestos only**

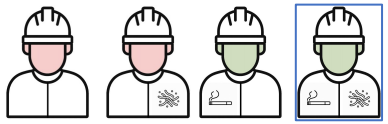
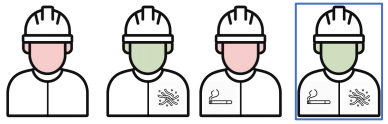
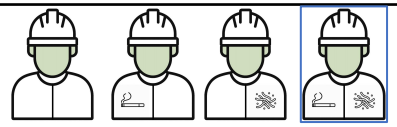


$\xi_{00} = 0.12\%$ $\xi_{10} = 0.95\%$ $\xi_{01} = 4.00\%$ $\xi_{11} = 4.51\%$

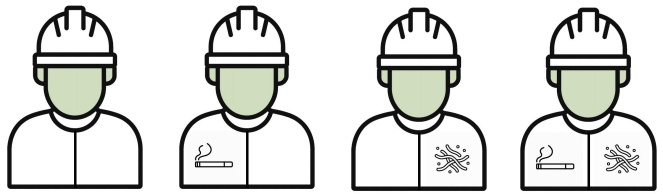
## Bob's claim

Observed case:
smoking + asbestos
+ lung cancer

**Total loss: £10,000**

| Story | Counterfactual pattern (worlds with cancer) | Numerical Results |
|---|---|---|
| **Synergy** |  | [0.00%,11.31%] |
| **Smoking only** |  | [0.00%,11.31%] |
| **Asbestos only** |  | [67.63%,78.94%] |
| **Parallel** |  | [7.10%,18.40%] |
| **Background** |  | 2.66% |

# Scenario II: strong independent asbestos risk



$\xi_{00} = 0.12\%$ $\xi_{10} = 0.95\%$ $\xi_{01} = 4.00\%$ $\xi_{11} = 4.51\%$

| Types | Active-only rule (identifiable) | Balanced rule (identifiable) | Workplace rule (sharp bounds) |
|---|---|---|---|
| 0001 | Smoking: 50%<br>Asbestos: 50% | Smoking: 30%<br>Asbestos: 30% | Smoking: 40%<br>Asbestos: 60% |
| 0101 | Smoking: 100%<br>Asbestos: 0% | Smoking: 60%<br>Asbestos: 20% | Smoking: 80%<br>Asbestos: 20% |
| 0011 | Smoking: 0%<br>Asbestos: 100% | Smoking: 10%<br>Asbestos: 50% | Smoking: 0%<br>Asbestos: 100% |
| 0111 | Smoking: 50%<br>Asbestos:50% | Smoking: 40%<br>Asbestos:40% | Smoking: 30%<br>Asbestos:70% |
| 1111 | Smoking: 0%<br>Asbestos: 0% | Smoking: 10%<br>Asbestos: 10% | Smoking: 0%<br>Asbestos: 10% |

| | Active-only rule (identifiable) | Balanced rule (identifiable) | Workplace-oriented rule (sharp bounds) |
|---|---|---|---|
| **Smoking** | **14.86%** | **17.78%** | **[10.04%, 11.18%]** |
| **Asbestos** | **82.48%** | **44.83%** | **[86.43%, 87.56%]** |
| **Background** | **2.66%** | **37.38%** | **2.39%** |

# Takeaways

Causal attribution asks how to divide responsibility for an adverse outcome that has already occurred.

### 1 Retrospective question

Unlike treatment effect analysis, the target is causes of effects: given that Bob developed lung cancer, what share should each factor bear?

### 2 Latent causal types

The ideal answer depends on Bob's potential outcomes under all exposure combinations, but only one world is observed.

### 3 Average causal responsibility

We average type-specific legal responsibility shares using posterior probabilities of the possible causal types.

### 4 Identification and bounds

Balanced rules are identified from observed risks; otherwise, sharp bounds quantify the remaining counterfactual uncertainty.

# Thank you !