Multi-Resolution Inference: An Engineering (Engineered?) Foundation for Statistical Inference

Keli Liu and Xiao-Li Meng

Department of Statistics, Harvard University

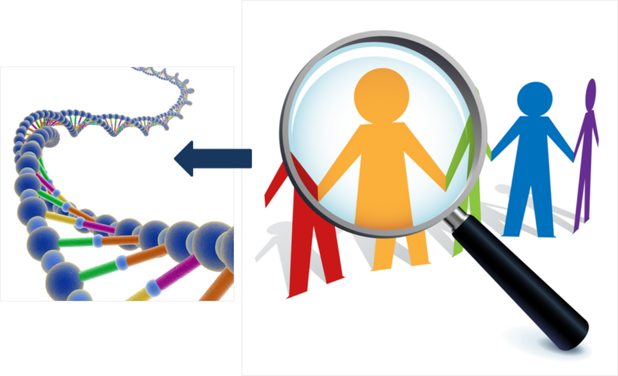
Liu, Keli and Xiao-Li Meng (2014). "A Fruitful Resolution To Simpson's Paradox via Multi-Resolution Inference." *The American Statistician*, 68: 17-29.

Meng, Xiao-Li (2014). "A Trio of Inference Problems That Could Win You a Nobel Prize in Statistics (if you help fund it)." *In the Past, Present, and Future of Statistical Science (Eds: X. Lin, et. al.)*, 535-560.

What Does Big Data Mean For You?

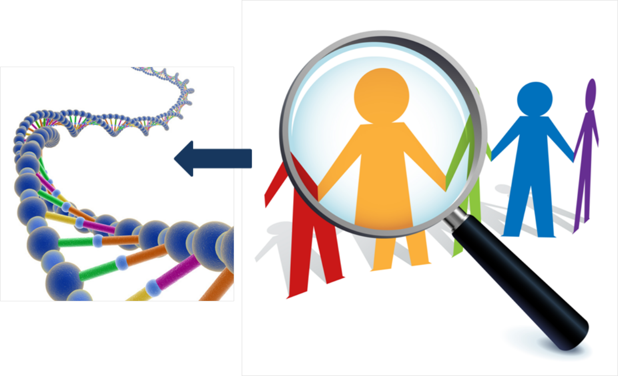
What Does Big Data Mean For You?

We see you and others more clearly.



What Does Big Data Mean For You?

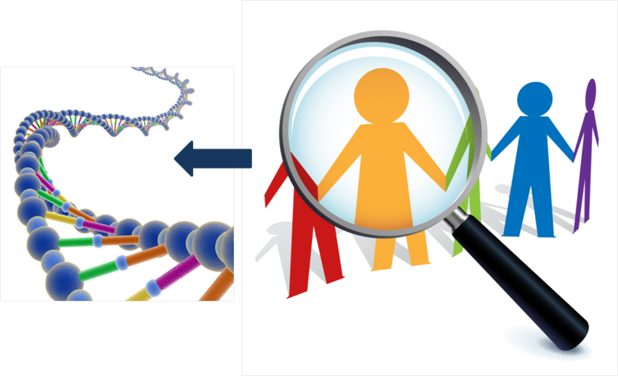
We see you and others more clearly.



- individualized medicine

What Does Big Data Mean For You?

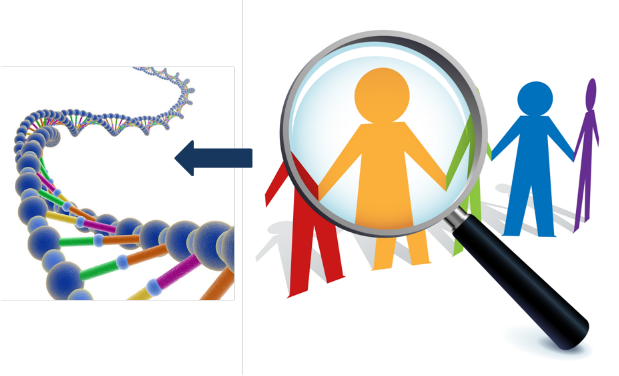
We see you and others more clearly.



- individualized medicine
- individualized education

What Does Big Data Mean For You?

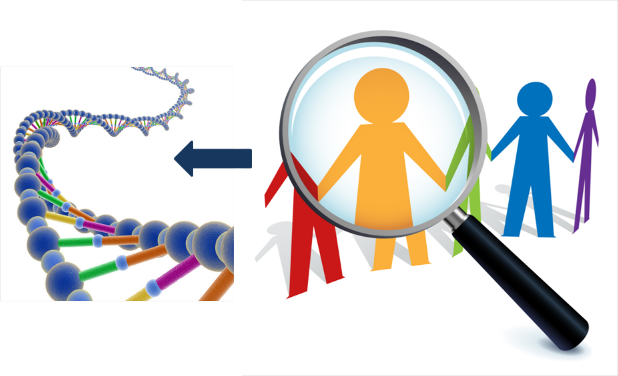
We see you and others more clearly.



- individualized medicine
- individualized education
- individualized news

What Does Big Data Mean For You?

We see you and others more clearly.



- individualized medicine
- individualized education
- individualized news
- individualized marketing

The Gift and Curse of Big Data

Gift: Treatment for you based only on data from people like you.

The Gift and Curse of Big Data

Gift: Treatment for you based only on data from people like you.

Curse: No one is perfectly like you.

What Does This Mean for Inference?

No two different objects are alike in all respects but one. It is obviously true; but it has no bearing on induction, where we deal with objects which we well know are, like all existing things, alike in numberless respects and unlike in numberless other respects.

Charles Peirce

The Statistically Payneful Question of Individualized Medicine

Ms.Payne, ☺* comes to us to be treated for kidney stones.

The Statistically Payneful Question of Individualized Medicine

Ms.Payne, ☺* comes to us to be treated for kidney stones.

Two treatments available: A and B. We need to make a treatment decision for **Ms.Payne** based on data from **others**.

The Statistically Payneful Question of Individualized Medicine

Ms.Payne, ☺* comes to us to be treated for kidney stones.

Two treatments available: A and B. We need to make a treatment decision for **Ms.Payne** based on data from **others**.

1. **Ideal Question**: If we could have all the data we wanted, how would we make a treatment decision for ☺*?

The Statistically Payneful Question of Individualized Medicine

Ms.Payne, ☺* comes to us to be treated for kidney stones.

Two treatments available: A and B. We need to make a treatment decision for **Ms.Payne** based on data from **others**.

1. **Ideal Question**: If we could have all the data we wanted, how would we make a treatment decision for ☺*?

2. **Estimation Question**: Given the actual data, what quantities needed for the ideal decision making process can we estimate? What quantities can we not?

The Statistically Payneful Question of Individualized Medicine

Ms.Payne, ☺* comes to us to be treated for kidney stones.

Two treatments available: A and B. We need to make a treatment decision for **Ms.Payne** based on data from **others**.

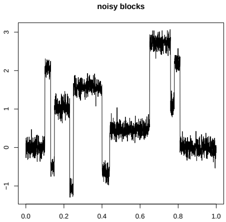
1. **Ideal Question**: If we could have all the data we wanted, how would we make a treatment decision for ☺*?
2. **Estimation Question**: Given the actual data, what quantities needed for the ideal decision making process can we estimate? What quantities can we not?
3. **Inferential Question**: Given information constraints, how do we best approximate the ideal decision making process by an achievable one?

A Wavelets Inspired Reformulation

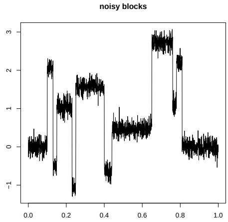
Decompose aggregate signal into signals at varying resolutions.

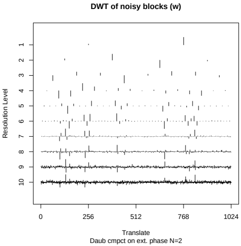
A Wavelets Inspired Reformulation

Decompose aggregate signal into signals at varying resolutions.



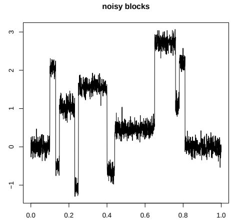
Decompose aggregate signal into signals at varying resolutions.

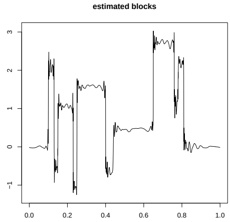


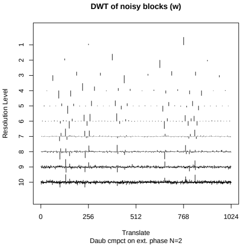


A Wavelets Inspired Reformulation

Decompose aggregate signal into signals at varying resolutions.

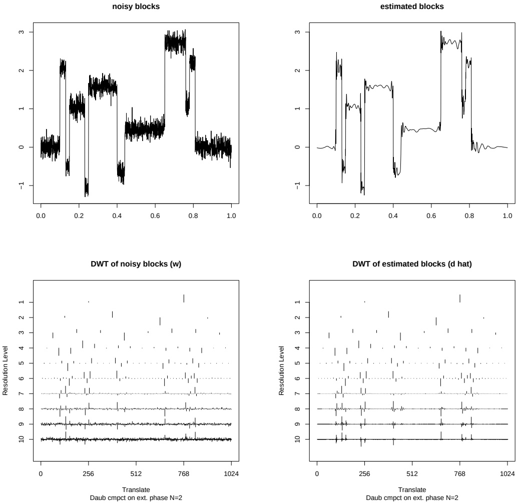






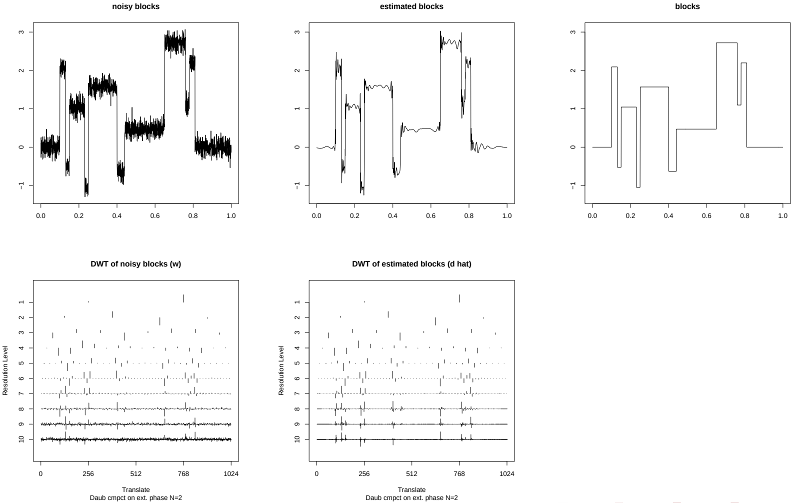
A Wavelets Inspired Reformulation

Decompose aggregate signal into signals at varying resolutions.



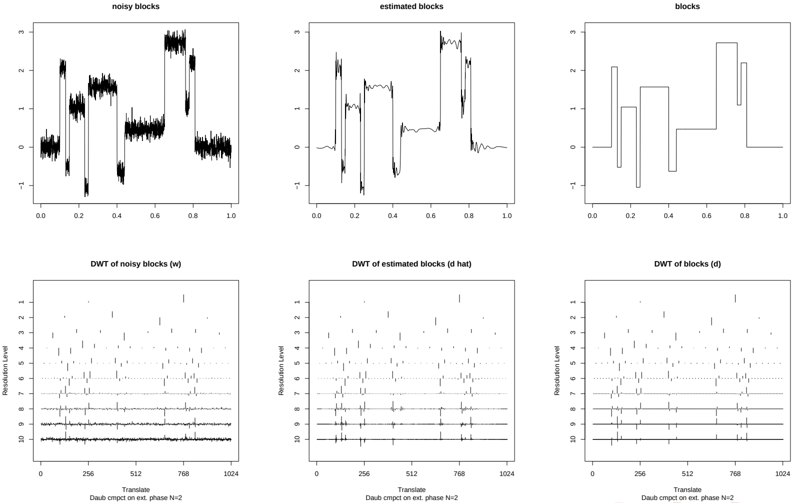
A Wavelets Inspired Reformulation

Decompose aggregate signal into signals at varying resolutions.



A Wavelets Inspired Reformulation

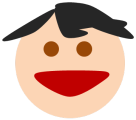
Decompose aggregate signal into signals at varying resolutions.



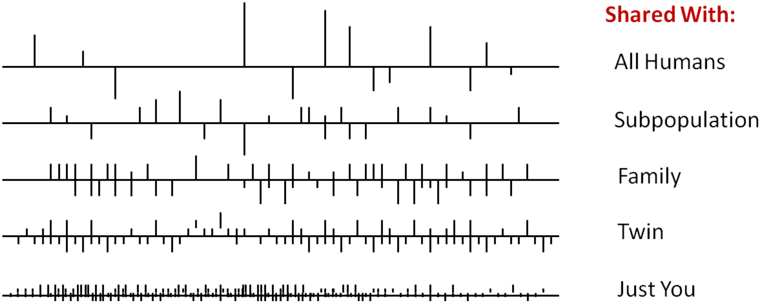
A Wavelet View of Individuality

An individual is a collection of signals (idiosyncracies).

A Wavelet View of Individuality



An individual is a collection of signals (idiosyncracies).



Kidney Stone Treatment

C. R. Charig, D. R. Webb, S. R. Payne, O. E. Wickham (March 1986)

Br Med J (Clin Res Ed) 292 (6524): 879-882.

Treatment A	Treatment B
78% (273/350)	**83%** (289/350)

Kidney Stone Treatment

C. R. Charig, D. R. Webb, S. R. Payne, O. E. Wickham (March 1986)

Br Med J (Clin Res Ed) 292 (6524): 879-882.

Treatment A	Treatment B
78% (273/350)	**83%** (289/350)

	Treatment A	Treatment B
Small Stone	**93%** (81/87)	87% (234/270)
Large Stone	**73%** (192/263)	69% (55/80)

Kidney Stone Treatment

C. R. Charig, D. R. Webb, S. R. Payne, O. E. Wickham (March 1986)

Br Med J (Clin Res Ed) 292 (6524): 879-882.

Treatment A	Treatment B
78% (273/350)	83% (289/350)

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

- Open Surgery (A) vs. Percutaneous Nephrolithotomy (B).

Kidney Stone Treatment

C. R. Charig, D. R. Webb, S. R. Payne, O. E. Wickham (March 1986)

Br Med J (Clin Res Ed) 292 (6524): 879-882.

Treatment A	Treatment B
78% (273/350)	**83%** (289/350)

	Treatment A	Treatment B
Small Stone	**93%** (81/87)	87% (234/270)
Large Stone	**73%** (192/263)	69% (55/80)

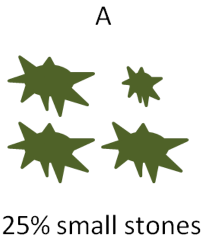
- Open Surgery (A) vs. Percutaneous Nephrolithotomy (B).
- **Problem:** Stone size and treatment are confounded.

Comparing Apples to Oranges

- 78% and 83% are population success rates. Which *population*?

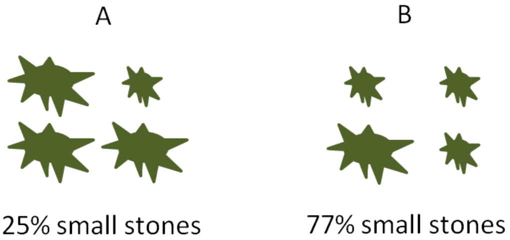
Comparing Apples to Oranges

- 78% and 83% are population success rates. Which *population*?



Comparing Apples to Oranges

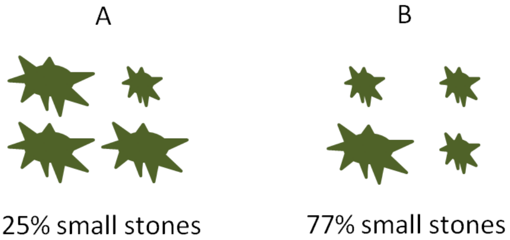
- 78% and 83% are population success rates. Which *population*?



- **Conditioning** forces populations under comparison to be more similar. *More conditioning = more similar.*

Comparing Apples to Oranges

- 78% and 83% are population success rates. Which *population*?



- **Conditioning** forces populations under comparison to be more similar. *More conditioning* = *more similar.*
- **Are more "similar" individuals always more comparable?**

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

- What if *side effect* of treatment is hair loss and the above are *post*-treatment pictures?

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

- What if *side effect* of treatment is hair loss and the above are *post*-treatment pictures?

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

- What if *side effect* of treatment is hair loss and the above are *post*-treatment pictures?

Two Lessons

1. Individuals **apparently** similar may be **intrinsically** dissimilar.

Beauty Is Only Skin Deep

- Are Mark and Ben similar?

- What if *side effect* of treatment is hair loss and the above are *post*-treatment pictures?

Two Lessons

1. Individuals **apparently** similar may be **intrinsically** dissimilar.
2. Definition of intrinsic similarity depends on the treatment.

Another Look at Kidney Stones

- Outcome is relief ($Y = 1$) or persistence ($Y = 0$) of pain.

Another Look at Kidney Stones

- Outcome is relief ($Y = 1$) or persistence ($Y = 0$) of pain.
- A and B are two types of pain medication whose indirect effects include changing kidney stone size.

Another Look at Kidney Stones

- Outcome is relief ($Y = 1$) or persistence ($Y = 0$) of pain.
- A and B are two types of pain medication whose indirect effects include changing kidney stone size.

Treatment A	Treatment B
78% (273/350)	83% (289/350)

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

Another Look at Kidney Stones

- Outcome is relief ($Y = 1$) or persistence ($Y = 0$) of pain.
- A and B are two types of pain medication whose indirect effects include changing kidney stone size.

Treatment A	Treatment B
78% (273/350)	83% (289/350)

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

- **Which treatment is better assuming randomization?**

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.
- Is ☺ is like ☺′?

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.
- Is ☺ is like ☺′?
  - Suppose we measured characteristics, C, of ☺ and ☺′.

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.
- Is ☺ is like ☺′?
  - Suppose we measured characteristics, C, of ☺ and ☺′.
  - ☺ is **intrinsically dissimilar** from ☺′ if $C(☺) \neq C(☺')$.

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.
- Is ☺ is like ☺′?
  - Suppose we measured characteristics, C, of ☺ and ☺′.
  - ☺ is **intrinsically dissimilar** from ☺′ if $C(☺) \neq C(☺')$.
- **Problem:** We only measure characteristics of the *realized state*, $☺_t \equiv (☺, t) \in \Omega \times \mathbb{T}$, where t is the treatment assigned.

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.
- Is ☺ is like ☺′?
  - Suppose we measured characteristics, C, of ☺ and ☺′.
  - ☺ is **intrinsically dissimilar** from ☺′ if $C(☺) \neq C(☺')$.
- **Problem:** We only measure characteristics of the *realized state*, $☺_t \equiv (☺, t) \in \Omega \times \mathbb{T}$, where t is the treatment assigned.

A Platonic Analogy

$$\text{form} : ☺ :: \text{particular} : ☺_t$$

A Page from Plato's Book

- Ω is population of interest with *individuals* ☺.

- Is ☺ is like ☺′?
  - Suppose we measured characteristics, C, of ☺ and ☺′.
  - ☺ is **intrinsically dissimilar** from ☺′ if $C(☺) \neq C(☺')$.

- **Problem:** We only measure characteristics of the *realized state*, $☺_t \equiv (☺, t) \in \Omega \times \mathbb{T}$, where t is the treatment assigned.

A Platonic Analogy

form : ☺ :: particular : $☺_t$

- No direct access to ☺, forms—only see functions $C(☺_t)$.

Intrinsic = Invariance to Treatment

- A characteristic, $C(☺_t)$, of the realized state is also an *intrinsic characteristic* of ☺ if $C(☻_t) = C(☻)$.

Intrinsic = Invariance to Treatment

- A characteristic, $C(☺_t)$, of the realized state is also an *intrinsic characteristic* of ☺ if $C(☻_t) = C(☻)$.

- If A ($t = 0$) and B ($t = 1$) are open surgery and nephrolithotomy:

Intrinsic = Invariance to Treatment

- A characteristic, $C(☺_t)$, of the realized state is also an *intrinsic characteristic* of ☺ if $C(☻_t) = C(☻)$.

- If A ($t = 0$) and B ($t = 1$) are open surgery and nephrolithotomy:

Intrinsic = Invariance to Treatment

- A characteristic, $C(☺_t)$, of the realized state is also an *intrinsic characteristic* of ☺ if $C(☺_t) = C(☺)$.

- If A ($t = 0$) and B ($t = 1$) are open surgery and nephrolithotomy:

- If A and B are drugs with size reduction indirect effect:

Intrinsic = Invariance to Treatment

- A characteristic, $C(☺_t)$, of the realized state is also an *intrinsic characteristic* of ☺ if $C(☺_t) = C(☺)$.

- If A ($t = 0$) and B ($t = 1$) are open surgery and nephrolithotomy:

- If A and B are drugs with size reduction indirect effect:

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.
- Observed data are characteristics, $C(☺_t)$, of realized state, $☺_t$—*not necessarily intrinsic*.

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.
- Observed data are characteristics, $C(☺_t)$, of realized state, $☺_t$—*not necessarily intrinsic*.

Potential outcomes (Rubin 2005)

- $C_0(☺) \equiv C(☺_0)$ and $C_1(☺) \equiv C(☺_1)$, are intrinsic.

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.
- Observed data are characteristics, $C(☺_t)$, of realized state, $☺_t$—*not necessarily intrinsic*.

Potential outcomes (Rubin 2005)

- $C_0(☺) \equiv C(☺_0)$ and $C_1(☺) \equiv C(☺_1)$, are intrinsic.
- Treatment determines which potential outcome is observed:

$$C(☺_t) = t \cdot C_1(☺) + (1-t) \cdot C_0(☺)$$

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.
- Observed data are characteristics, $C(☺_t)$, of realized state, $☺_t$—*not necessarily intrinsic*.

Potential outcomes (Rubin 2005)

- $C_0(☺) \equiv C(☺_0)$ and $C_1(☺) \equiv C(☺_1)$, are intrinsic.
- Treatment determines which potential outcome is observed:

$$C(☺_t) = t \cdot C_1(☺) + (1 - t) \cdot C_0(☺)$$

- Not Intrinsic: post-treatment kidney stone size.

What Intrinsic Characteristics Are Observed?

- Random sampling of ☺ from Ω and randomized treatment assignment to ☺.
- Observed data are characteristics, $C(☺_t)$, of realized state, $☺_t$—*not necessarily intrinsic*.

Potential outcomes (Rubin 2005)

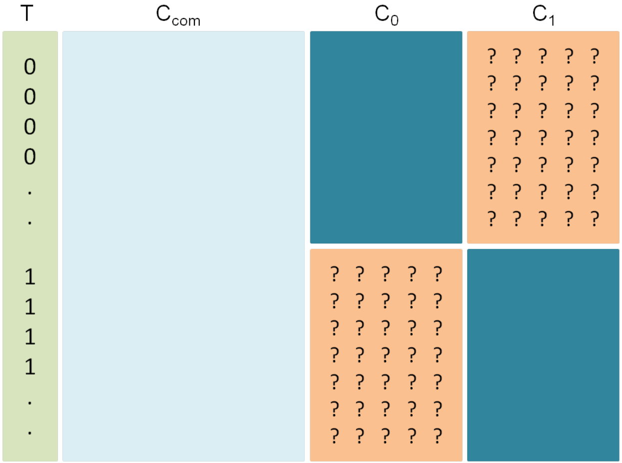
- $C_0(☺) \equiv C(☺_0)$ and $C_1(☺) \equiv C(☺_1)$, are intrinsic.
- Treatment determines which potential outcome is observed:

$$C(☺_t) = t \cdot C_1(☺) + (1 - t) \cdot C_0(☺)$$

- Not Intrinsic: post-treatment kidney stone size.
- Intrinsic: post-treatment kidney stone size under treatment 1.

Observed Intrinsic Characteristics

Observed Intrinsic Characteristics



- **Without assumptions**, e.g. $C_0 = C_1$, C_{1-t} is always missing in group $T = t$.

The Resolution Perspective

- Which individuals are comparable to Ms.Payne, ☺*?

The Resolution Perspective

- Which individuals are comparable to Ms.Payne, ☺*?
- Let $C(☺)$ measure an intrinsic characteristic.

The Resolution Perspective

- Which individuals are comparable to Ms.Payne, $☺^*$?
- Let $C(☺)$ measure an intrinsic characteristic.
- The $☺^*$ **relevant subpopulation** is

$$\Omega_C(☺^*) = \{☺ : C(☺) = C(☺^*)\}.$$

The Resolution Perspective

- Which individuals are comparable to Ms.Payne, $☺^*$?
- Let $C(☺)$ measure an intrinsic characteristic.
- The $☻^*$ **relevant subpopulation** is

$$\Omega_C(☺^*) = \{☺ : C(☺) = C(☺^*)\}.$$

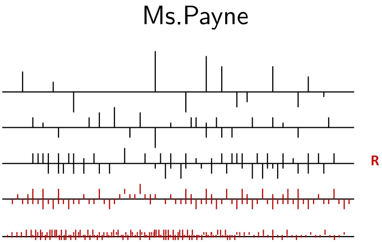
- $R =$dim(C) is the resolution.

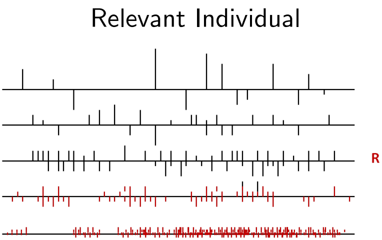
The Resolution Perspective

- Which individuals are comparable to Ms.Payne, ☺*?
- Let C(☺) measure an intrinsic characteristic.
- The ☺* **relevant subpopulation** is

$$\Omega_C(☺^*) = \{☺ : C(☺) = C(☺^*)\}.$$

- R =dim(C) is the resolution.





The Ideal Question and Resolution

- Let $Y_t(☺)$ denote success (1) or failure (0) of treatment t on individual ☺.

The Ideal Question and Resolution

- Let $Y_t(☺)$ denote success (1) or failure (0) of treatment t on individual ☺.

The Ideal Question

Is $Y_1(☺^*) - Y_0(☺^*)$ less than, greater than, or equal to 0?

The Ideal Question and Resolution

- Let $Y_t(☺)$ denote success (1) or failure (0) of treatment t on individual ☺.

The Ideal Question

Is $Y_1(☺^*) - Y_0(☺^*)$ less than, greater than, or equal to 0?

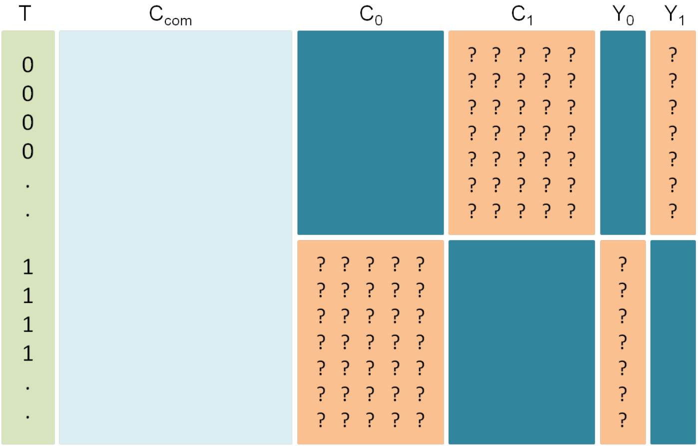
- The ideal relevant subpopulation contains only Ms.Payne herself (and her exact clones)

$$\Omega_C(☺^*) = \{☺^*\}.$$

C consists of all possible intrinsic characteristics.

Study Data

Study Data



- We will always condition on C_{com}, denoted by $C = \emptyset$.

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

$$Y_0 | ☺ \in \Omega_C(☺^*) \qquad Y_1 | ☺ \in \Omega_C(☺^*)$$

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

	$Y_0 \mid ☺ \in \Omega_C(☺^*)$	$Y_1 \mid ☺ \in \Omega_C(☺^*)$
$C = \emptyset$	**Yes**	**Yes**

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

	$Y_0 \mid ☺ \in \Omega_C(☺^*)$	$Y_1 \mid ☺ \in \Omega_C(☺^*)$
$C = \emptyset$	**Yes**	**Yes**
$C = C_0$	**Yes**	**No**

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

	$Y_0 \mid ☺ \in \Omega_C(☺^*)$	$Y_1 \mid ☺ \in \Omega_C(☺^*)$
$C = \emptyset$	**Yes**	**Yes**
$C = C_0$	**Yes**	**No**
$C = C_1$	**No**	**Yes**

The Estimation Question and Data Resolution

- **Problem:** We do not observe $Y_1(☺^*) - Y_0(☺^*)$.
- **Nor** do we observe $Y_1(☺) - Y_0(☺)$ for any ☺.
- **But** We observe $Y_t(☺)$, for $☺ \neq ☺^*$.

	$Y_0 \mid ☺ \in \Omega_C(☺^*)$	$Y_1 \mid ☺ \in \Omega_C(☺^*)$
$C = \emptyset$	**Yes**	**Yes**
$C = C_0$	**Yes**	**No**
$C = C_1$	**No**	**Yes**
$C = (C_0, C_1)$	**No**	**No**

The Gap Between Our Ideal and the Data

- Let $\text{Ave}_C(Y_t)$ be average of $Y_t(☺)$ for $☺ \in \Omega_C(☺^*)$.

The Gap Between Our Ideal and the Data

- Let $\text{Ave}_C(Y_t)$ be average of $Y_t(☺)$ for $☺ \in \Omega_C(☺^*)$.

- If $\Omega_C(☺^*) = \{☺^*\}$, then $\text{Ave}_C(Y_1) - \text{Ave}_C(Y_0)$ reduces to $Y_1(☺^*) - Y_0(☺^*)$, our ideal estimand.

The Gap Between Our Ideal and the Data

- Let $\text{Ave}_C(Y_t)$ be average of $Y_t(☺)$ for $☺ \in \Omega_C(☺^*)$.

- If $\Omega_C(☺^*) = \{☺^*\}$, then $\text{Ave}_C(Y_1) - \text{Ave}_C(Y_0)$ reduces to $Y_1(☺^*) - Y_0(☺^*)$, our ideal estimand.

- We want to choose C as rich as possible...

The Gap Between Our Ideal and the Data

- Let $\text{Ave}_C(Y_t)$ be average of $Y_t(☺)$ for $☺ \in \Omega_C(☺^*)$.

- If $\Omega_C(☺^*) = \{☺^*\}$, then $\text{Ave}_C(Y_1) - \text{Ave}_C(Y_0)$ reduces to $Y_1(☺^*) - Y_0(☺^*)$, our ideal estimand.

- We want to choose C as rich as possible...

- **BUT**, if C is too rich, we cannot estimate $\text{Ave}_C(Y_t)$ from data.

Which Philosophy for Big Data?

Frequentist OR Bayes

Which Philosophy for Big Data?

Frequentist OR Bayes

More fundamentally...

Which Philosophy for Big Data?

Frequentist OR Bayes

More fundamentally...

Coherent Joint Model OR Incoherent Piecemeal Models

Which Philosophy for Big Data?

Frequentist OR Bayes

More fundamentally...

Coherent Joint Model OR Incoherent Piecemeal Models

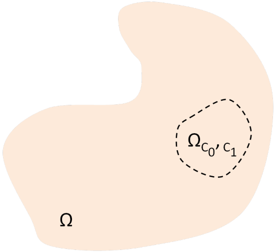
- **How to do principled corner cutting?**

Component-Wise Maximal Resolution

Estimate $\text{Ave}_C(Y_0)$

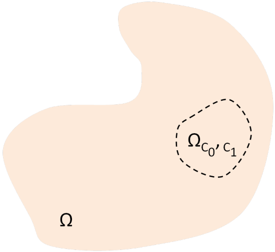
Component-Wise Maximal Resolution

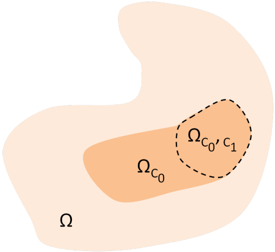
Estimate $\text{Ave}_C(Y_0)$



Component-Wise Maximal Resolution

Estimate $\text{Ave}_C(Y_0)$

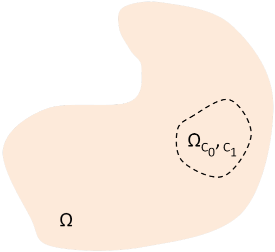


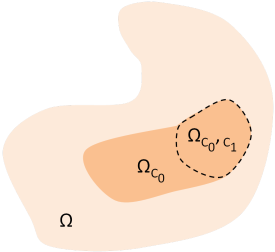


Component-Wise Maximal Resolution

Estimate $\text{Ave}_C(Y_0)$

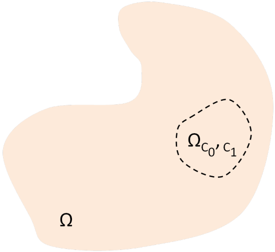
Estimate $\text{Ave}_C(Y_1)$

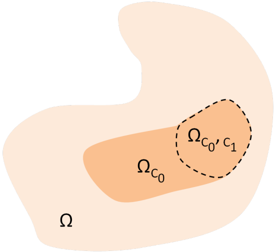




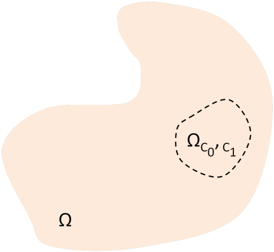
Component-Wise Maximal Resolution

Estimate $\text{Ave}_C(Y_0)$



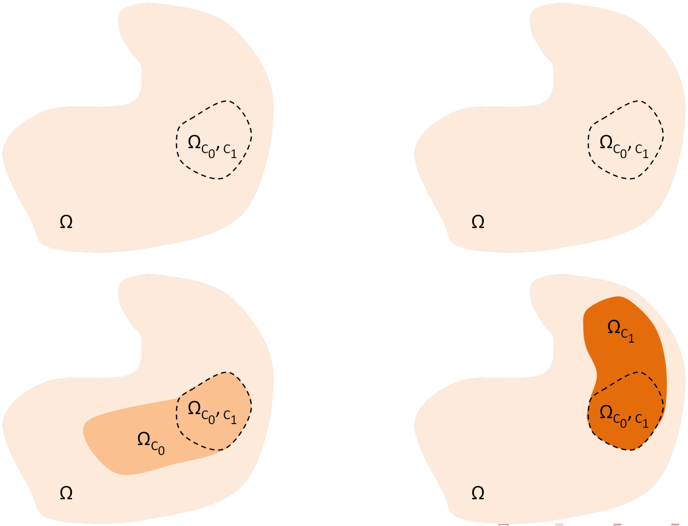


Estimate $\text{Ave}_C(Y_1)$



Component-Wise Maximal Resolution

Estimate $\text{Ave}_C(Y_0)$ Estimate $\text{Ave}_C(Y_1)$



Estimand vs Estimator

- For our **estimand**, we need to compare average of Y_0 to average of Y_1 over the same population, Ω_C.

Estimand vs Estimator

- For our **estimand**, we need to compare average of Y_0 to average of Y_1 over the same population, Ω_C.

- For our **estimator**, choose $\tilde{C}_0$ and $\tilde{C}_1$ to best approximate C. *No need to constraint* $\tilde{C}_0 = \tilde{C}_1$.

Estimand vs Estimator

- For our **estimand**, we need to compare average of Y_0 to average of Y_1 over the same population, Ω_C.

- For our **estimator**, choose $\tilde{C}_0$ and $\tilde{C}_1$ to best approximate C. *No need to constraint* $\tilde{C}_0 = \tilde{C}_1$.

- $R =$ **dim**(C) is resolution of our estimand, $\text{Ave}_C(Y_1) - \text{Ave}_C(Y_0)$.

Estimand vs Estimator

- For our **estimand**, we need to compare average of Y_0 to average of Y_1 over the same population, Ω_C.

- For our **estimator**, choose $\tilde{C}_0$ and $\tilde{C}_1$ to best approximate C. *No need to constraint* $\tilde{C}_0 = \tilde{C}_1$.

- $R = \mathbf{dim}(C)$ is resolution of our estimand, $\text{Ave}_C(Y_1) - \text{Ave}_C(Y_0)$.

- $\tilde{R} = [\mathbf{dim}(\tilde{C}_1) + \mathbf{dim}(\tilde{C}_0)]/2$ is resolution of our estimator $\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$.

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C **R** $\tilde{\mathbf{R}}$ $\tilde{\mathbf{C}}_0$ $\tilde{\mathbf{C}}_1$

The Inferential Queisiton and Resolution

Estimand	Estimator
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C	$\mathbf{R}$	$\tilde{\mathbf{R}}$	$\tilde{\mathbf{C}}_0$	$\tilde{\mathbf{C}}_1$
$C = \emptyset$	0	0	$\emptyset$	$\emptyset$

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C	**R**	$\tilde{\mathbf{R}}$	$\tilde{\mathbf{C}}_0$	$\tilde{\mathbf{C}}_1$
$C = \emptyset$	0	0	$\emptyset$	$\emptyset$
$C = C_0$	1	0.5	C_0	$\emptyset$

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C	**R**	$\tilde{\mathbf{R}}$	$\tilde{\mathbf{C}}_0$	$\tilde{\mathbf{C}}_1$
$C = \emptyset$	0	0	$\emptyset$	$\emptyset$
$C = C_0$	1	0.5	C_0	$\emptyset$
$C = C_1$	1	0.5	$\emptyset$	C_1

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C	**R**	$\tilde{\mathbf{R}}$	$\tilde{\mathbf{C}}_0$	$\tilde{\mathbf{C}}_1$
$C = \emptyset$	0	0	$\emptyset$	$\emptyset$
$C = C_0$	1	0.5	C_0	$\emptyset$
$C = C_1$	1	0.5	$\emptyset$	C_1
$C = (C_0, C_1)$	2	1	C_0	C_1

The Inferential Quesiton and Resolution

Estimand	**Estimator**
$\text{Ave}_C(Y_1 - Y_0)$	$\text{Ave}_{\tilde{C}_1}(Y_1) - \text{Ave}_{\tilde{C}_0}(Y_0)$

Operational C	**R**	$\tilde{\mathbf{R}}$	$\tilde{\mathbf{C}}_0$	$\tilde{\mathbf{C}}_1$
$C = \emptyset$	0	0	$\emptyset$	$\emptyset$
$C = C_0$	1	0.5	C_0	$\emptyset$
$C = C_1$	1	0.5	$\emptyset$	C_1
$C = (C_0, C_1)$	2	1	C_0	C_1

How best to choose an operational C**?**

Revisiting The Signal/Noise Dichotomy

Revisiting The Signal/Noise Dichotomy

The Inferential Question

How should we choose $\Omega_C(☺^*)$?

Revisiting The Signal/Noise Dichotomy

The Inferential Question

How should we choose $\Omega_C(☺^*)$?

- This is actually an age old statistical question...

Revisiting The Signal/Noise Dichotomy

The Inferential Question

How should we choose $\Omega_C(☺^*)$?

- This is actually an age old statistical question...

$$Y_t(☺^*) = \text{Ave}_C(Y_t) + \epsilon = \text{signal} \quad + \quad \text{noise.}$$

Revisiting The Signal/Noise Dichotomy

The Inferential Question

How should we choose $\Omega_C(☺^*)$?

- This is actually an age old statistical question...

$$Y_t(☺^*) = \text{Ave}_C(Y_t) + \epsilon = \text{signal} \quad + \quad \text{noise.}$$

In reality, signal and noise are two sides of the same coin.

$$\text{noise} = \text{unmodelled signal} \quad + \quad \text{intrinsic noise}$$

Noise Is Just Resolution Bias

Keli Liu and Xiao-Li Meng

- Let $C^{(1)}, C^{(2)}, \ldots$ be a sequence of vectors: (i) $C^{(r)}$ contains intrinsic characteristics (ii) $C^{(r)}$ is nested in $C^{(r+1)}$.

Noise Is Just Resolution Bias

- Let $C^{(1)}, C^{(2)}, \ldots$ be a sequence of vectors: (i) $C^{(r)}$ contains intrinsic characteristics (ii) $C^{(r)}$ is nested in $C^{(r+1)}$.
- r is the index of resolution.

Noise Is Just Resolution Bias

- Let $C^{(1)}, C^{(2)}, \ldots$ be a sequence of vectors: (i) $C^{(r)}$ contains intrinsic characteristics (ii) $C^{(r)}$ is nested in $C^{(r+1)}$.
- r is the index of resolution.

An ANOVA Decomposition

$$\text{variance at } r = \text{variance at } \infty + \sum_{s=r+1}^{\infty} E_r(\text{ signal at } s\ )^2$$

Noise Is Just Resolution Bias

- Let $C^{(1)}, C^{(2)}, \ldots$ be a sequence of vectors: (i) $C^{(r)}$ contains intrinsic characteristics (ii) $C^{(r)}$ is nested in $C^{(r+1)}$.
- r is the index of resolution.

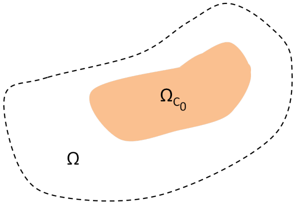
An ANOVA Decomposition

$$\text{variance at } r = \text{variance at } \infty + \sum_{s=r+1}^{\infty} E_r(\text{ signal at } s\ )^2$$

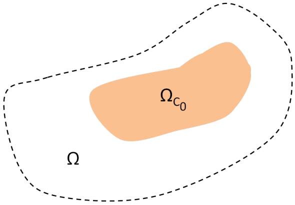
Signal at Resolution s for $s > r$.

$$\text{signal at } s = \text{Ave}_{C^{(s+1)}}(Y) - \text{Ave}_{C^{(s)}}(Y)$$

2 × Rule

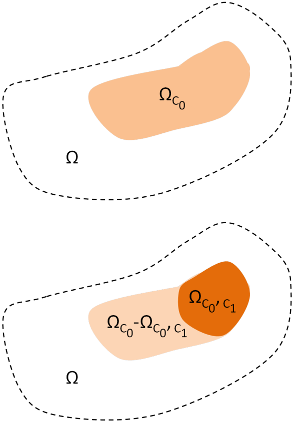


2 × Rule



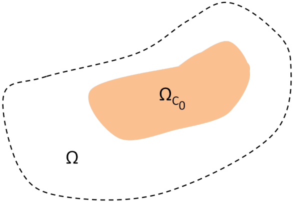
- We observe C_0 when predicting Y_0.

2 × Rule

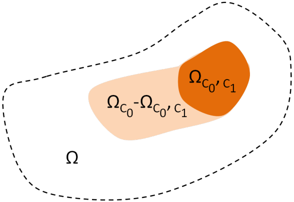


- We observe C_0 when predicting Y_0.

2 × Rule

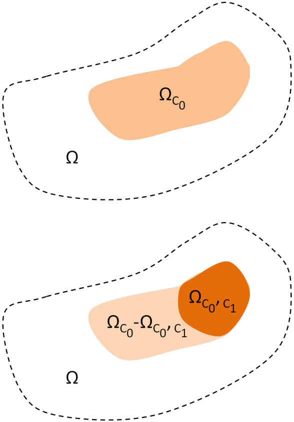


- We observe C_0 when predicting Y_0.



- We do not observe C_1.

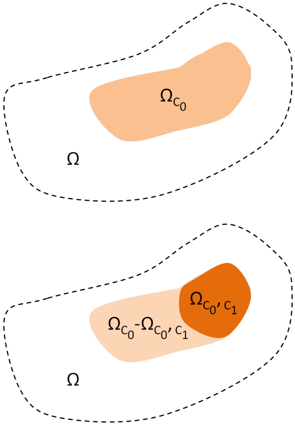
2 × Rule



- We observe C_0 when predicting Y_0.

- We do not observe C_1.
- Missing information in C_1 induces bias if individuals in $\Omega_{C_0} - \Omega_{C_0,C_1}$ are different from those in Ω_{C_0,C_1}.

2 × Rule



- We observe C_0 when predicting Y_0.
- We do not observe C_1.
- Missing information in C_1 induces bias if individuals in $\Omega_{C_0} - \Omega_{C_0,C_1}$ are different from those in Ω_{C_0,C_1}.

- Discriminatory power in the observed data must be twice as substantial as discriminatory power in the missing data.

There Is Actually Math Involved...

- $\hat{Y}_t \equiv \mathrm{E}(Y_t|C_t)$, $R_t \equiv Y_t - \mathrm{E}(Y_t|C_t)$, $\sigma_t^2 \equiv V[\hat{Y}_t|C_{\text{com}}]$,

There Is Actually Math Involved...

- $\hat{Y}_t \equiv \mathrm{E}(Y_t | C_t)$, $R_t \equiv Y_t - \mathrm{E}(Y_t | C_t)$, $\sigma_t^2 \equiv V[\hat{Y}_t | C_{\mathrm{com}}]$,

$$\beta_{t|1-t}^{\mathrm{obs}} \equiv \frac{\mathrm{Cov}\left(\hat{Y}_t, \hat{Y}_{1-t} \middle| C_{\mathrm{com}}\right)}{V\left(\hat{Y}_{1-t} \middle| C_{\mathrm{com}}\right)} \qquad \beta_{t|1-t}^{\mathrm{mis}} \equiv \frac{\mathrm{Cov}\left(R_t, \hat{Y}_{1-t} \middle| C_{\mathrm{com}}\right)}{V\left(\hat{Y}_{1-t} \middle| C_{\mathrm{com}}\right)}$$

There Is Actually Math Involved...

- $\hat{Y}_t \equiv \mathrm{E}(Y_t|C_t)$, $R_t \equiv Y_t - \mathrm{E}(Y_t|C_t)$, $\sigma_t^2 \equiv V[\hat{Y}_t|C_{\text{com}}]$,

$$\beta^{\text{obs}}_{t|1-t} \equiv \frac{\mathrm{Cov}\left(\hat{Y}_t, \hat{Y}_{1-t} \middle| C_{\text{com}}\right)}{V\left(\hat{Y}_{1-t} \middle| C_{\text{com}}\right)} \qquad \beta^{\text{mis}}_{t|1-t} \equiv \frac{\mathrm{Cov}\left(R_t, \hat{Y}_{1-t} \middle| C_{\text{com}}\right)}{V\left(\hat{Y}_{1-t} \middle| C_{\text{com}}\right)}$$

The $2\times$ Rule

- $\mathrm{Ave}_{\tilde{C}_1}(Y_1) - \mathrm{Ave}_{\tilde{C}_0}(Y_0)$ beats $\mathrm{Ave}_{C_{\text{com}}}(Y_1) - \mathrm{Ave}_{C_{\text{com}}}(Y_0)$ in MSE if and only if

$$2[w\beta^{\text{mis}}_{1|0} + (1-w)\,\beta^{\text{mis}}_{0|1}] \le \left[w\left(1-\beta^{\text{obs}}_{1|0}\right) + (1-w)\left(1-\beta^{\text{obs}}_{0|1}\right)\right]$$

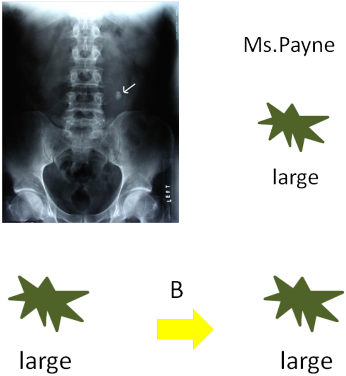
where $w = \sigma_0^2 / \left(\sigma_0^2 + \sigma_1^2\right)$.

Some Personalized Information

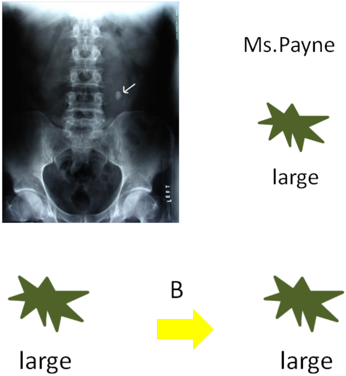
Ms.Payne

large

Some Personalized Information



Some Personalized Information



Problem

- The clinical trial data do not allow us to calculate the average of $Y_B - Y_A$ conditional on size $B =$ large.

What Comparison Should We Make?

Treatment A	Treatment B
78% (273/350)	**83%** (289/350)

	Treatment A	Treatment B
Small Stone	**93%** (81/87)	87% (234/270)
Large Stone	**73%** (192/263)	69% (55/80)

Checking Our Yields

Treatment A	Treatment B
78% (273/350) ⟷	83% (289/350)

⟷ Resolution 0

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	83% (289/350)

↔ Resolution 0

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	83% (289/350)

↔ Resolution 0

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

↔ Resolution 1

- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$

Keli Liu and Xiao-Li Meng

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	83% (289/350)

Resolution 0

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

Resolution 1

- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$
- **Resolution 1**: $0.69 - 0.73 = -0.04 \implies A$

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	83% (289/350)

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)

Resolution 0

Resolution 0.5

Resolution 1

- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$
- **Resolution 0.5**: $0.69 - 0.73 = -0.04 \implies A$

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	**83%** (289/350)

	Treatment A	Treatment B
Small Stone	**93%** (81/87)	87% (234/270)
Large Stone	**73%** (192/263)	69% (55/80)

Resolution 0

Resolution 0.5

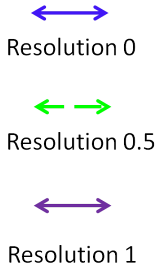
Resolution 1

- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$
- **Resolution 0.5**: $0.69 - 0.73 = -0.04 \implies A$
- **Resolution 1**: $0.69 - 0.78 = -0.09 \implies A$

Checking Our Yields

Treatment A	Treatment B
78% (273/350)	83% (289/350)

	Treatment A	Treatment B
Small Stone	93% (81/87)	87% (234/270)
Large Stone	73% (192/263)	69% (55/80)



- **Resolution 0**: $0.83 - 0.78 = 0.05 \implies B$
- **Resolution 0.5**: $0.69 - 0.73 = -0.04 \implies A$
- **Resolution 1**: $0.69 - 0.78 = -0.09 \implies A$
- **Plausible Range**: $[-0.31, 0.43]$

A Multi-Resolution View of Big Data

Population Resolution

A Multi-Resolution View of Big Data

Population Resolution

Individual Resolution