

Amenazas a la validez interna de los experimentos aleatorizados

Cómo (intentar) evitar que tu experimento se convierta en un desastre

Danilo Freire

danilo.freire@emory.edu

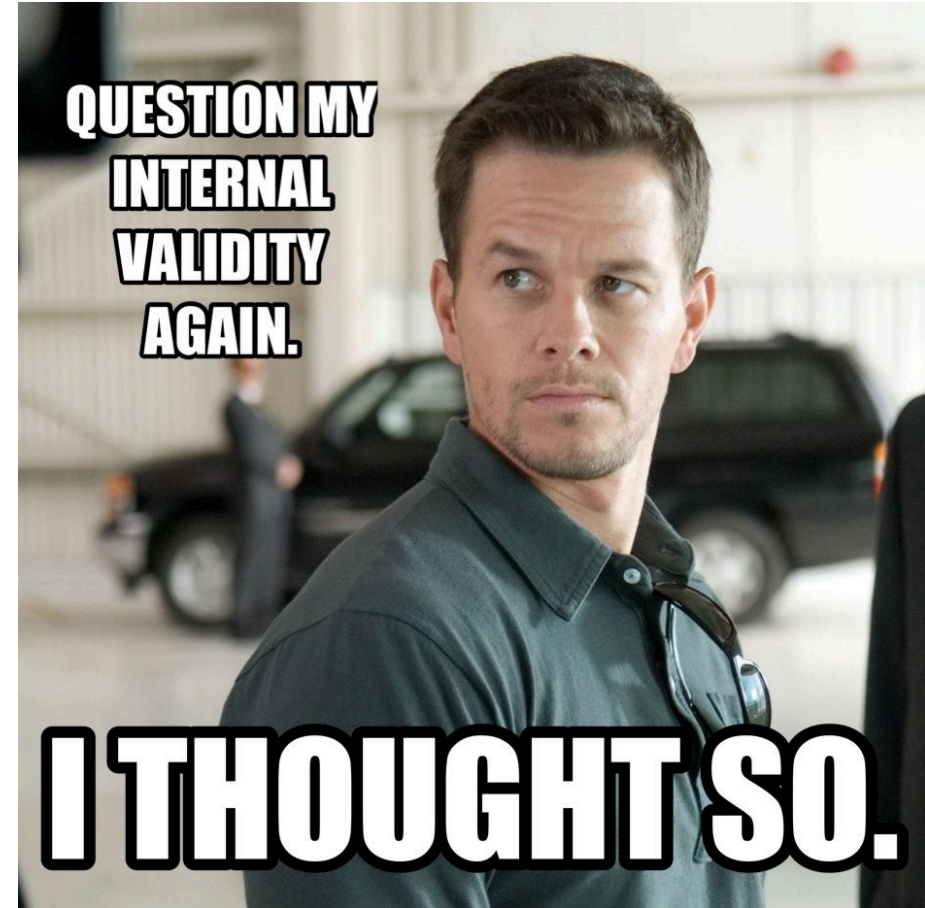
Emory University

Revisión de los supuestos principales



Puntos clave de esta presentación

- Deserción (variables de resultado con datos faltantes)
- Incumplimiento
- Derrame
- Efecto Hawthorne
- Tratamiento diferencial de los distintos brazos de tratamiento



Deserción

Deserción (variables de resultado con datos faltantes)

Otro problema de datos faltantes

- **Deserción**: Pérdida de participantes antes de la finalización del estudio
- Dos tipos principales:
 - **Deserción aleatoria**: La ausencia de datos no está relacionada con el tratamiento/resultado
 - **Deserción no aleatoria/sistemática**: La ausencia de datos se correlaciona con otras variables
- Impactos:
 - Reducción del **poder estadístico** 📉
 - Potencial **sesgo de selección** que amenaza la validez
 - **Generalización** comprometida 🌐



**Non-random
attrition**

**Random
attrition**

Deserción en experimentos

Datos de resultado faltantes

- Cuando la **deserción** se correlaciona con el tratamiento, **puede sesgar las estimaciones**
- Varios factores pueden conducir a la **deserción**:
 - **Incumplimiento**: La intervención es demasiado engorrosa, injusta o ineficaz
 - **Fatiga de encuesta**: Los participantes pierden interés con el tiempo
 - **Eventos imprevistos**: Problemas de salud, cambios de trabajo, etc.
 - **Errores en la recolección de datos**: Intencionales o no
- La **deserción obliga a los académicos a hacer suposiciones** sobre los datos faltantes
- La principal es que **la ausencia de datos es ignorable**, es decir, no está relacionada con el resultado
- Otro enfoque es suponer que **la ausencia de datos está relacionada con variables observadas**
- En este caso, se pueden utilizar **algunos métodos estadísticos** para modelar el mecanismo de ausencia de datos
- Finalmente, podemos **intentar recolectar más datos** para reducir la **deserción**, pero esto no siempre es factible

Casos especiales de deserción

Casos especiales de deserción

Datos faltantes independientes de los resultados potenciales

- ¡La **deserción aleatoria** es el mejor de los casos!
- Esta afirmación generalmente **no se verifica directamente**, sino que se asume
- Formalmente,
$$Y_i(z) \perp R_i(z)$$
- En algunos casos, esto puede ser cierto, como un mal funcionamiento de la computadora que causa la pérdida de datos de algunos participantes
- ¡De hecho, podemos probar eso!
 - Simplemente haga una regresión de r_i sobre las covariables y la asignación del tratamiento y no debería encontrar una relación significativa
- Si la **deserción** es aleatoria, podemos **ignorarla** y continuar con nuestro análisis
- Sin embargo, todavía tiene problemas:
 - **Poder reducido** debido a un tamaño de muestra más pequeño
 - Problemas de **generalización**
- Pero al menos podemos **estimar el ATE** sin sesgo 🤔
- **¿Cuáles son otros ejemplos de deserción aleatoria?** ¿Se le ocurre alguno?

Casos especiales de deserción

Datos faltantes relacionados con variables observadas

- Dado que la **deserción aleatoria** es rara, a menudo tenemos que lidiar con **deserción no aleatoria**
- Un tipo especial es la **ausencia independiente de los resultados potenciales dado X** , o **$MIPO|X$** (missing independent of potential outcomes given X)
 - $Y_i(z) \perp R_i(z)|X_i$
- Puede incluir más covariables en X_i si cree que están relacionadas con la **deserción**
- Imaginamos que hay una intervención que mejora los resultados de los estudiantes, pero algunos estudiantes no asisten a la escuela el día de la evaluación
- Supongamos que estudiantes con poca asistencia tienen más probabilidades de ausentarse el día de la evaluación y más probabilidades de beneficiarse de una intervención
- **$MIPO|X$** significa que si uno **divide la muestra experimental por asistencia previa**, la **ausencia es aleatoria dentro de cada subgrupo**
- Entre los estudiantes cuyo historial de asistencia es deficiente, no existe una relación entre la ausencia a la escuela el día de la evaluación y los resultados potenciales de los sujetos
- Lo mismo ocurre con los estudiantes con un buen historial de asistencia

Ponderación por probabilidad inversa

Una forma de manejar los datos faltantes relacionados con variables observadas

- Podemos compensar los datos faltantes **ponderando los datos observados**
- La idea es **sobreponderar las observaciones que son similares a las faltantes**
- Permítanme darles un ejemplo sencillo:
 - Supongamos que tenemos 40 personas en nuestro experimento
 - 30 son hombres y 10 son mujeres
 - El promedio general sería $\frac{30}{40} \times$ promedio de los hombres + $\frac{10}{40} \times$ promedio de las mujeres
- Ahora supongamos que 15 de los 30 hombres abandonan
- “Controlando por género” (es decir, ponderando):
 - Los 15 hombres restantes producen una estimación insesgada del promedio entre los 30 hombres
 - ¡Así que podemos **contarlos dos veces**!
 - Esta es la esencia de la **ponderación por probabilidad inversa** (IPW - Inverse Probability Weighting)

Ponderación por probabilidad inversa

La fórmula

- IPW estima el ATE cuando se cumple $MIPO|X$
- Para estimar el ATE, necesitamos $E[Y_i(1)]$ y $E[Y_i(0)]$ (¡como ya saben!)
- Cuando se cumple $MIPO|X$, $E[Y_i(1)]$ es un promedio ponderado:

$$E[Y_i(1)] = \frac{1}{N} \sum_{i=1}^N \frac{Y_i(1)r_i(1)}{\pi_i(z = 1, x)},$$

- $\pi_i(z = 1, x)$: proporción de sujetos no ausentes entre los tratados con perfil de covariable x
- Los resultados faltantes no tienen efecto en la suma
- Resultados informados ponderados por $1/\pi_i(z = 1, x)$
- Las ponderaciones reemplazan los valores faltantes con copias de los valores no faltantes
- El esquema de ponderación se llama **ponderación por probabilidad inversa** porque **las observaciones se ponderan por la probabilidad inversa de ser observadas**

Desventajas de IPW

Cuando $MIP O|X$ es incorrecto

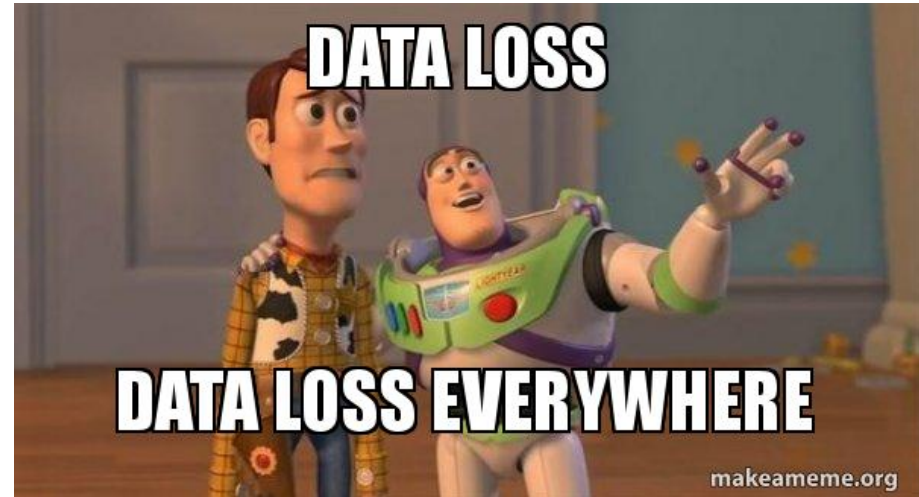
- El supuesto incorrecto de $MIP O|X$ conduce a estimaciones engañosas
- IPW asigna las mayores ponderaciones a los subgrupos con alta deserción
- La estimación sesgada de un subgrupo puede influir desproporcionadamente en el ATE general
- El promedio de IPW puede estar más sesgado que los datos no ponderados
- La reponderación aumenta la variabilidad del muestreo
- En la práctica, los investigadores deben hacer suposiciones sobre la deserción
- Evaluar $MIP O|X$: trabajo de detective y especulación 🕵️
- Más información sobre IPW en [este conjunto de diapositivas](#)

¿Qué podemos hacer? 🤔🧐

¿Qué podemos hacer?

Un checklist

- Compruebe si las **tasas de deserción son similares** en los grupos de tratamiento y control.
- Compruebe si las covariables de los grupos de tratamiento y control tienen **padrones similares**.
- **No elimine las observaciones con datos faltantes** en las variables de resultado.
- A veces, cuando faltan datos en la variable de resultados, **podemos acotar** ("bound") **nuestras estimaciones** del efecto del tratamiento.



Análisis de cotas extremas

Extreme Bounds Analysis (EBA)

- Imagina los resultados faltantes de aquellos que abandonaron
- No sabemos cuáles son esos resultados, pero podemos considerar dos escenarios extremos:
- Mejor escenario posible: Asumir que los desertores en el grupo de tratamiento habrían tenido los mejores resultados posibles, y los desertores en el grupo de control habrían tenido los peores resultados posibles (en relación con los datos observados)
- Peor escenario posible: Asumir que los desertores en el grupo de tratamiento habrían tenido los peores resultados posibles, y los desertores en el grupo de control habrían tenido los mejores resultados posibles
- Al calcular el ATE bajo estos supuestos extremos, obtenemos un rango (cota superior y cota inferior) que (con suerte 😊) contiene el ATE verdadero

Un ejemplo sencillo

- Supongamos que nuestro programa tiene los siguientes resultados si no hay deserción:
 - Promedio para el grupo de tratamiento: $\frac{(7+10+6+6)}{4} = \frac{29}{4} = 7.25$
 - Promedio para el grupo de control: $\frac{(3+7+5+6)}{4} = \frac{21}{4} = 5.25$
 - ATE: $7.25 - 5.25 = 2$
- Ahora, supongamos que solo observamos los siguientes resultados:
 - Promedio para el grupo de tratamiento: $\frac{(7+10+?+?)}{4} = ?$
 - Promedio para el grupo de control: $\frac{(?+7+5+6)}{4} = ?$
- Asumamos que los resultados recopilados muestran que los valores posibles para los datos faltantes están entre 0 y 10
- ¡Veamos cómo calcular las cotas para el ATE!

Un ejemplo sencillo

- Para encontrar la **cota superior** de la estimación del efecto del tratamiento, **sustituya 10 por los valores faltantes en el grupo de tratamiento y 0 por el valor faltante en el grupo de control**
- Cota superior: $\frac{(7+10+10+10)}{4} - \frac{(0+7+5+6)}{4} = \frac{37}{4} - \frac{18}{4} = 4.75$
- Para encontrar la cota inferior de la estimación del efecto del tratamiento, **sustituya 0 por los valores faltantes en el grupo de tratamiento y 10 por el valor faltante en el grupo de control**
- Cota inferior: $\frac{(7+10+0+0)}{4} - \frac{(10+7+5+6)}{4} = \frac{17}{4} - \frac{28}{4} = -2.75$
- Entonces, el **ATE está entre -2.75 y 4.75...** ¡y de hecho lo es (2)! 🎉
- Pero como puedes ver, **las cotas son bastante amplias...**
- [Lee \(2009\)](#) sugirió que podemos **recortar las cotas** para hacerlas más informativas (y más estrechas)
- Más información sobre el análisis de cotas extremas en [este conjunto de diapositivas](#)

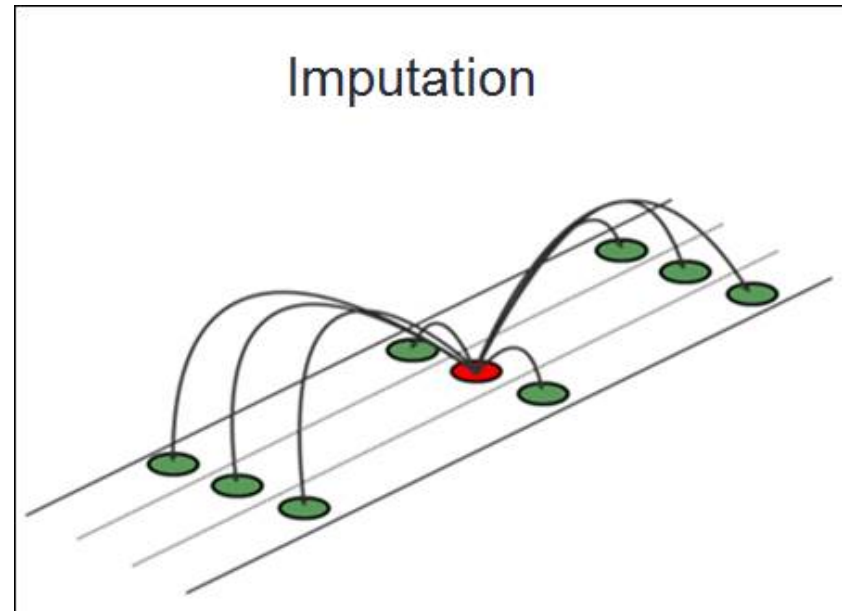
¿Qué podemos hacer?

- Pero la mejor solución es tratar de anticipar y prevenir la deserción.
- Personas que **no sepan a qué condición** del tratamiento fueron asignadas.
- Prometerle al grupo de control que **va a recibir el tratamiento** na vez concluida la investigación.
- **Planificar ex ante** para poder contactar a los sujetos en para la recolección final de datos.
- Contar con presupuesto para haver **seguimiento intensivo** a una muestra aleatoria de los sujetos que desierten



Los datos faltantes en covariables de contexto no son tan problemáticos

- La ausencia de datos en **covariables de contexto** (es decir, variables cuyos valores no cambian como resultado del tratamiento) para algunas observaciones **es menos problemática**.
- Podemos aprender sobre el efecto causal de un experimento incluso sin esas covariables
- Podemos usar la covariable de contexto según lo planeado **imputando los datos faltantes**
- También podemos condicionar nuestro análisis directamente a esos datos faltantes



Incumplimiento

Incumplimiento

Incumplimiento

- En algunas ocasiones **las unidades que son asignadas al tratamiento terminan no recibéndolo**. Es decir, no cumplen la asignación.
- Si todas las unidades asignadas al grupo de control no toman el tratamiento, pero sólo algunas unidades asignadas al tratamiento sí lo toman, estamos frente a un caso de incumplimiento unilateral.
- El efecto de ser asignado al tratamiento no es el mismo que el de recibir el tratamiento.
- El efecto de recibir el tratamiento es comunmente conocido como “efecto local promedio del tratamiento” (**LATE**) o efecto promedio del tratamiento para el cumplidor (**CACE**).
- La palabra “local” hace referencia a la idea de que el efecto solo aplica para las personas que toman el tratamiento cuando son asignadas a éste (el tipo de personas).

Un ejemplo motivador

Campaña para aumentar la participación electoral

- Imagina que estás interesado en estudiar el efecto de las campañas electorales en la participación de los votantes
 - ¡Quizás si tocas las puertas de las personas y les hablas sobre la importancia de votar, serán más propensas a hacerlo!
- Diseñas un experimento en el que asignas aleatoriamente a 1000 personas para que reciban la campaña (grupo de tratamiento) y a 1000 para que no la reciban (grupo de control)
- Sin embargo... usualmente solo el 25% de las personas en el grupo de tratamiento reciben realmente la campaña
 - El resto no está en casa, se niega a hablar, etc.
- Así que tenemos 250 personas tratadas y 1000 en el grupo de control
- ¿Qué harías? 🤔

Algunas opciones

- La primera opción que tenemos es **simplemente comparar los dos grupos** como si nada hubiera pasado
- El problema con este enfoque es que podría haber un **sesgo de selección** en el grupo de tratamiento
- La segunda opción, relacionada con la primera, es **asumir que las diferencias entre los dos grupos son aleatorias**
- El problema con este enfoque es que **no podemos probar esa suposición**
- No podemos saber si las diferencias entre los dos grupos son aleatorias o no
- **¡Lo más probable es que no lo sean!**
- La tercera opción es **mantener la asignación aleatoria** y comparar los dos grupos como si todos hubieran seguido la asignación del tratamiento
- En lugar de comparar a las personas que realmente recibieron la campaña con las que no la recibieron, **comparamos a las personas que fueron asignadas a recibir la campaña con las que no fueron asignadas a recibirla**
- La diferencia aquí es **semántica**:
 - Podríamos recuperar el ATE verdadero si solo le hubiéramos dado a la gente **la opción** de recibir o no la campaña
- La cuarta opción es usar **variables instrumentales (VI)** (o mínimos cuadrados en dos etapas - 2SLS)
- El beneficio de las VI es que nos permite **recuperar el efecto real del programa** en lugar de solo el efecto de que se ofrezca el programa
- La desventaja es que **las VI no nos permiten recuperar el ATE verdadero** en toda la población, solo en la subpoblación de cumplidores

CACE y VI 🧐

Efecto causal promedio del cumplidor (CACE)

El efecto del tratamiento en los cumplidores

- Como no podemos estimar correctamente el ATE con incumplimiento, nos centramos en el **efecto causal promedio del cumplidor (CACE)**
- El CACE intenta responder a esta pregunta: “**Para aquellos individuos que realmente escucharon el mensaje, ¿cuál es el efecto del mensaje en su probabilidad de votar?**”
- Sea la asignación experimental del sujeto i , z_i .
- Cuando $z_i = 1$, el sujeto es asignado al grupo de tratamiento, y cuando $z_i = 0$, el sujeto es asignado al grupo de control.
- Sea $d_i(z)$ una representación de si el sujeto i es realmente tratado, dada la asignación z_i .
- Para abreviar, escribamos $d_i(z = 1)$ como $d_i(1)$ y $d_i(z = 0)$ como $d_i(0)$.
- Para el incumplimiento unilateral, $d_i(0)$ es siempre 0 para todas las personas en los grupos de control, pero $d_i(1)$ puede ser 0 o 1.
- Formalmente, el CACE se define como:

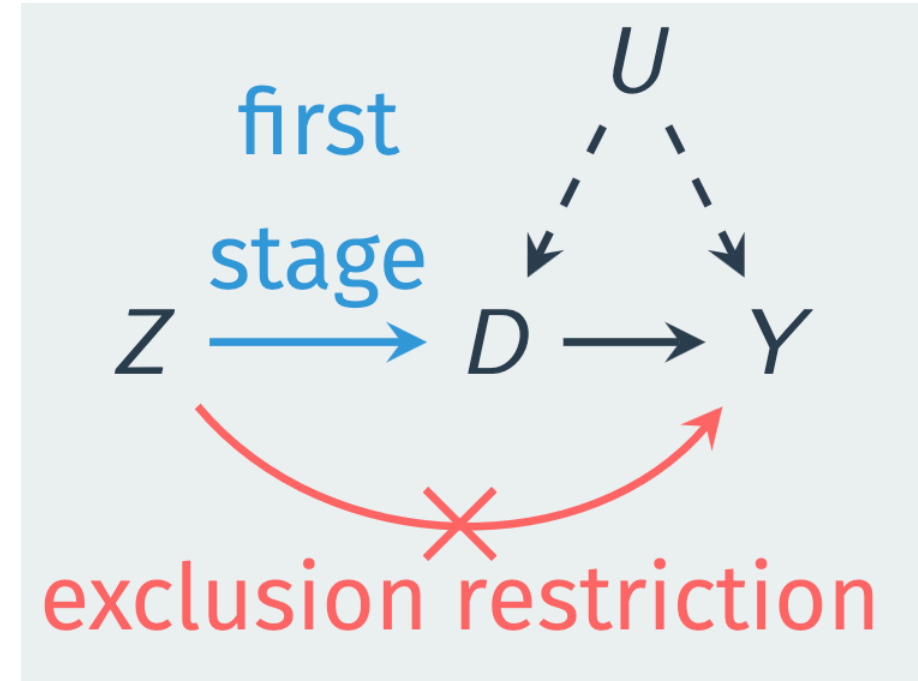
$$CACE \equiv \frac{\sum_{i=1}^N (Y_i(1) - Y_i(0))d_i(1)}{\sum_{i=1}^N d_i(1)} = E[(Y_i(d = 1) - Y_i(d = 0)) | d_i(1) = 1]$$

- En otras palabras, es **el efecto del tratamiento, pero solo proveniente de los cumplidores, dividido por el número de cumplidores**
- El CACE también se conoce como **Efectos Locales Promedio del Tratamiento (LATE)**. [Más en este link.](#)

CACE y variables instrumentales

El efecto del tratamiento en los cumplidores

- Lo bueno del CACE/LATE es que tenemos un estimador consistente para él
- Equivalente a los **estimadores de mínimos cuadrados en dos etapas**
 - Regresar D_i sobre Z_i para obtener los valores ajustados \hat{D}_i
 - Regresar Y_i sobre \hat{D}_i
- Pero recuerde los supuestos:
 - La **no interferencia** se violará, por ejemplo, si el sujeto j es contactado y se lo cuenta a i
 - La **excluibilidad** también puede violarse si los controles son tratados por otra campaña de contacto
 - La **primera etapa** requiere que al menos una persona sea tratada



Fuente: Blackwell (2021)

Estimando las regresiones

- Volvamos a nuestro ejemplo de campaña electoral
- Tenemos las siguientes regresiones para estimar el CACE:
 - Primera etapa: $TRATADO_i = \alpha_0 + \alpha_1 ASIGNADO + \epsilon_i$
 - Segunda etapa: $VOTÓ_i = \beta_0 + \beta_1 TRATADO + u_i$
- La primera etapa estima el efecto de la asignación del tratamiento sobre el estado del tratamiento
- La segunda etapa estima el efecto del estado del tratamiento sobre el resultado
- El coeficiente β_1 es el estimador de **mínimos cuadrados en dos etapas (MC2E)** del CACE
- Más información sobre el CACE y las variables instrumentales en [este link](#)

Derrame

Derrame e interferencia entre unidades

Derrame e interferencia entre unidades

- A veces tendremos sospechas de que la asignación al tratamiento de una unidad **afecta los resultados de otras unidades**.
- Si el tratamiento de una unidad afecta el resultado de otra unidad, estamos **violando uno de los supuestos básicos de inferencia causal**.
- Podemos seleccionar unidades que están **lejos unas de otras** para evitar la “transmisión” del tratamiento entre unidades.




¡No todos los derrames son malos!

- Como hemos visto, muchos fenómenos sociales son **interdependientes**
- Pero algunos están explícitamente **diseñados para aprovechar los derrames**
 - **Contaminación**: El efecto de vacunarse en la probabilidad de contraer una enfermedad depende de si otros están vacunados
 - **Efectos de red**: El valor de un producto o servicio aumenta a medida que más personas lo usan (redes sociales, telecomunicaciones)
 - **Vigilancia en puntos calientes**: El efecto de una mayor presencia policial en un área puede reducir el crimen en áreas cercanas
 - **Disuasión**: El efecto de un castigo más severo en un individuo puede disuadir a otros de cometer delitos

Hot spots policing in a high-crime environment: an experimental evaluation in Medellín



Daniela Collazos¹ · Eduardo García² · Daniel Mejía³ · Daniel Ortega⁴ · Santiago Tobón⁵ 

Published online: 13 February 2020
© Springer Nature B.V. 2020

Abstract

Objectives Test direct, spillover, and aggregate effects of hot spots policing on crime in a high-crime environment.

Methods We identified 967 hot spot street segments and randomly assigned 384 to a six-month increase in police patrols. To account for the complications resulting from a large experimental sample in a dense network of streets, we use randomization inference for hypothesis testing. We also use non-experimental streets to test for spillovers onto non-hot spots and examine aggregate effects citywide.

Results Our results show an improvement in short-term security perceptions and a reduction in car thefts, but no direct effects on other crimes or satisfaction with policing services. We see larger effects in the least secure places, especially for short-term security perceptions, car thefts, and assaults. **We find no evidence of crime displacement but rather a decrease in car thefts in nearby hot spots and a decrease in assaults in nearby non-hot spots.** We estimate that car thefts decreased citywide by about 11%.

Fuente: Collazos et al. (2021)

Estudiar los efectos secundarios

- Esto no es un problema si diseñas un estudio que permita investigar el derrame, es decir, que los resultados de una unidad pueden depender del tratamiento de otras unidades.
- Para investigar los efectos de derrame:

Un ejemplo

Campaña para aumentar la participación electoral

- Imagina que estamos realizando un experimento para probar el efecto de una campaña para salir a votar en la participación electoral
- Tenemos un **diseño multinivel**:
 - Los **hogares** se asignan aleatoriamente a tratamiento o control
 - Dentro de los hogares, los **individuos** también se asignan aleatoriamente
- Nos centramos en **hogares de dos votantes** donde los residentes comparten dirección
- Grupos de asignación aleatoria:
 - 5,000 hogares: ambos votantes objetivo
 - 5,000 hogares: ninguno objetivo
 - 10,000 hogares: uno aleatoriamente objetivo
- Crea cuatro grupos de 10,000 personas:
 - **Correo recibido por ambos**
 - **Ninguno recibe correo**
 - **Correo con compañero de casa no tratado**
 - **No tratado con compañero de casa tratado**
- Características del **diseño multinivel**:
 - **Aleatorización en dos etapas** (hogar y luego individuo)
 - Ampliable a **niveles adicionales como código postal**

Un ejemplo

Campaña para aumentar la participación electoral

- Los resultados potenciales dependen tanto del estado de tratamiento propio como del compañero de casa
- El sistema de notación revisado tiene en cuenta las influencias de tratamiento dual (Y_{ab} donde):
 - a = estado de tratamiento del compañero de casa (0=control, 1=tratado)
 - b = estado de tratamiento propio (0=control, 1=tratado)

Notación	Compañero de casa	Propio	Interpretación
Y_{00}	Control	Control	Resultado base
Y_{01}	Control	Tratado	Efecto directo del tratamiento
Y_{10}	Tratado	Control	Efecto de derrame
Y_{11}	Tratado	Tratado	Efectos combinados

- Supuesto clave: Contención estricta en el hogar
 - Resultados solo afectados por intervenciones dentro del hogar
 - Sin interferencia de asignaciones de tratamiento externas

Amenazas

Definiendo los efectos causales

- De los **cuatro resultados potenciales**, surgen estimandos clave:
 - $Y_{01} - Y_{00}$: **Efecto directo** del tratamiento cuando el compañero de casa no es tratado
 - $Y_{11} - Y_{10}$: **Efecto del tratamiento** cuando el compañero de casa recibe correo
 - $Y_{10} - Y_{00}$: **Efecto de derrame** en no tratados con compañero de casa tratado
 - $Y_{11} - Y_{01}$: **Interacción de derrame** entre pares tratados
- Consideraciones críticas:
 - Los efectos dependen del **estado de tratamiento del compañero de casa**
 - Los estimandos capturan **influencias combinadas** de:
 - ⇒ Impactos directos de la intervención
 - ⇒ Derrames por comunicación
 - ⇒ Efectos de recursos compartidos
 - No aísla **mecanismos** específicos detrás de los derrames

Desafíos de implementación

Consideraciones prácticas

- Obstáculos comunes de implementación:
 - Contaminación del tratamiento entre grupos
 - Deserción diferencial entre condiciones
 - Complejidades del monitoreo del cumplimiento
 - Asignación de recursos para seguimiento multinivel
- Consideraciones éticas:
 - Consentimiento informado para miembros de la red
 - Protección de la privacidad para datos del hogar
 - Implicaciones de equidad de los efectos de derrame
- Limitaciones de diseño a abordar:
 - Requisitos de tamaño de muestra para efectos de clúster
 - Desafíos de medición para impactos indirectos
 - Aspectos temporales del momento del derrame
 - Estimación correcta de errores estándar, ya que los sujetos probablemente están correlacionados de muchas maneras

Nota: El pre-registro es crucial para diseños complejos

Efecto Hawthorne 🧐📷

Efecto Hawthorne 🙇

- El efecto Hawthorne: **saber que están siendo observados/estudiados** puede llevar a los sujetos a comportarse de manera diferente.
- Esto podría crear un **error de medición sistemático** tanto en los grupos de tratamiento como en los de control.
- También podría resultar de una **mayor atención prestada al grupo de tratamiento**, deshaciendo la creación de grupos de tratamiento y control equivalentes creados a través de la asignación aleatoria



Buenas practicas

- Recolectar datos de la forma más discreta posible.
- En la medida de lo posible, nadie más que el propio sujeto debe conocer a qué condición de tratamiento fue asignado.
- Encuestadores/investigadores deberían desconocer la condición de tratamiento de los sujetos.
- No debemos hacer mediciones adicionales para el grupo de tratamiento.

No excluibilidad 🙅

Diferencias entre los grupos de tratamiento y control más allá del tratamiento

- Manejar los grupos de tratamiento y control de manera diferente significa que las diferencias observadas en las variables de resultado entre estos grupos pueden deberse al tratamiento y/o al manejo diferente.
- Los ejemplos incluyen el uso de diferentes cuestionarios para la recolección de datos o rondas adicionales de recolección de datos para el grupo de tratamiento con la intención de obtener información sobre los mecanismos.
- Recuerde diseñar su estudio y sus encuesta para tratar todos los brazos de tratamiento por igual, excepto por el tratamiento en sí.

