

# Seminario regional: introducción a la evaluación de impacto de políticas y programas públicos

**Lesbia Maris**

Dirección de Evaluación de Impacto y Aprendizaje de Políticas  
CAF – Banco de desarrollo de América Latina

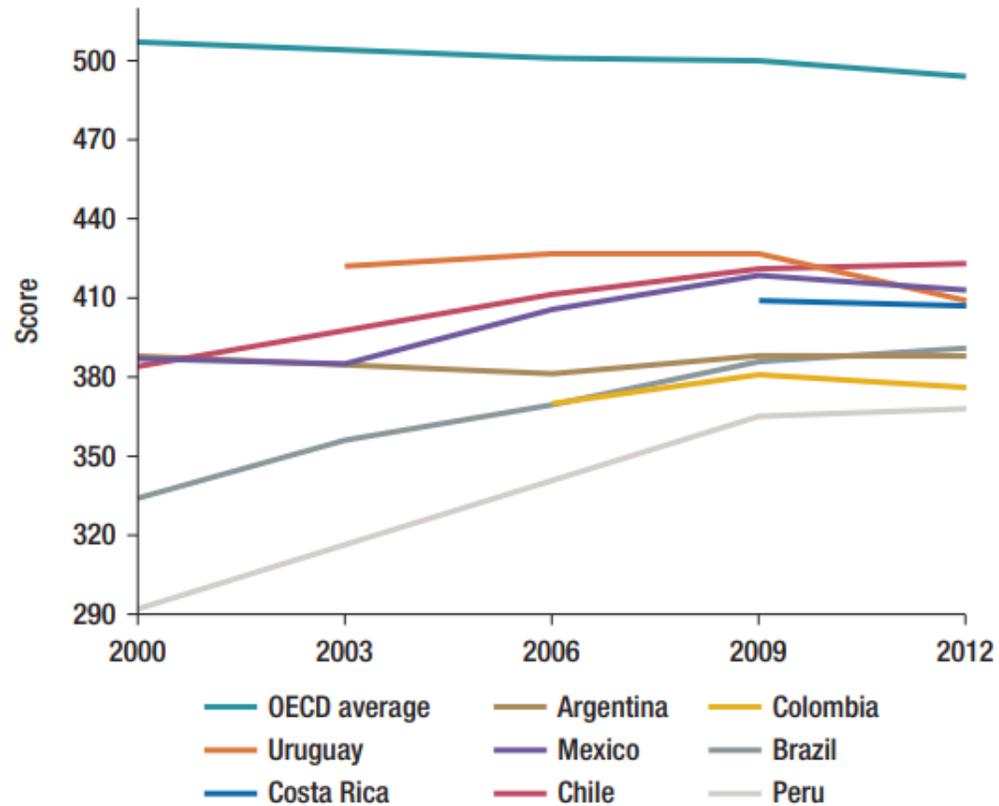
**Ciudad de México**

23, 24 y 25 de agosto de 2017



# ¿Por qué estamos aquí?

## Cambio en el desempeño en Matemática entre 2000 y 2012 (PISA)



Source: OECD 2013.

Note: OECD = Organisation for Economic Co-operation and Development; PISA = Program for International Student Assessment.



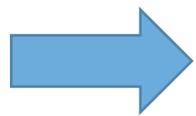
## ¿Cómo mejoramos la provisión de servicios públicos?

- Mejores políticas
- Mejor implementación
- Más recursos

Necesitamos saber qué funciona y qué no

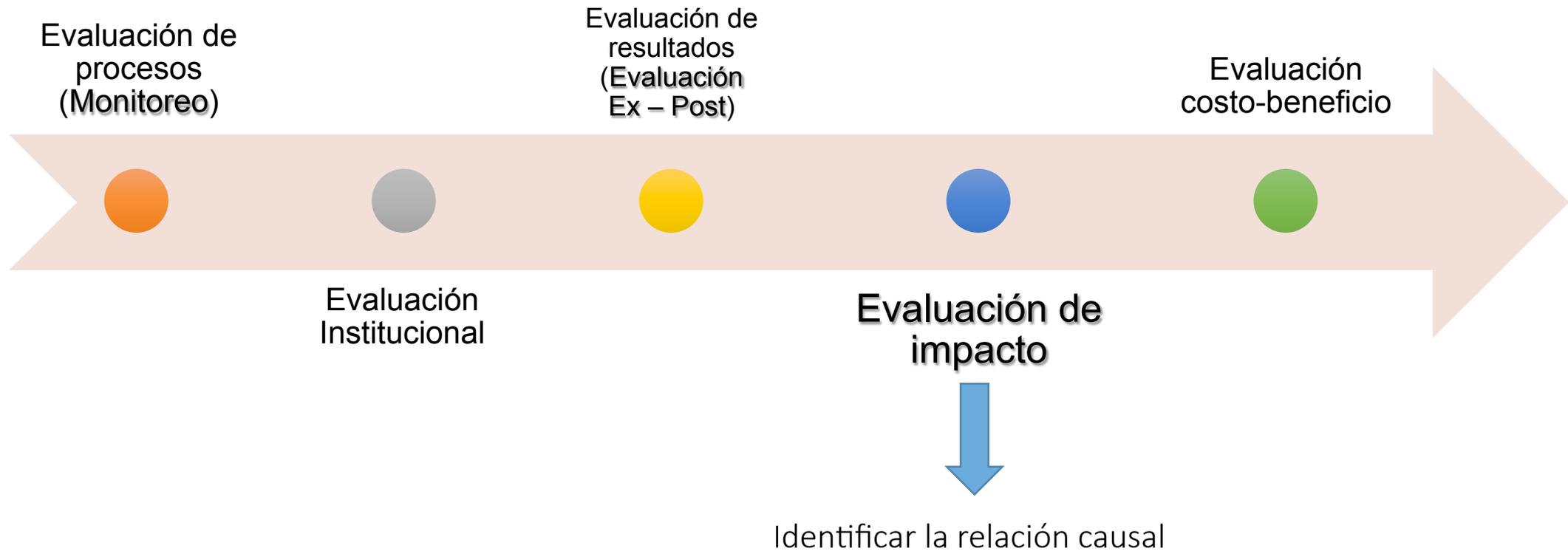
# ¿Sabemos cómo gestionar recursos públicos? ¿sabemos lo que funciona?

- En general, NO
- Por diversas razones, la abundancia de experiencia en la aplicación de iniciativas NO se suele traducir en conocimiento útil para la gestión



Oportunidades de aprendizaje no aprovechadas

# Tenemos algunas opciones para aprender...



# Ciclo de innovación y aprendizaje

1

*Identificación de oportunidades de mejora*

2

*Diagnóstico de causas y validación*

3

*Posibles soluciones*

5

*Incorporación de aprendizajes*

4

*Pilotaje de alternativas y medición de impacto*



# ¿Qué es evaluación de impacto?

- La evaluación de impacto busca **cuantificar el efecto** sobre los beneficiarios que se le puede **atribuir** a un programa.
- Para esto trata **establecer una relación causal** entre el programa y los resultados, utilizando **métodos rigurosos de diseño** y de estudio de la **evidencia empírica**



# La evaluación de impacto versus otras evaluaciones...

- Es importante hacer la diferencia entre **Evaluación de Impacto y Evaluación de Resultados**.
- Mientras que la Evaluación de Resultados **no distingue** entre la parte del resultado observado que se debe estrictamente al programa y la que se debe a diferencias previas entre beneficiarios y no-beneficiarios. La Evaluación del Impacto mide la **diferencia entre lo que pasó** habiendo implementado el programa y **lo que hubiera pasado** sin él.

# Un ejemplo: programa de entrenamiento para el trabajo

## ⇒ El objetivo del programa

Entrenar a desempleados o trabajadores informales en actividades prácticas para expandir sus oportunidades laborales, incrementar su ingreso y fortalecer sus valores cívicos. Las actividades son: ventas y bartending

## ⇒ El objetivo de la evaluación

Entender si el programa está cumpliendo con su objetivo: incrementar participación y empleabilidad

# Un ejemplo: programa de entrenamiento para el trabajo

⇒ ¿Cómo hallar el contra-factual?

- 1) Se construyó grupo de participantes elegibles a través de encuesta.
- 2) 40% de los elegibles fue asignado a un curso y el resto (60%) fue asignado a no participar. A los del grupo de control, se les ofreció participar en una segunda fase de registro.

*Esto asegura que el grupo que participa es en promedio parecido al grupo que no participa y que cualquier diferencia en su desempeño se debe SOLO al programa*

⇒ ¿Cómo se halló el impacto?

Nivel de empleo del grupo de participantes – nivel de empleo del grupo de no participantes

# ¿Cómo hacer una evaluación de impacto?

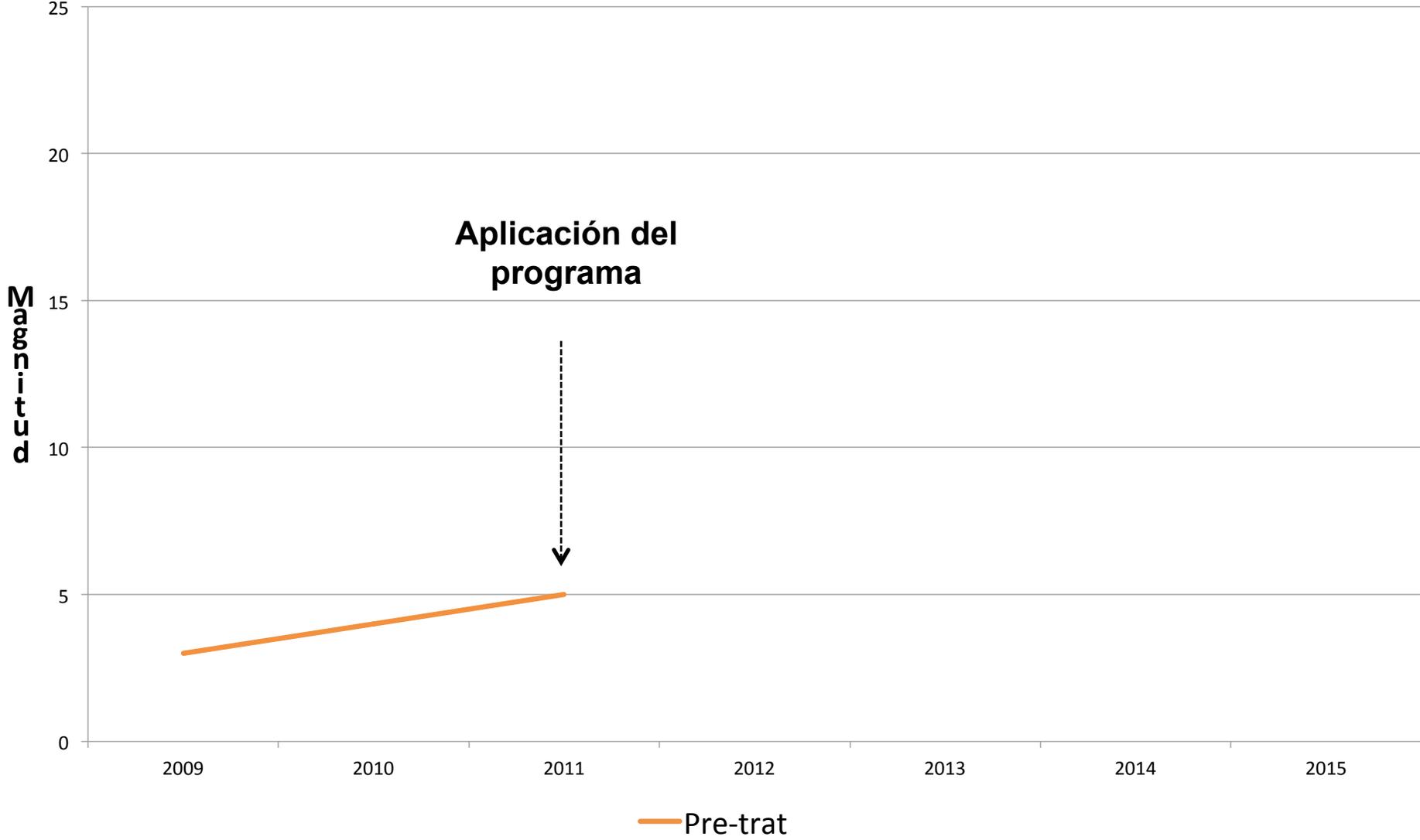
El reto fundamental es lograr encontrar el contra-factual siguiente: “¿cuál hubiese sido el resultado observado si el tratamiento/programa no se hubiese aplicado?”

Los **requisitos principales** para lograr esto son:

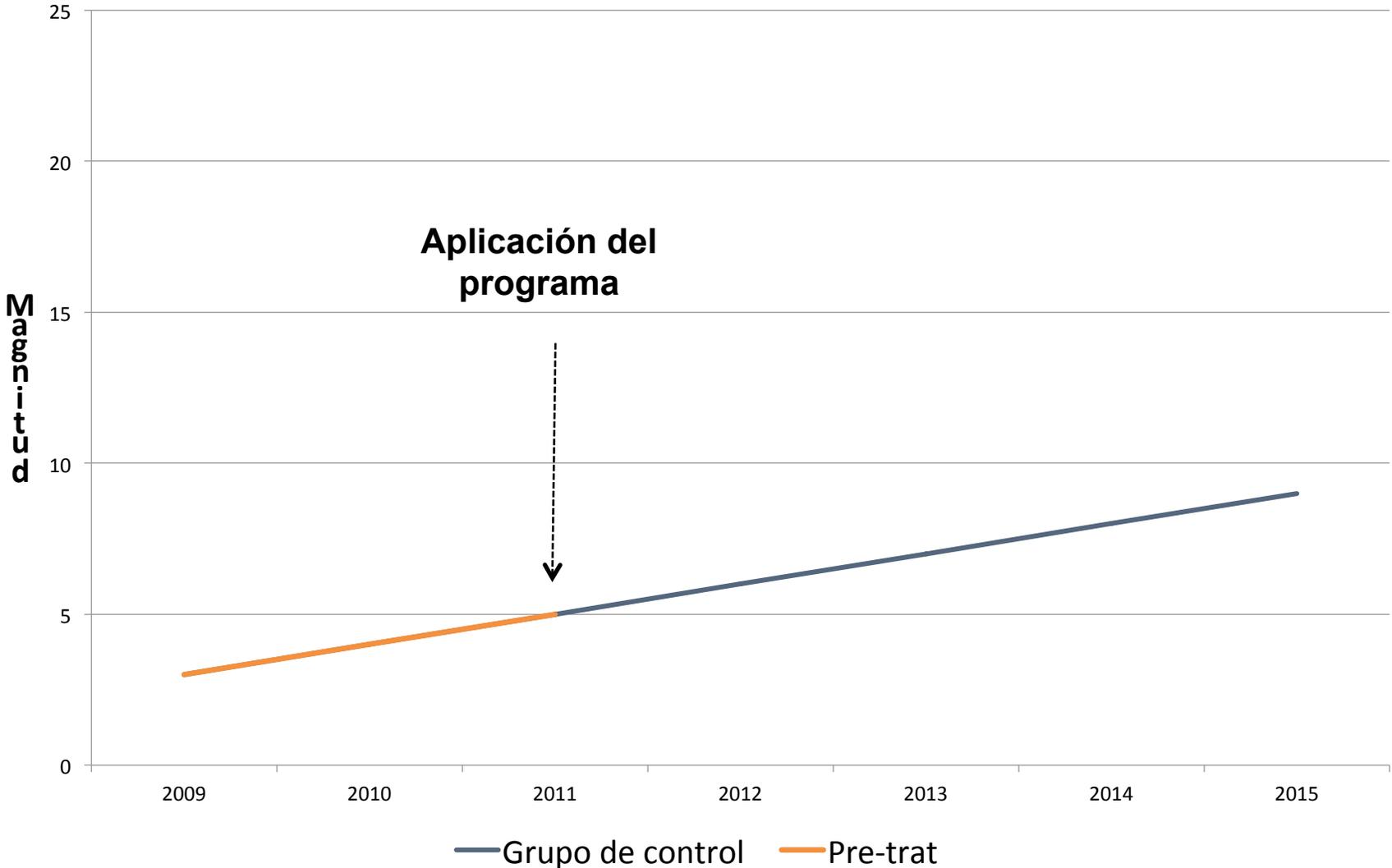
1. Construir un grupo de control adecuado
2. Medir las variables de interés tanto en este grupo de control como en el grupo de tratamiento, como mínimo, después del programa

Un grupo de control adecuado es **estadísticamente indistinguible** del grupo que recibe la política (grupo de tratamiento) antes de recibirla. La mejor manera de lograr esto es a través de una **selección aleatoria** de los que reciben el programa.

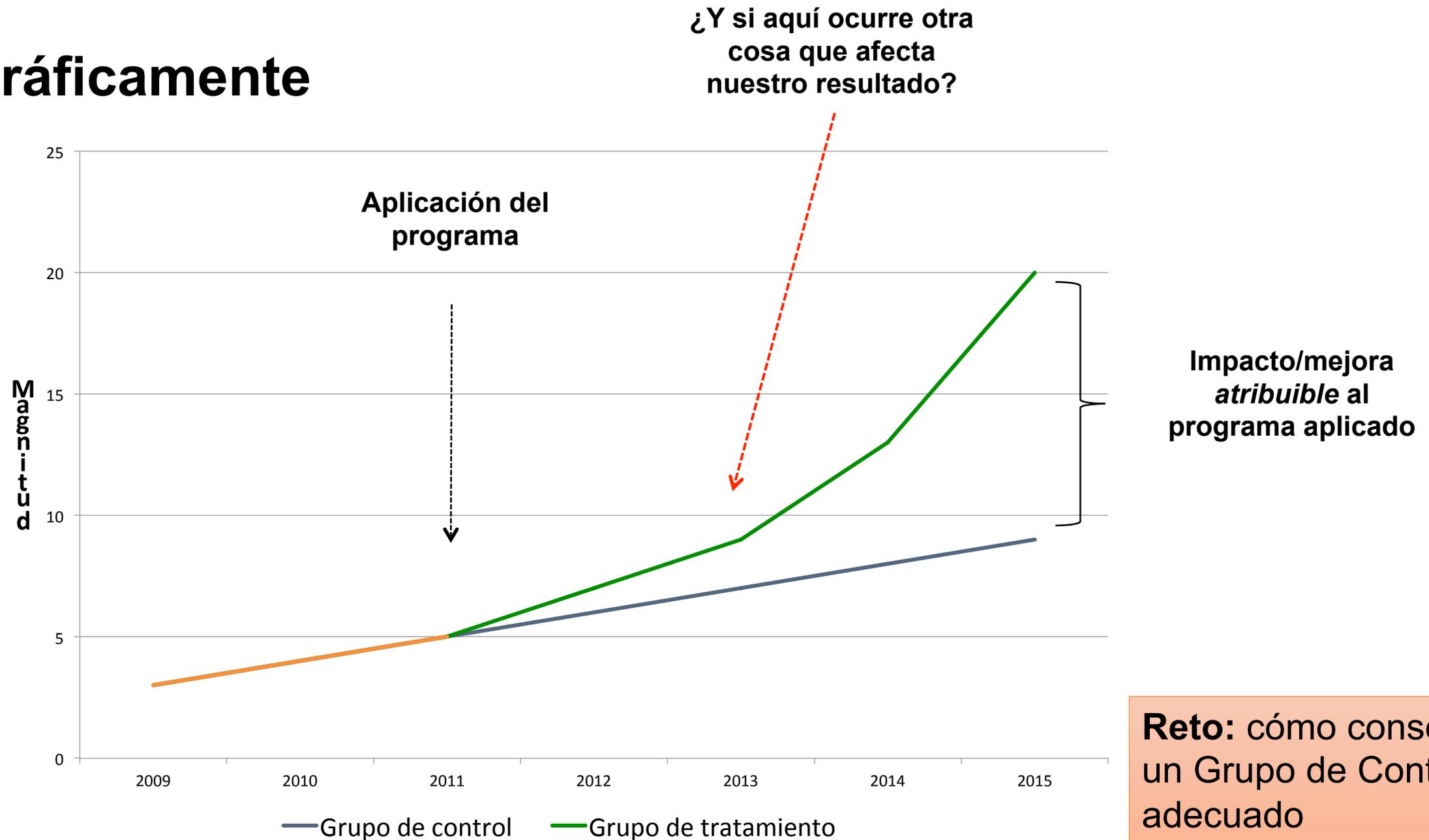
# Gráficamente



# Gráficamente



# Gráficamente



# El problema de inferencia causal

- Conocer el efecto de una intervención requiere saber ¿cuál hubiese sido el resultado si la intervención no hubiese ocurrido? (contra-factual)
- Imposible de responder a nivel individual:  
La persona recibe o no recibe la intervención, y el tiempo no se puede retroceder

# El problema de inferencia causal formalmente

- Sea  $Y_i$  el resultado (outcome) de interés para la unidad  $i$  (persona, empresa, etc)
- $Y_{0i}$  sería el resultado cuando  $X=0$  (no ha recibido tratamiento) y  $Y_{1i}$  cuando  $X=1$  (ha recibido el tratamiento)
- Para cada individuo podemos pensar en el par  $(Y_{0i}, Y_{1i})$  como posibles resultados observados dependiendo de si recibe el tratamiento-política ( $X_i=1$ ) o no ( $X_i=0$ )

# Efecto Causal

El efecto causal de X sobre Y es por lo tanto

$$T_i = Y_{1i} - Y_{0i}$$

Donde T es el “efecto tratamiento”

El problema es que nunca se observa  $Y_{1i} - Y_{0i}$

dado que no podemos observar a una persona en ambos estados

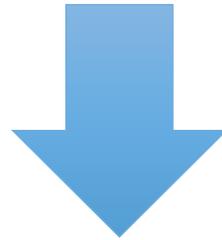
En cambio, se observa o  $Y_{1i}$  o  $Y_{0i}$

Este es el “problema fundamental de inferencia causal”: en vista de que no es posible observar ambos resultados para el individuo, no se puede observar el efecto causal de X sobre Y en un individuo

# Efecto Causal

Rara vez podemos observar a la misma persona  $i$  en ambos estados al mismo tiempo:

*observamos el factual, no observamos el contra-factual*



El problema de la evaluación de **impacto es un problema de ausencia de información**

# ¿A qué queremos que se dedique nuestro hijo?

Tenemos la opción de inscribir a nuestro hijo en una academia de beisbol o en una academia de música

Niños	$Y_{0i}$ Ingresos si se dedican a jugar beisbol	$Y_{1i}$ Ingresos si se dedican a ser músicos	Impacto (Efecto de ir a una escuela de música)
Niño 1	10	15	5
Niño 1	15	15	0
Niño 3	20	30	10
Niño 4	20	15	-5
Niño 5	10	20	10
Niño 6	15	10	-5
<b>Promedio</b>	<b>16</b>	<b>18</b>	<b>2</b>

# Entonces, si no observamos el contra-factual, ¿cómo hacemos?...

- Queremos medir el impacto de dedicarse a actividades musicales
- Tenemos dos grupos:
  - Un grupo de egresados de un conservatorio de entre 20 y 25 años, que vive en Ciudad de México, todos varones, provenientes de estratos humildes y empleados.
  - Un grupo de varones que seleccionamos al azar con las mismas características (de entre 20 y 25 años, que vive en Ciudad de México, todos varones, provenientes de estratos humildes y empleados).

¿Por qué simplemente no los comparamos?

# Solución estadística

- Qué datos tenemos:
  - El resultado (ingreso) para cada individuo del tratamiento y del control después del programa
- Podríamos comparar ambos grupos
  - Calculamos el ingreso promedio del grupo de tratamiento:  $E(Y_{1i} | X_i = 1)$
  - Calculamos el ingreso promedio del grupo de control:  $E(Y_{0i} | X_i = 0)$
  - Luego los restamos:

**Promedio grupo de tratamiento – promedio grupo de control**

“Naive” estimation

**Esto es incorrecto... ¿Por qué?**

# Solución estadística

Hay cosas que NO observamos:

- qué hubiese pasado con un individuo del tratamiento si hubiese estado en el control

$$E(Y_{0i} | X_i = 1) \neq E(Y_{0i} | X_i = 0)$$

- qué hubiese pasado con un individuo del control si hubiese estado en el tratamiento

$$E(Y_{1i} | X_i = 0) \neq E(Y_{1i} | X_i = 1)$$

**Observamos  
factuales  
NO observamos  
contra-factuales**

# Sesgo de selección

El grupo de participantes puede ser diferente al grupo de no participantes (incluso en ausencia del programa)



**Estamos comparando dos grupos diferentes**



$$E(Y_{0i} | X_i = 1) \neq E(Y_{0i} | X_i = 0)$$

$$E(Y_{1i} | X_i = 0) \neq E(Y_{1i} | X_i = 1)$$

# Sesgo de selección

## Músicos:

- Habilidad musical (que puede estar relacionada con habilidad matemática)
- Padres más educados
- Padres que exponen a sus hijos a expresiones de arte
- Padres interesados en sus hijos
- Niños motivados por la música y el arte

## Deportistas:

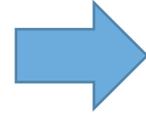
- Habilidad motora
- No necesariamente padres más educados
- Padres que exponen a sus hijos a actividades deportivas
- Padres interesados en sus hijos
- Niños motivados por el deporte y la actividad física

Si observamos que los músicos tienen mayores ingresos (en promedio)

¿Atribuimos este impacto a la academia de música o a algunas de estas diferencias? ¿o a una combinación?

# Atendiendo el sesgo de selección

El grupo de control “ideal” es aquel que es estadísticamente indistinguible del grupo que recibe la política, antes de recibirla



Para atender el sesgo de selección hay que atender la “selección”

# Sesgo de selección

Si nuestros grupos fueran parecidos desde el principio, en promedio, podríamos aproximarnos al contra-factual (que NO observamos) usando información que ***SÍ tenemos***:

Si antes del programa, sucede que el grupo de tratamiento es parecido al grupo de control, en promedio, entonces:

$$E(Y_{0i} | X_i = 1) = E(Y_{0i} | X_i = 0)$$

Porque, dado que son grupos parecidos, ocurriría que:

$$E(Y_{0i} | X_i = 1) = E(Y_{0i} | X_i = 0) = E(Y_{0i})$$

# Solución estadística

Aunque sigue habiendo cosas que NO observamos (ahora las podemos estimar):

- qué hubiese pasado con un individuo del tratamiento si hubiese estado en el control

$$E( Y_{0i} | X_i = 1 ) = E( Y_{0i} | X_i = 0 ) = E( Y_{0i} )$$

- qué hubiese pasado con un individuo del control si hubiese estado en el tratamiento

$$E( Y_{1i} | X_i = 0 ) \neq E( Y_{1i} | X_i = 1 ) = E( Y_{1i} )$$

# ¿Cómo construimos un buen contra-factual?

El secreto:

Lograr que la única diferencia posible entre dos grupos sea que unos recibieron un tratamiento y otros no (todo lo demás, debe ser igual)



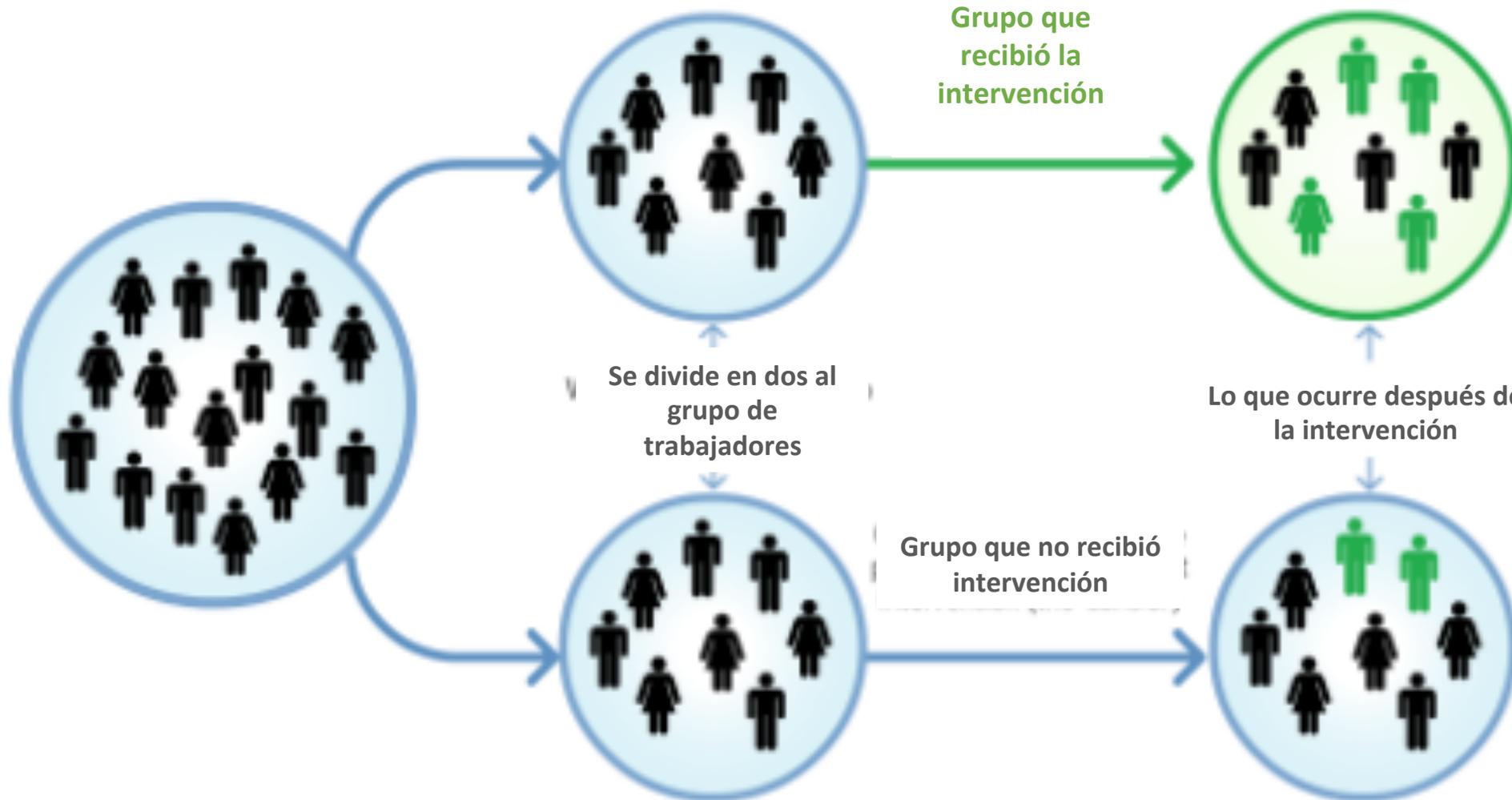
¿Cómo logramos esto?

<p><b>Experimentos</b></p>	<p><b>Cuasi-experimentos</b> <b>Experimentos naturales</b></p>	<p><b>Métodos no experimentales:</b></p> <ul style="list-style-type: none"><li>• Diferencias en diferencias<ul style="list-style-type: none"><li>• Emparejamiento</li></ul></li><li>• Regresión discontinua</li><li>• Variables instrumentales (IV)</li></ul>
----------------------------	--	---

# ¿Por qué experimentos?

- Se puede resolver el problema de la evaluación (sesgo de selección).
- La implementación puede ser simple: **fácil y poco costosa**.
- La lógica del experimento es relativamente simple y esto puede facilitar la difusión de resultados. Es más probable que sirva para convencer a alguien de continuar (o discontinuar) un programa.

# ¿Cómo la aleatorización nos ayuda a resolver el sesgo de selección?



**IMPORTANTE:**  
Cualquier individuo del grupo original tiene la misma probabilidad de ser seleccionado para participar en el programa

# Formalmente....

Simplemente tomamos la diferencia entre dos medias:

$$E(Y_{1i} | X = 1) = E(Y_{1i})$$

$$E(Y_{0i} | X = 0) = E(Y_{0i})$$

$$E(Y_{1i}) - E(Y_{0i}) = \tau$$

Impacto

$$\tau = \frac{\sum_{i=1}^{n_1} (Y_{1i} | X_i = 1)}{n_1} - \frac{\sum_{i=1}^{n_0} (Y_{0i} | X_i = 0)}{n_0}$$

Este es un cálculo  
mecánico

# Algunos ejemplos interesantes de experimentos

¿**Qué medio de comunicación es más efectivo** para aumentar el pago de obligaciones tributarias?

- Podemos enviar correos o visitar a los contribuyentes
- Tenemos una lista de 20.000 contribuyentes con su información de contacto (emails y direcciones) y deudas pendientes
- Contribuyentes variados en: ubicación, tamaño de deuda, tipo de deuda, antigüedad de deuda
- Construimos tres grupos (parecidos) de contribuyentes:
  - Unos que no reciben comunicación
  - Unos que reciben correos
  - Unos que reciben visitas

# Algunos ejemplos interesantes de experimentos

¿Cómo umentar la matriculación escolar en zonas rurales?

- Podemos permitir que las familias tomen estas decisiones o podemos incentivarlas mediante transferencias condicionadas.
- Tenemos una lista de casi 500 localidades que cumplen con condiciones de vulnerabilidad
- La localidades varían en varias dimensiones: tamaño, actividad económica, ingresos, etc.
- Construimos dos grupos parecidos de localidades:
  - Unos que no recibirán las transferencias
  - Unas que sí las recibirán

# Aleatorizando en la vida real

## ¿Cómo hacemos el sorteo?

Excel, sorteo público (explícitamente organizado o no), tirar una moneda, paquete estadístico

## Cuándo podemos aleatorizar:

- **Sobre-subscripción:** la capacidad del programa es menor a la demanda. Ideal cuando no se puede atender a toda la población interesada.
- **Entrada escalonada:** se tiene que atender a todos los interesados pero no al mismo tiempo, sino que el tratamiento va por grupos. Aleatoriza el orden en que cada grupo recibe el tratamiento. No permiten estimar efectos de largo plazo.
- **Aleatorización al interior de un grupo:** dentro de cada grupo atiendes a un subgrupo con características similares. El sub-grupo en cada caso se escoge aleatoriamente.
- **Diseño de inducción (o incentivo):** no se puede decidir aleatoriamente quién participa, pero sí quién recibe un aliento o incentivo para participar. El incentivo tiene que funcionar, sino, sería un instrumento débil.

# Para tener en cuenta...

- ¿Están equilibrados los grupos? (¿se parecen?)
- ¿Estamos aleatorizando grupos?
  - El número de personas dentro de cada grupo es menos importante que el número de grupos.
  - Hay una parte del error que es común al grupo (esto afecta la estimación).
- ¿Asumimos que todo el mundo efectivamente toma el tratamiento?: generalmente **NO**, entonces **necesitamos más gente**. ¿y en el grupo de control?
- ¿Asumimos que el saber que están siendo evaluados **NO** cambia los comportamientos?
- A veces hay que estratificar (aleatorizar a lo interno de grupos con ciertas características): esto aumenta la precisión de los estimados (disminuye el error estándar).

# Límites de los experimentos

- Solo nos permiten conocer el **impacto promedio**: queda mucha información importante que no podemos conocer a través de un experimento (esto es muy importante desde el punto de vista de la política pública).
- No nos permite conocer **qué hay detrás de los impactos**. Para esto necesitamos modelos estructurales (que requieren hacer muchas suposiciones).
- **Validez externa**: si replicáramos el programa en otro contexto, ¿obtendríamos los mismos resultados?
- Costos de la evaluación (explícitos e implícitos)
- ¿Nos sentimos cómodos seleccionando la participación por sorteo?
- **Timing**: disponibilidad de los resultados y ciclo de políticas públicas.

# Hay vida más allá del muro (la aleatorización)...

- Incluso si aleatorizamos, es posible que tengamos pocas unidades (entonces es probable que los grupos difieran).
- Si tenemos datos de línea de base, seguramente querríamos utilizarlos
- Si la selección no es aleatoria, ¿qué opciones tenemos?

# Diferencias en diferencias

Se hacen dos comparaciones:

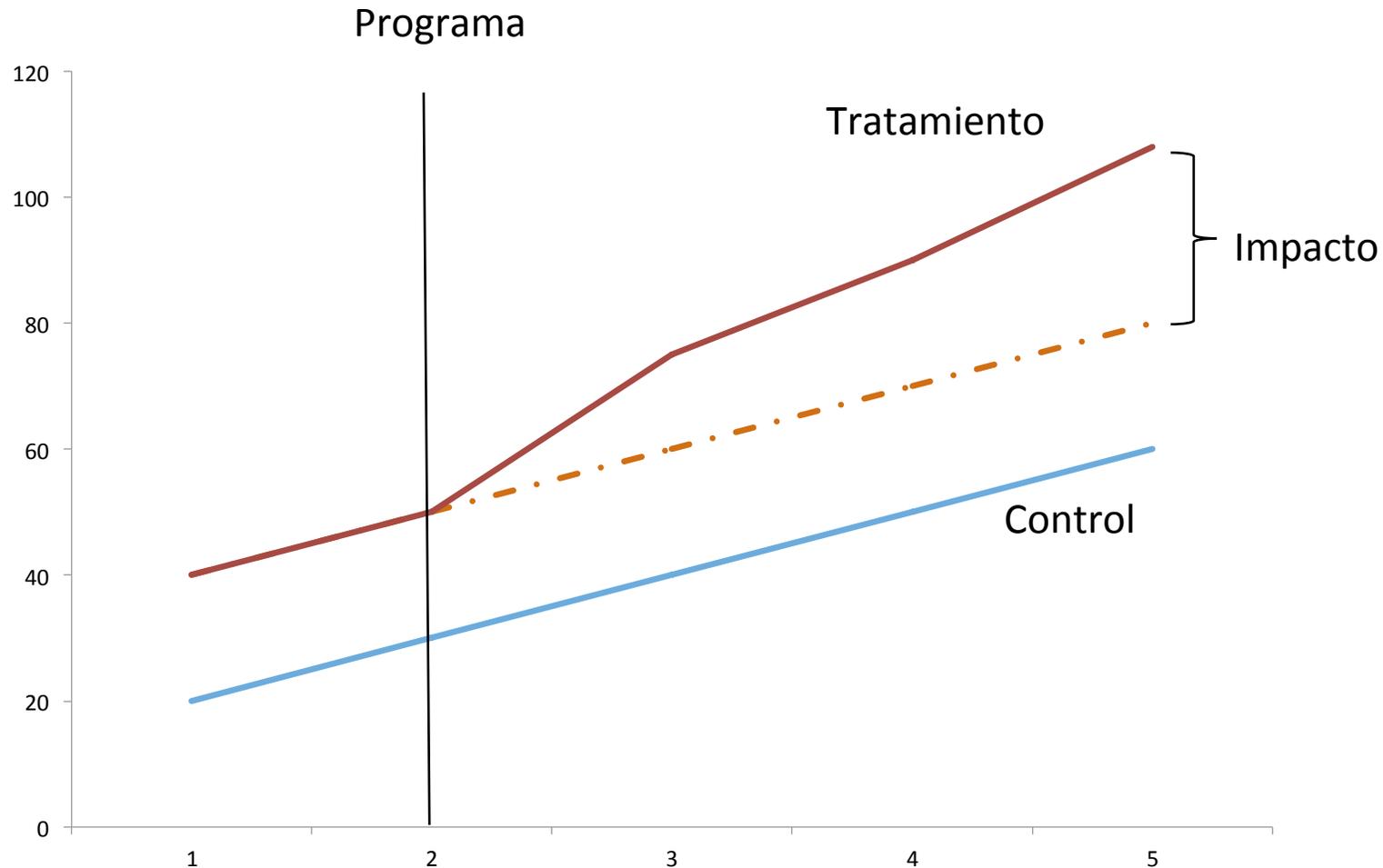
- En el tiempo: antes y después
- Entre grupos: control y tratamiento

1. Buscamos un **grupo de control adecuado**
2. Calculamos el **cambio en la variable de interés entre los dos períodos** tanto para el grupo de control como para el grupo “experimental”
3. Calculamos la diferencia entre el cambio en el tiempo de ambos grupos:

	Tratamiento	Control
T=0	$Y_{0t}$	$Y_{0c}$
T=1	$Y_{1t}$	$Y_{1c}$

Lo que calcularíamos aquí es:  $(Y_{1t} - Y_{0t}) - (Y_{1c} - Y_{0c})$

# Diferencias en diferencias



Para que esto sea un estimador insesgado del impacto:  
asumimos  
tendencias  
paralelas

# Un ejemplo: un mayor despliegue policial y el robo de vehículos en Argentina

## La pregunta de evaluación:

¿Una mayor despliegue policial ayuda a reducir delitos?

## La variable de impacto:

Delitos (robo de vehículos)

## ¿Cuál es el sesgo de selección?

Las zonas con mayores delitos suelen recibir mayor atención policial

## La intervención:

Mayor despliegue policial (causado por los ataques terroristas a centro judío)

## Los datos:

Datos geo-referenciados de delitos

Para zonas que recibieron mayor despliegue y para zonas que no recibieron mayor despliegue

Datos para antes del despliegue y para después

# Calculando el impacto

	Tratamiento	Control	Diferencia
Antes	15	20	-5
Después	10	25	-15
Diferencia	-5	5	-10

# ¿Qué datos requerimos?

- Datos de corte transversal:

- Dos grupos: control y tratamiento
- Dos mediciones en el tiempo: antes y después
- **NO** medimos a la misma gente en ambos períodos
- En ambos períodos levantamos información de una muestra representativa de ambos grupos
- De cada individuo solo tenemos una observación

- Datos de panel:

- Dos grupos: control y tratamiento
- Dos mediciones en el tiempo: antes y después
- Medimos a **la misma gente en ambos períodos**
- De cada individuo al menos dos observaciones

# Problemas con Difference-in-Difference

- “Ashenfelter dip”: el programa puede aplicarse al grupo con peor desempeño justo antes de la intervención, esto sesga hacia arriba el estimador DD.
- Si hay cosas que cambian en el tiempo de forma diferente para ambos grupos, no es muy útil.

# Elementos a tener en cuenta para hacer una evaluación de impacto

- Identificar claramente:
  - ✓ cuál es la pregunta de evaluación, qué queremos aprender
  - ✓ cuál es la intervención que se quiere evaluar (o su componente)
  - ✓ quién o qué recibe la intervención: una persona, una familia, una empresa, un contribuyente
- ¿Cómo se cuantifica el beneficio: variables y forma de medir?
- ¿Hay datos administrativos?
- ¿Cómo se construye el contra-factual? (experimentos, DiD, emparejamiento, RDD, IV, etc.)
- Hay compromiso institucional