

# Seminario sobre la evaluación de impacto de políticas públicas para las micro, pequeñas y medianas empresas en América Latina y El Caribe

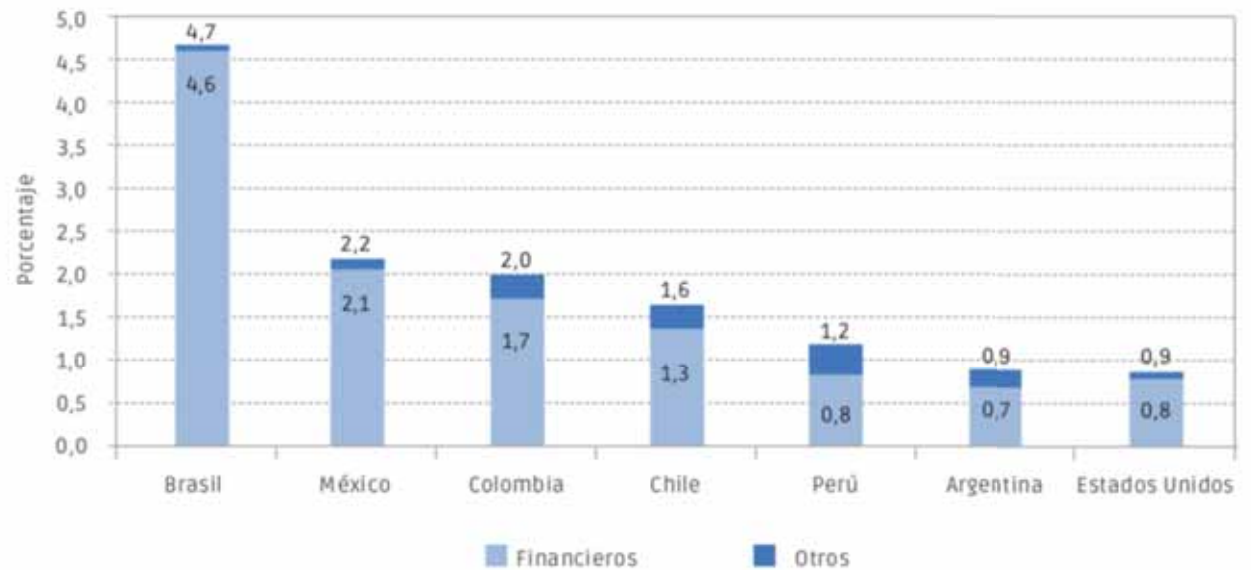


Lesbia Maris – DEIAP (CAF)  
Septiembre 2018



## Evaluación de Impacto y Aprendizaje de Políticas

**Gráfico 5.1** Esfuerzo público total para el apoyo al emprendimiento y la productividad en países seleccionados de América Latina y Estados Unidos como porcentaje del PIB (2011)



Fuente: RED 2013 (CAF)

**¿Hay más y mejores  
emprendimientos productivos?**

**¿Se han logrado los objetivos?**

**¿Ha aumentado la productividad?**

**¿Qué sabemos sobre los programas que implementamos?**

## Una política pública efectiva

---



Más oportunidades, un mejor futuro.

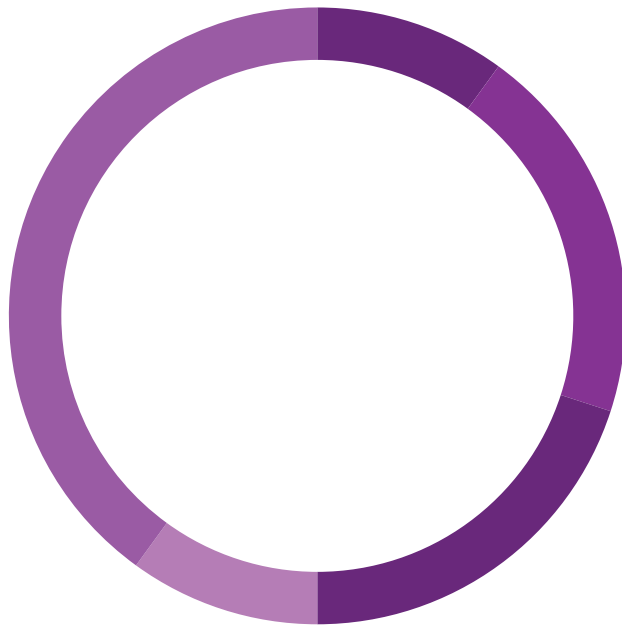
**Recursos  
económicos**

**Recursos  
humanos**

**Conocimiento**

**Restricciones  
institucionales**

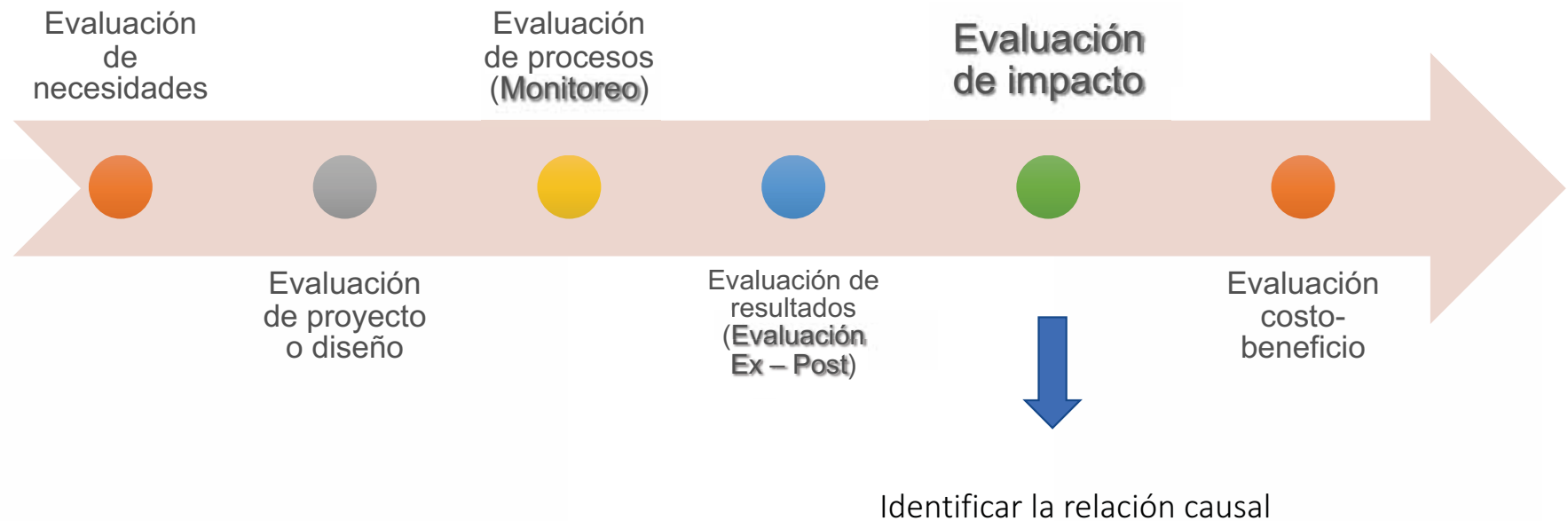
**Restricciones  
políticas**



## Hoja de ruta

- 1** *Introducción sobre el aprendizaje de políticas*
- 2** *La evaluación de impacto*
- 3** *Evaluaciones experimentales*
- 4** *Métodos no experimentales*
- 5** *La evaluación de las políticas de emprendimiento y productividad*
- 6** *Taller práctico de evaluación de impacto*

## ¿A qué nos referimos cuando hablamos de aprendizaje?



# Tipos de evaluaciones



## ¿Qué es una evaluación de impacto?

Es la cuantificación de los efectos que **se pueden atribuir** a un programa o proyecto

**Cuantificar**

**Demostrar que se debe sólo a mi proyecto**

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



# ¿Qué será más efectivo para aumentar la asistencia escolar?

---



**Medicinas para eliminar  
gusanos intestinales**



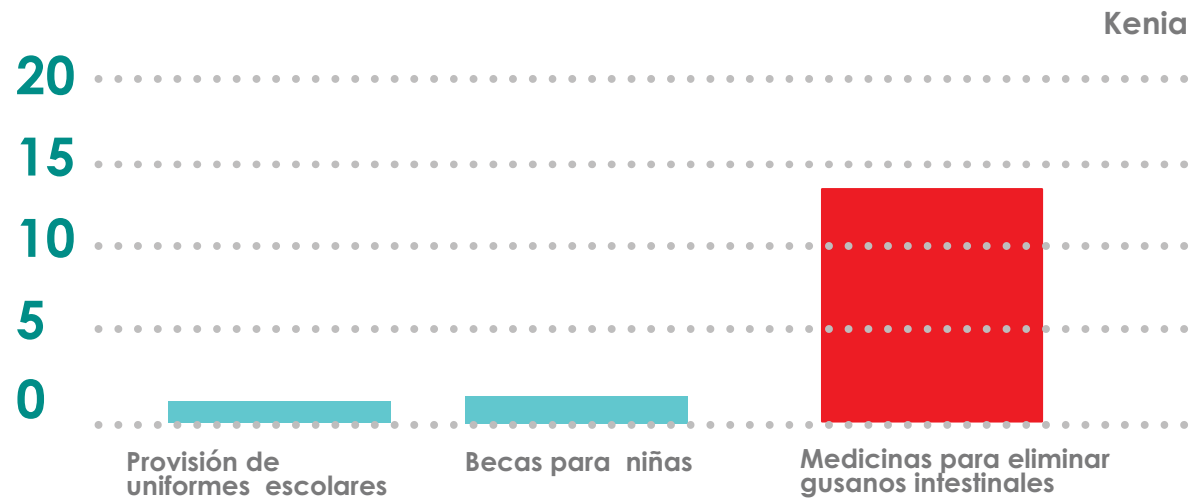
**Becas para niñas**



**Provisión de  
uniformes escolares**

Más oportunidades, un mejor futuro.

## ¿Cuántos años de educación podemos aumentar con 100 dólares?



J-PAL Policy Bulletin. 2012. "Deworming: A Best Buy for Development." Cambridge, MA: Abdul Latif Jameel Poverty Action Lab

Más oportunidades, un mejor futuro.

## ¿Qué tipo de preguntas responde la evaluación?

¿Cuál es la magnitud de los beneficios generados por un programa?

Envío de notificaciones de pago a contribuyentes morosos en Guatemala

**Los resultados fueron positivos** para todos los tipos de mensajes, incrementaron el pago de impuestos en 43%

¿Cuál es la mejor forma de hacer el delivery de una política/servicio?

Sin embargo, **unos mensajes fueron más efectivos que otros**

Si se hubiese enviado la carta vinculada a normas sociales a toda la muestra, en 11 semanas hubiera generado **más de 750 mil dólares de ingresos adicionales -- 36 veces el costo de enviar las cartas**

## ¿Qué tipo de preguntas responde la evaluación?

¿Qué grupo se beneficia más?

Fútbol con  
Corazón



**Efectos diferenciados por grupos de edad**

**5 a 7 años** no se encontró resultado

**8 a 13 años** mejoras en la actitud hacia el colegio, autoestima, percepción de roles de género y uso del tiempo

**14 a 17 años** resultados negativos

¿En qué condiciones el programa tiene mayor beneficio?



La instalación de **cámaras de videovigilancia** resultó ser más efectiva en disminuir el crimen en estacionamientos

Ha demostrado ser más efectiva en Reino Unido que en otros países

## ¿Qué tipo de preguntas responde la evaluación?

### ¿Qué componente de un programa es más beneficioso?

Las capacitaciones vocacionales y empresariales dirigidas a mujeres son más efectivas si:

1. **Tienen un enfoque de género específico:** es decir, si el programa toma en cuenta las restricciones de movilidad de las mujeres, los patrones de uso del tiempo
2. **Incluye formación en habilidades para la vida**



# ¿Cómo medimos el impacto de una política?

El **reto fundamental** es responder a la siguiente pregunta: “¿cuál hubiese sido el resultado observado si el tratamiento/programa no se hubiese aplicado?” (**el contra-factual**)

Con esta información, cuando medimos el impacto, lo que estamos haciendo es la siguiente comparación:

¿Qué pasó  
después de  
implementarla  
política?

VS.

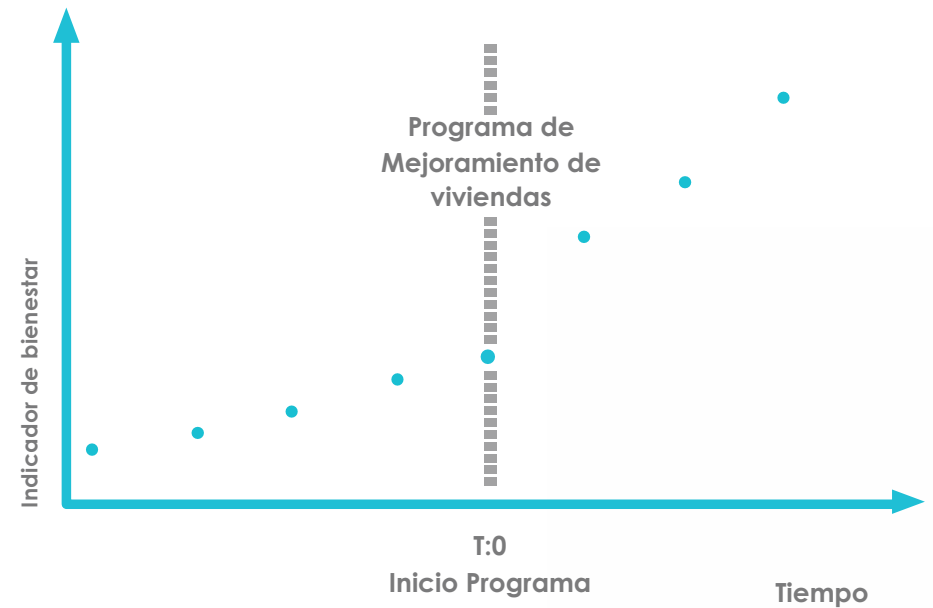
¿Qué hubiese pasado si  
no hubiésemos  
implementado la política?

## Un ejemplo...

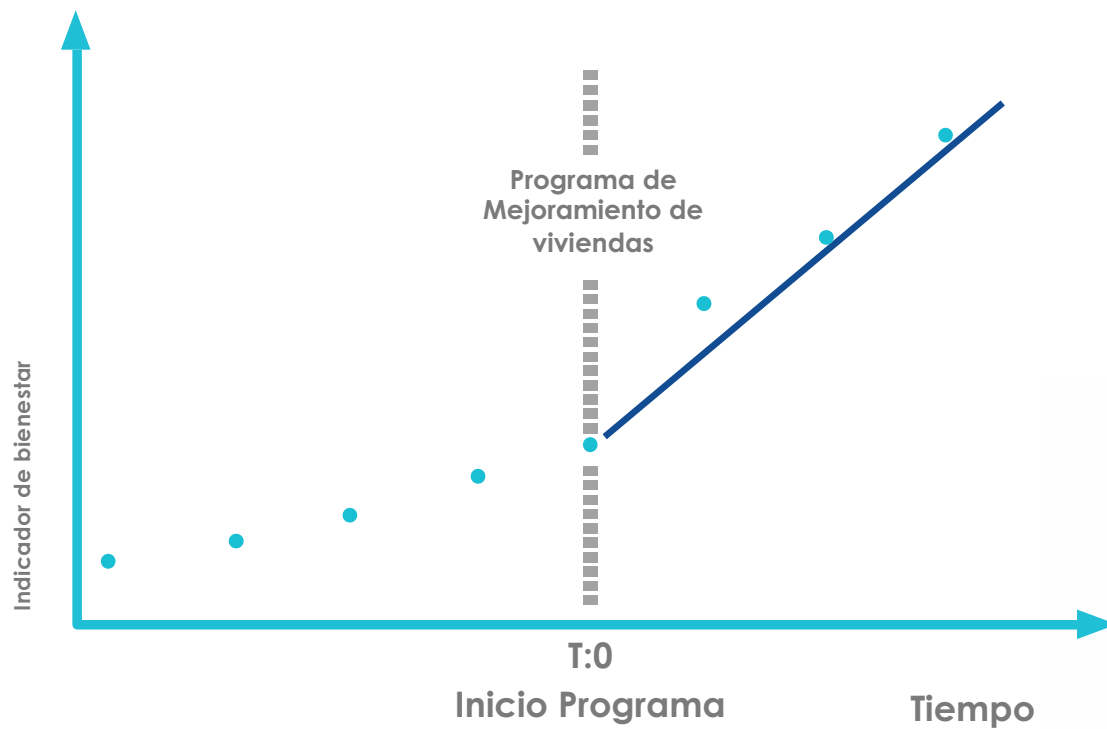
Supongamos que se busca medir el impacto de un programa de entrega de créditos a pequeñas y medianas empresas.

Según el siguiente gráfico, el impacto del programa sobre el tamaño de las ventas es:

- A. Positivo
- B. Nulo
- C. No se sabe



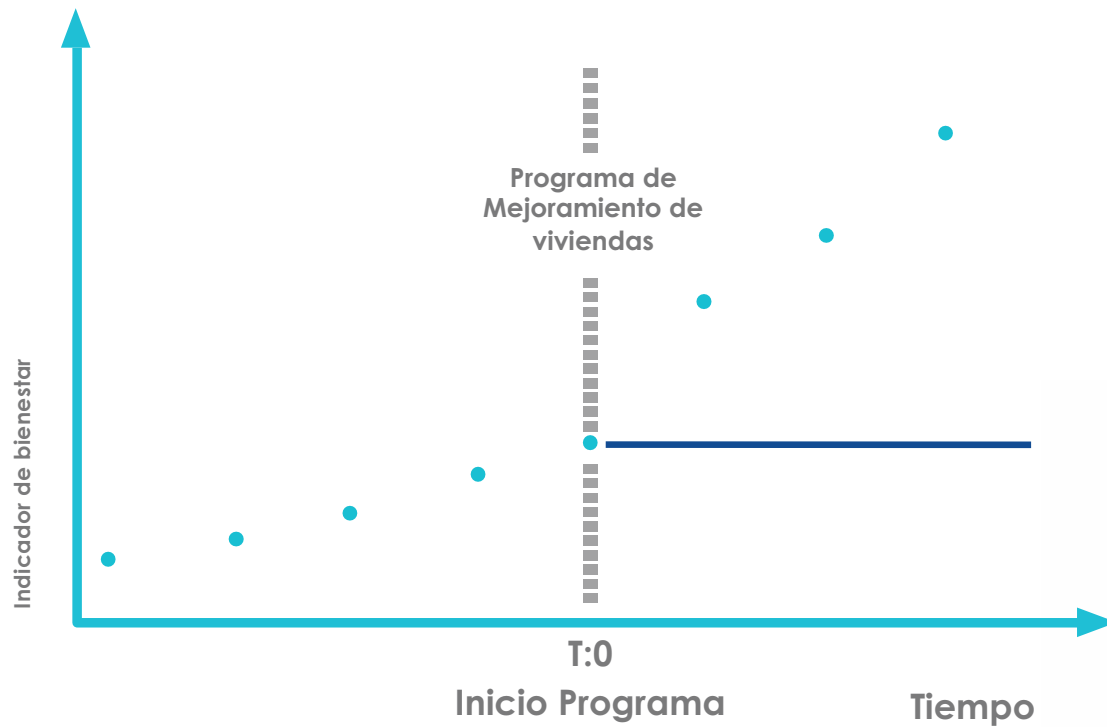
Todo depende de cómo ha evolucionado el bienestar del grupo de control



Más oportunidades, un mejor futuro.



Todo depende de cómo ha evolucionado el bienestar del grupo de control



Más oportunidades, un mejor futuro.

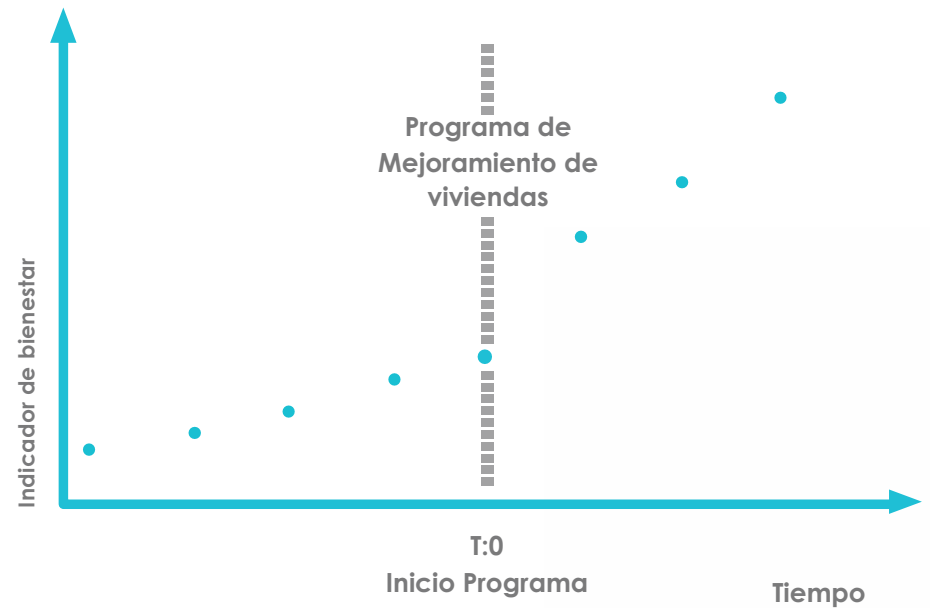
## Un ejemplo...

Supongamos que se busca medir el impacto de un programa de entrega de créditos a pequeñas y medianas empresas.

Según el siguiente gráfico, el impacto del programa sobre el tamaño de las ventas es:

- A. Positivo
- B. Nulo
- C. No se sabe**

Más oportunidades, un mejor futuro.



No podemos saber ¿qué hubiese pasado con aquellos que recibieron las viviendas en el caso de no haberlas recibido?

## ¿Por qué?

- ✓ Necesitaríamos observar al **mismo individuo**, con y sin el programa
- ✓ ...pero es **imposible observar** al mismo individuo en los dos estados (con y sin programa...)



## Concepto: contrafactual

---

Escenario hipotético que representa **lo que le hubiera ocurrido** a la **unidad de análisis** en **ausencia** de la intervención.

No es cualquier unidad de análisis

## ¿Qué es una evaluación de impacto?

Es una identificación **sistemática** de los efectos que **se pueden atribuir** a un programa o proyecto

**Cuantificar**

**Demostrar que se debe sólo a mi proyecto**

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



## ¿Qué la diferencia de otras metodologías?

### Monitoreo

Permite verificar si las actividades se están realizando según lo planificado (**Hotspots – Colombia**)

### Evaluación de resultados

Permite medir los resultados a corto y largo plazo del programa de acuerdo a los resultados esperados establecidos antes del inicio del programa (**AVANZA-Venezuela**)

Más oportunidades, un mejor futuro.

### Evaluación cualitativa

Captura información cualitativa sobre el funcionamiento de un programa (**Formación ciudadana – Colombia**)

# Ejercicio 1: Una pregunta de evaluación de impacto

1. ¿Cuál es la **intervención / Programa**?
2. ¿Cuáles son los **resultados** que podemos observar después de implementar el programa?
3. ¿Cuál es la **pregunta de evaluación de impacto** aquí?

NO participa en  
programa de  
alimentación escolar



Participa en  
programa de  
alimentación escolar



En casa recibieron la  
misma educación, la  
misma alimentación y  
los mismos cuidados



La gemela que  
participó creció más

Más oportunidades, un mejor futuro.



# Contrafactual

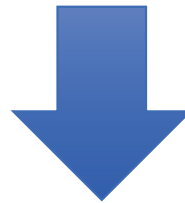
Encontrar individuos (CLONES) que no sean beneficiarios del programa, y que tienen características, en promedio, muy similares a los beneficiarios antes de comenzar el programa

- ✓ Los factores externos tienen que afectar de la misma forma al grupo de comparación (control) y al tratamiento
- ✓ Los efectos del programa deberían ser los mismos para ambos grupos

# Efecto Causal

Rara vez podemos observar a la misma persona 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**

# ¿Cómo obtenemos la información que nos falta?

1. ¿Buscamos a un grupo que no haya participado?

Puede ser que el grupo ya era diferente ANTES



# ¿Cómo obtenemos la información que nos falta?

2. ¿Hacemos una comparación antes-después?

Puede ser que se implementaran otros programas en el intermedio ¿Cómo saber cuál es el responsable de lo que observamos?

## Problemas comunes a la hora de construir el contra-factual

- Los programas suelen estar focalizados
- La participación en un programa suele ser voluntaria



**Sesgo de selección**

## Ejercicio 2: El contrafactual y el sesgo de selección

1. ¿Cuál es el contrafactual implícito en la pregunta de evaluación?
2. Escoja un grupo de comparación
3. ¿Existe riesgo de sesgo de selección?

# 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”

# ¿Cómo construimos un buen contra-factual?

## El secreto:

Identificar cómo se decide quién es beneficiado y quién no  
Lograr que los grupos que comparemos sean iguales (en promedio),  
**excepto que** uno recibió el programa y otro no

Por un sorteo	Por algún criterio sistemático parecido a un sorteo	Por algún criterio de focalización
<b>Experimentos</b>	<b>Experimentos naturales</b>	<b>Métodos no experimentales:</b> <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>



## Un experimento (RCT):

¿Cuál es la forma más efectiva de lograr que los contribuyentes paguen sus obligaciones pendientes?



**Programa:** Jornadas de cobro a contribuyentes morosos: visitas, envío de cartas, envío de emails

**Pregunta:** ¿Cuál forma de contacto es más efectiva en lograr el pago de obligaciones pendientes?

**Selección:** 20.000 contribuyentes fueron asignados por sorteo a recibir una visita, o una carta, o un correo o a no ser contactados

**Las visitas son las más efectivas...  
pero las más costosas**



## Concepto: Experimento (RCT)

---

Se decide por **SORTEO** quién participa en un programa y quién no, con el objetivo de medir el impacto del programa en los beneficiarios

# Algunos ejemplos de aleatorizaciones

## Asesorías y coaching para emprendedores

1. Período de aplicación de dos meses
2. Se inscriben 600 empresas
3. Solo 500 son elegibles
4. El programa solo puede atender a 200
5. Se escogen por sorteo 200 que participarán y 300 que quedarán como grupo de control



# Algunos ejemplos de aleatorizaciones

## UNICAS– Grupos de ahorro

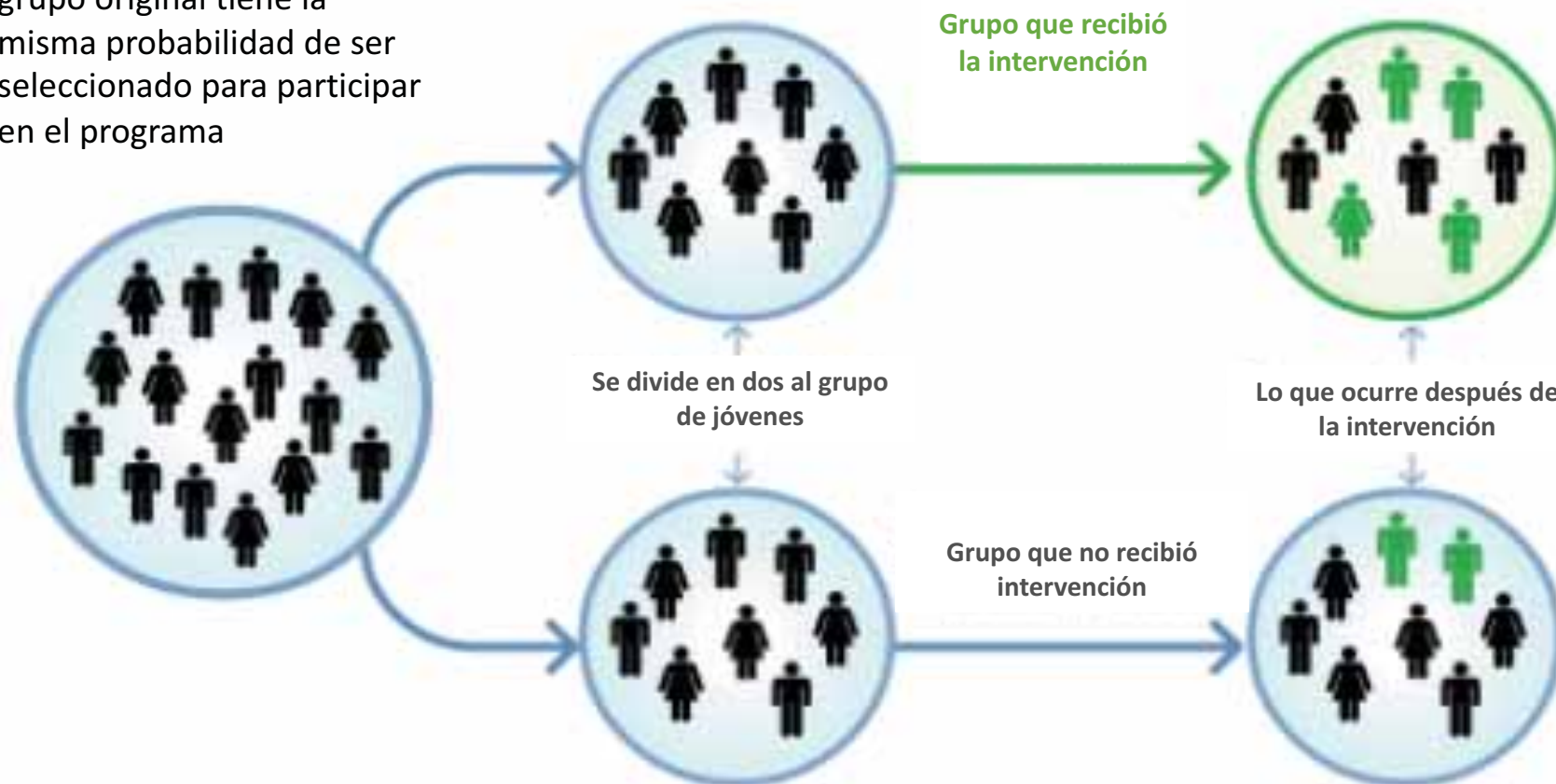


1. La promoción para la formación de UNICAS se hace en centros poblados
2. Se identifican 200 centros poblados
3. Se eligen por sorteo 100 de ellos para recibir la visita de los promotores y quedan 100 como grupo de comparación

# ¿Cómo funcionan los experimentos?

## IMPORTANTE:

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



# ¿Cómo funcionan los experimentos?

## Experimento de cobro de impuestos (DIAN)

	Probabilidad de pago
Grupo de control	5%
Grupo de tratamiento	22%
Diferencia (Impacto)	17 puntos porcentuales

Promedio de variable de resultado para grupo de control

Promedio de variable de resultado para grupo de tratamiento

**Impacto**



**22%**

Factual



**5%**

Aproximación al  
contrafactual (a través  
de un experimento)

# ¿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.

# Aleatorizando en la vida real

- **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.



# Algunos ejemplos interesantes de experimentos

¿Cómo fomentar la innovación y la creación de empleo a través de emprendimientos?

- A través de recursos financieros otorgados a las empresas ganadoras en una competencia de planes de negocio (Nigeria).
- Dado el volumen de aplicaciones, una parte de las ganadoras se pudo seleccionar por sorteo (729 tratadas y 1112 en el control).
- Las ganadoras recibieron un máximo de 4 pagos (dependiendo del monto requerido para el emprendimiento)
- Resultados:
  - Mayor probabilidad de sobrevivencia
  - Mayor empleo
  - Más innovación

Más oportunidades, un mejor futuro.

Fuente: McKenzie (2015)

# 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.

## Experimentos controlados aleatorios (RCT)

### Ventajas

- El RCT es la metodología ideal
- Es una manera justa de asignar los recursos

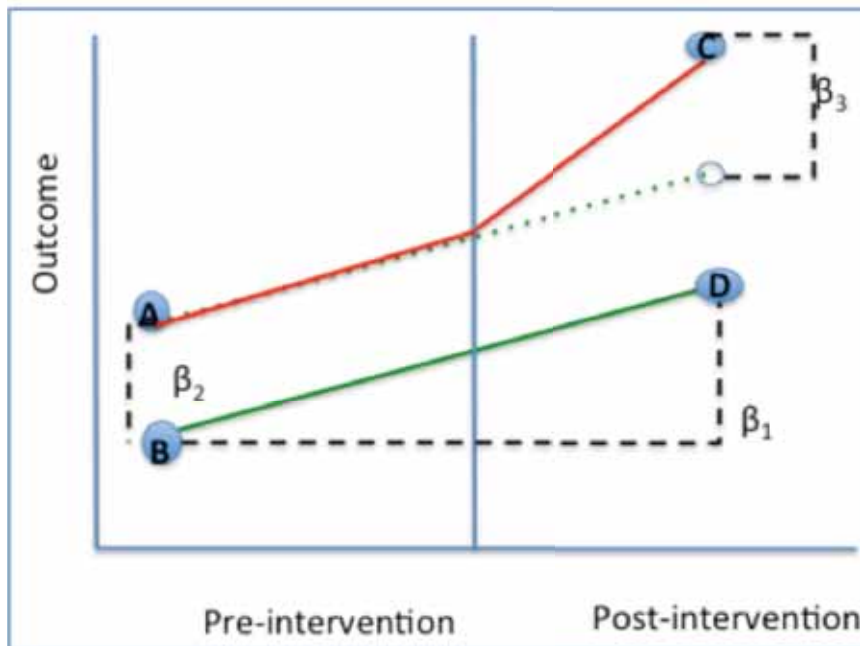
### Desventajas

- Políticamente más complicado
- Solo impacto promedio
- Caja negra
- Validez externa
- Incumplimiento de la asignación
- Cuidado con los efectos de “estar siendo medidos”

# 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?

# 1. Diferencias-en-diferencias (Diff-in-Diff)



## ¿Cuándo podemos utilizarla?

- 1 Tenemos datos para ambos grupos en varios momentos del tiempo (mientras más mejor)
- 2 La tendencia antes del programa es la misma para el grupo de tratamiento y para el grupo de control (tendencias paralelas)
- 3 Cuando estamos en presencia de experimentos naturales (cambios de legislación en algunos estados o solo para algún grupo poblacional)

Más oportunidades, un mejor futuro.

# Diferencias en diferencias

Se hacen dos comparaciones:

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

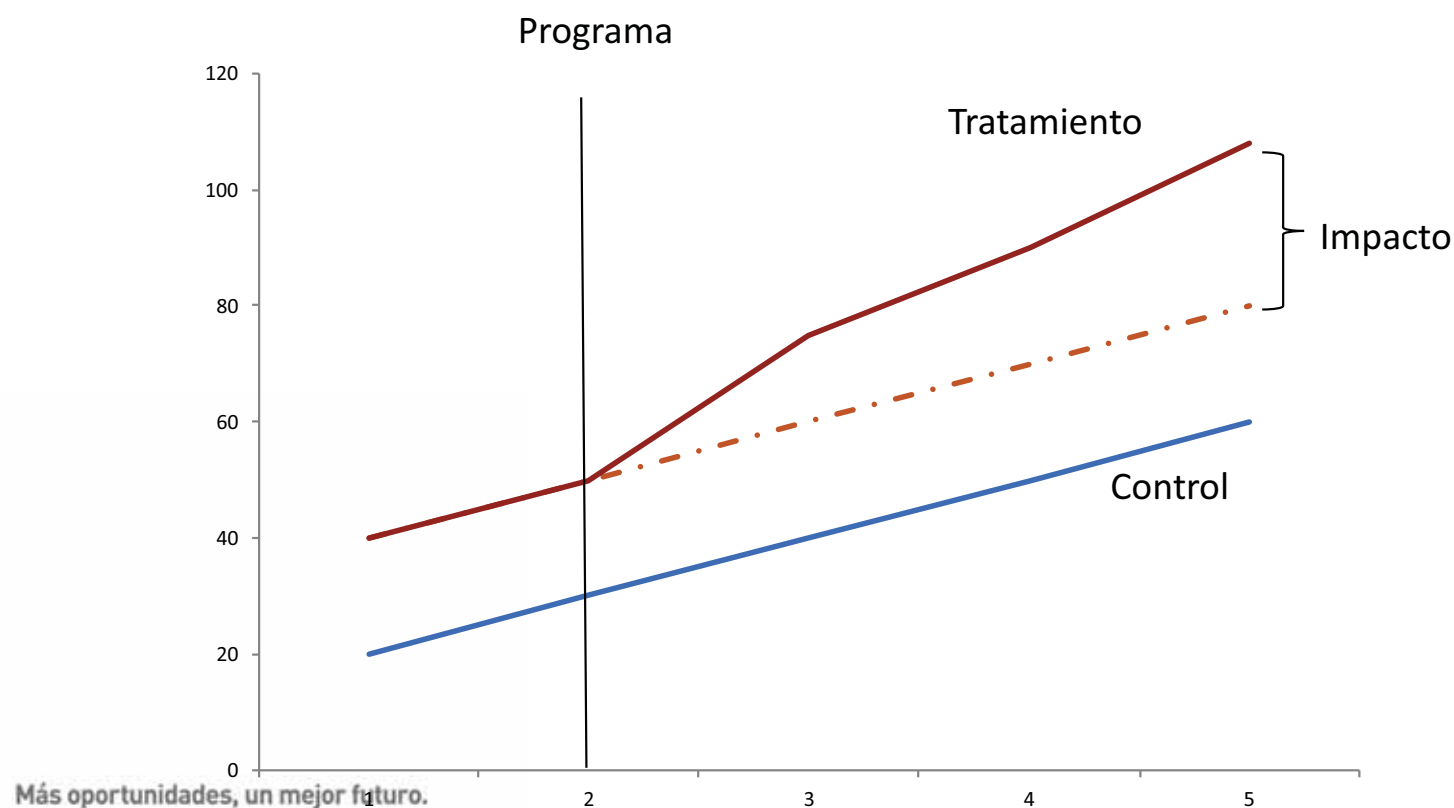
1. Buscamos un **grupo de control adecuado**
2. Calculamos el **cambio en la variable de interés entre los dos períodos** tanto para el grupo de control como para el grupo “experimental”
3. Calculamos la diferencia entre el cambio en el tiempo de ambos grupos:

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

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

Más oportunidades, un mejor futuro.

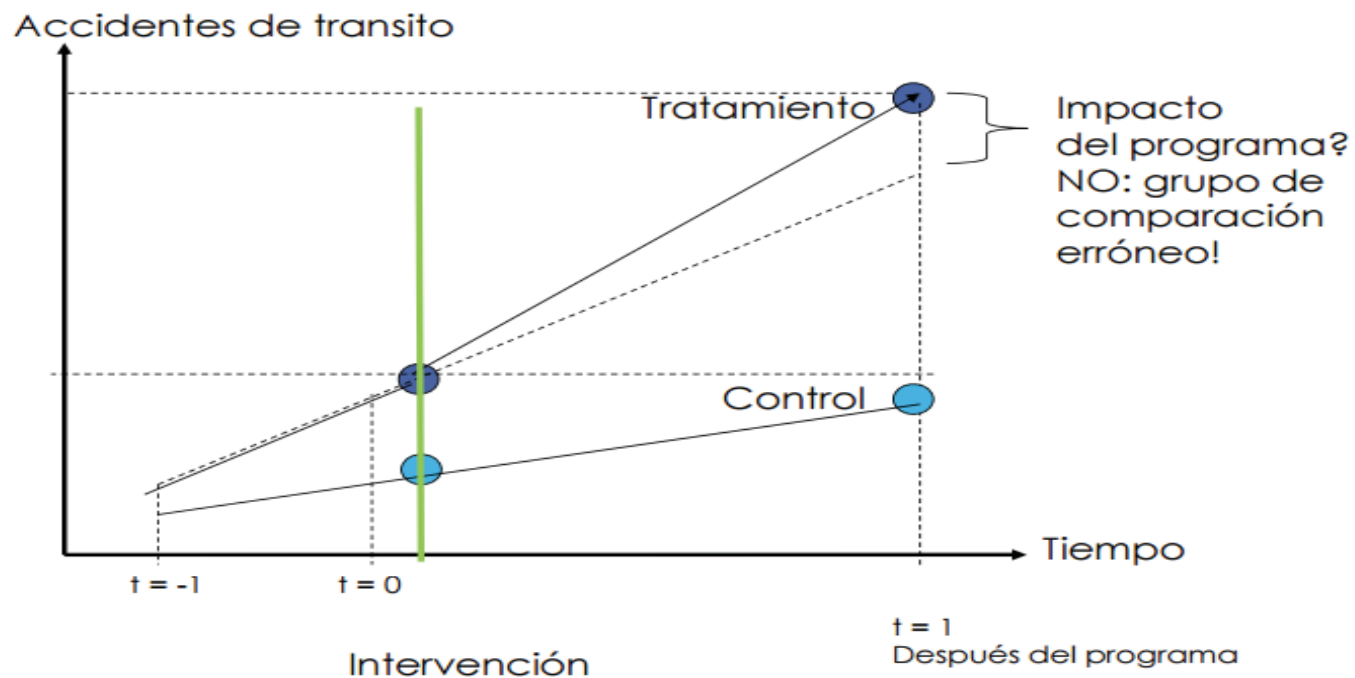
# Diferencias en diferencias



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



# ¿Cuándo es **erróneo** usar Diff-in-Diff?



**Tendencias Paralelas antes de cambio de la norma NO se cumplen**

Fuente: Adriana Camacho

Más oportunidades, un mejor futuro.

# 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)

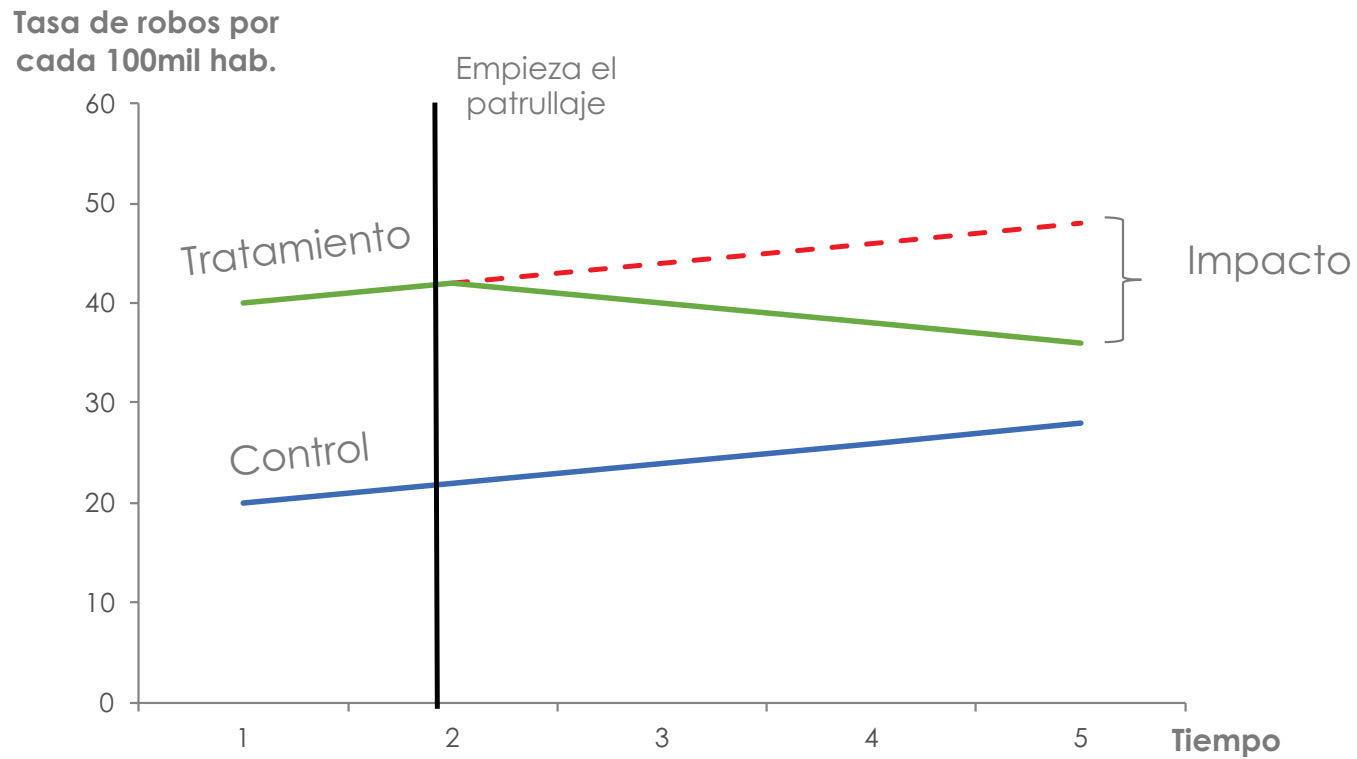
## Los datos:

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

## ¿Cómo construimos un grupo comparable?



Asumimos  
tendencias  
paralelas



**-5**

Evolución  
en zonas  
tratadas

**5**

Evolución  
en zonas  
no  
tratadas

**-10**

Diferencia en  
la evolución de  
ambas zonas

## La evolución de los robos de vehículos

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

Más oportunidades, un mejor futuro.

# ¿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



**Más oportunidades, un mejor futuro.**

[www.caf.com](http://www.caf.com) | [@AgendaCAF](https://twitter.com/AgendaCAF)