

TALLER PARA LA EVALUACIÓN DE PROYECTOS DE PREVENCIÓN SOCIAL DE LA VIOLENCIA Y LA DELINCUENCIA



SEGOB
SECRETARÍA DE GOBERNACIÓN



PROGRAMA PARA LA
CONVIVENCIA CIUDADANA

ABDUL LATIF JAMEEL
Poverty Action Lab
TRADUCIENDO LA INVESTIGACIÓN EN ACCIÓN

Sesión 4

Evaluación de Impacto

Vivian Bronsoler

Gerente de Políticas

Abdul Latif Jameel Poverty Action Lab



OBJETIVOS


Después de esta sesión los participantes serán capaces de:

- Entender el concepto de causalidad.
- Detallar los problemas asociados con atribuir causalidad en las ciencias sociales.
- Diferenciar una evaluación de impacto experimental de una evaluación no experimental.
- Entender las ventajas que presentan las evaluaciones experimentales.



ESQUEMA DE LA PRESENTACIÓN

1. ¿Qué se requiere para atribuir causalidad?
2. Estimación del grupo de comparación
3. ¿Por qué aleatorizar?
4. Críticas a las evaluaciones experimentales
5. Pasos clave para su diseño

The slide features decorative gear icons in the corners. In the bottom-left corner, there are two overlapping grey gears of different sizes. In the top-right corner, there are two overlapping grey gears, one larger than the other. The main text is centered on the slide.

1. ¿Qué se requiere para atribuir causalidad?



RELACIONES CAUSALES

El objetivo de muchos estudios en ciencias sociales y naturales es encontrar relaciones de causa-efecto entre variables o eventos.

¿Por qué es importante establecer relaciones causales?

- *Objetivo científico:* mejorar teorías, modelos y cómo entendemos el mundo.
- *Objetivo práctico:* mejorar la calidad de las decisiones (políticas públicas, sistemas de incentivos y políticas organizacionales, etc.)



PREGUNTAS

¿Cuál es el efecto de *fomentar actividades deportivas* en *rendimiento académico de los niños*?

¿Cuál es el efecto de *mejorar los espacios públicos* en el *capital social de las comunidades*?

¿Qué efecto tiene la *capacitación técnica* en la *reinserción social de los jóvenes*?

¿Cuál es el efecto de *brindar charlas sobre el consumo de alcohol* en *jóvenes*?





PROBLEMA DE ATRIBUCIÓN CAUSAL

Aún habiendo definido:

1. El programa que queremos implementar,
2. Las variables que miden el resultado

...contestar estas preguntas no es trivial.

¿Por qué?



ATRIBUCIÓN



Por ejemplo, queremos medir el **efecto del servicio militar obligatorio sobre la tasa de delincuencia**.

- Si comparamos países que tienen servicio militar obligatorio con los que no tienen, podemos estar comparando peras con manzanas.
- Los países pueden ser diferentes en cuanto a:
 - Ingresos promedios
 - Niveles de educación promedios
 - Desigualdad

(¿Por qué en algunos países es obligatorio?)

Estos otros factores también podrían afectar la tasa de delincuencia.

PROBLEMA DE ATRIBUCION CAUSAL



Otro ejemplo: Queremos saber si ir al colegio reduce las tasas de consumo de alcohol y tabaco de adolescentes.

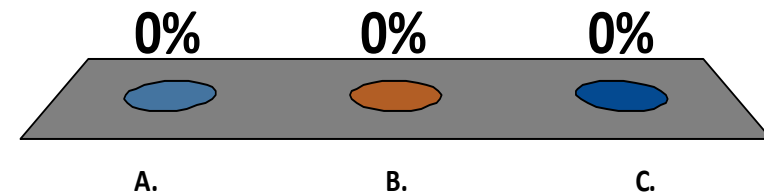
- ¿Podemos comparar la tasa de consumo de niños que van al colegio con la tasa de consumo de niños que no van al colegio?



PREGUNTAS

¿Cuáles podrían ser algunas diferencias entre adolescentes que van al colegio y adolescentes que no van al colegio?

- A. Los que van al colegio pueden ser más influenciados por compañeros de su misma edad.
- B. Los que no van al colegio pueden trabajar más y tener mayor poder adquisitivo.
- C. Todas las anteriores.
- D. No hay diferencias importantes.





PROBLEMA DE LA ATRIBUCIÓN CAUSAL

Este problema se conoce como

Problema de selección

- Las personas que eligen participar en el programa tienen características distintas a las personas que no lo hacen.
- Las personas que son elegidas para recibir un beneficio tienen características distintas a las personas que no fueron elegidas.



2. Estimación de grupo de comparación





PROGRAMA DE CAPACITACIÓN LABORAL A ADOLESCENTES

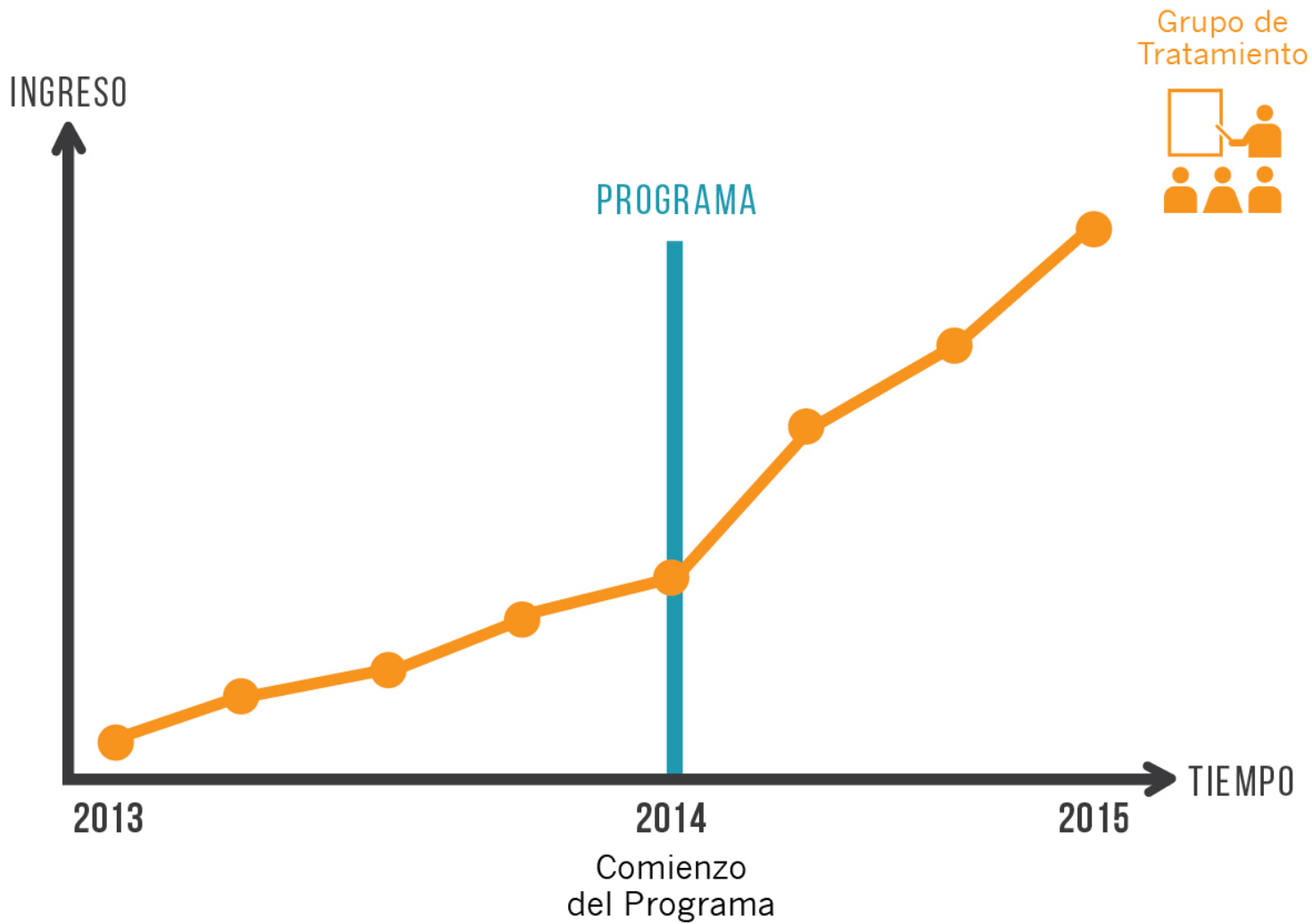


Programa de capacitación profesional para jóvenes vulnerables en Colombia.

- Busca incrementar las probabilidades de obtener empleo.
- Tiene por objetivo incrementar los ingresos provenientes del trabajo formal y lícito.



¿CUÁL FUE EL IMPACTO?



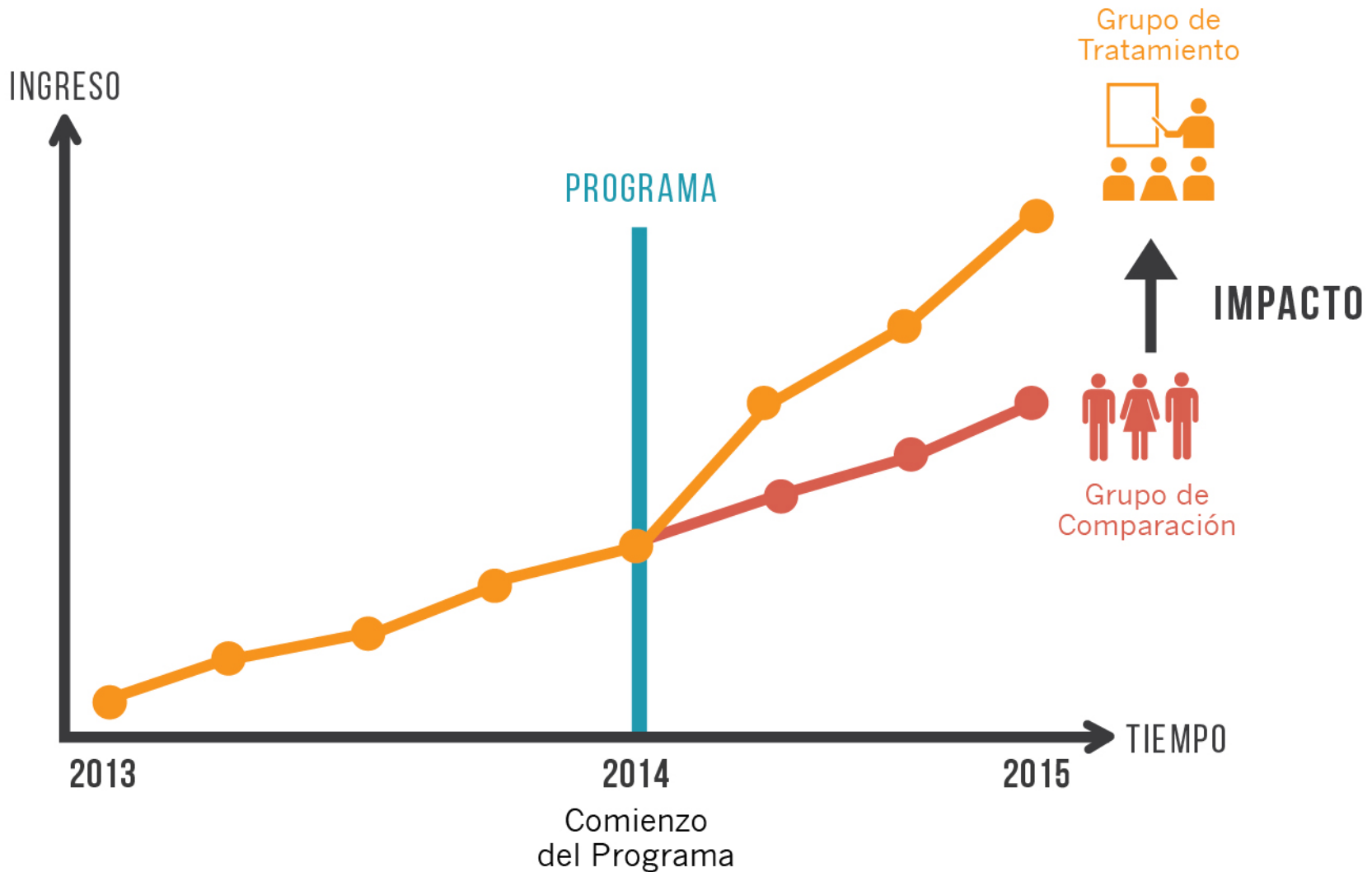


PREGUNTA

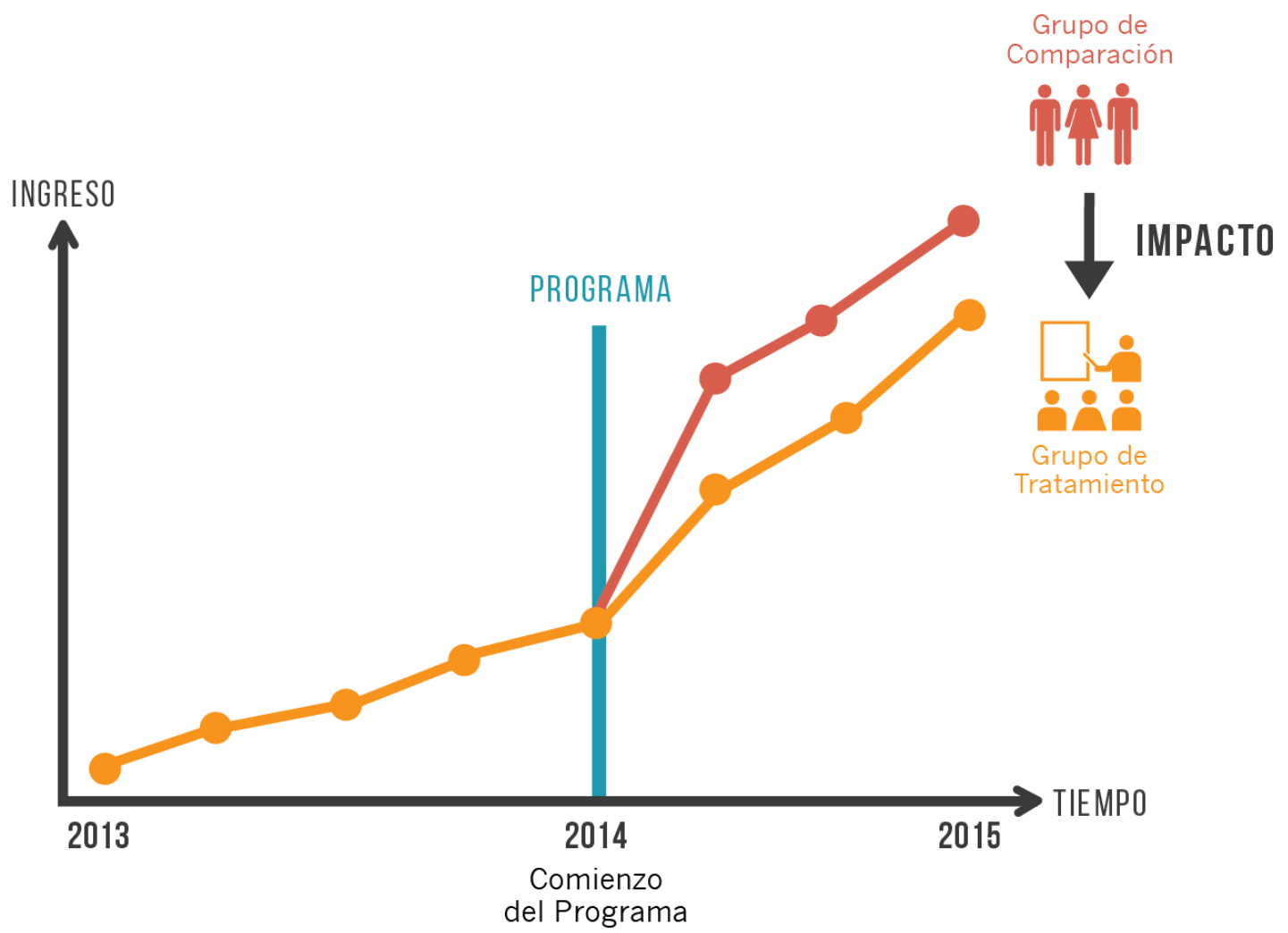
El impacto del programa fue:

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

¿CUÁL FUE EL IMPACTO?



¿CUÁL FUE EL IMPACTO?





¿QUÉ ES IMPACTO?

Ingreso laboral de los jóvenes que reciben la capacitación un tiempo después de participar en el programa

Ingreso laboral que esos mismos jóvenes hubieran ganado en ese mismo momento *en el caso hipotético de no haber participado en el programa*

Grupo de comparación

IMPACTO

- Es hipotético:
 - Representa el estado del mundo que se habría experimentado si no se hubiera aplicado el programa
- Problema:
 - El *contrafactual* (ese grupo de comparación) no se puede observar.
- Solución:
 - Necesitamos estimar o construir un grupo de comparación que sea lo más parecido al grupo hipotético.



ESTIMANDO EL GRUPO DE COMPARACIÓN

- Se hace mediante la selección de un grupo de personas que no participaron en el programa
- Este grupo se llama el **grupo de control** o el **grupo de comparación**
- La forma de seleccionar este grupo es una decisión clave en el diseño de cualquier evaluación de impacto



SELECCIONANDO EL GRUPO DE CONTROL

El objetivo es seleccionar un grupo que **en promedio sea igual** al grupo de participantes en todas las dimensiones, **excepto en si reciben o no el programa.**

- Igual proporción de hombres y mujeres
- Igual número de años de educación
- Igual situación económica
- Mismas expectativas
- Igual en términos de motivación
- Etc...



Métodos de evaluación experimentales

- Selección aleatoria del grupo de control
- Muy confiables si son bien implementados

Métodos no experimentales y cuasi experimentales

- Antes y después
- Diferencia simple
- Diferencias en diferencias
- “Matching”
- Regresión discontinua



MÉTODOS NO EXPERIMENTALES

Todos los métodos tratan de hacer lo mismo:

- Estimar qué hubiera pasado sin el programa por medio de la **construcción de un grupo de comparación**.
- Estimar la diferencia entre los resultados de los grupos de tratamiento y comparación.

La principal diferencia entre estos métodos es **cómo estiman el grupo de comparación**,

¿Quién conforma el grupo de comparación ?

MÉTODOS NO EXPERIMENTALES

- Confiabilidad de las metodologías:

En evaluaciones no experimentales, **se tiene que hacer muchos supuestos** para poder afirmar que los grupos de tratamiento y control son comparables.

¡Puede ser que los grupos no sean tan comparables!

¿Qué tan sensato es pensar que esos supuestos se cumplen en realidad?



CONTRAFACTUAL FALSO 1: ANTES Y DESPUÉS



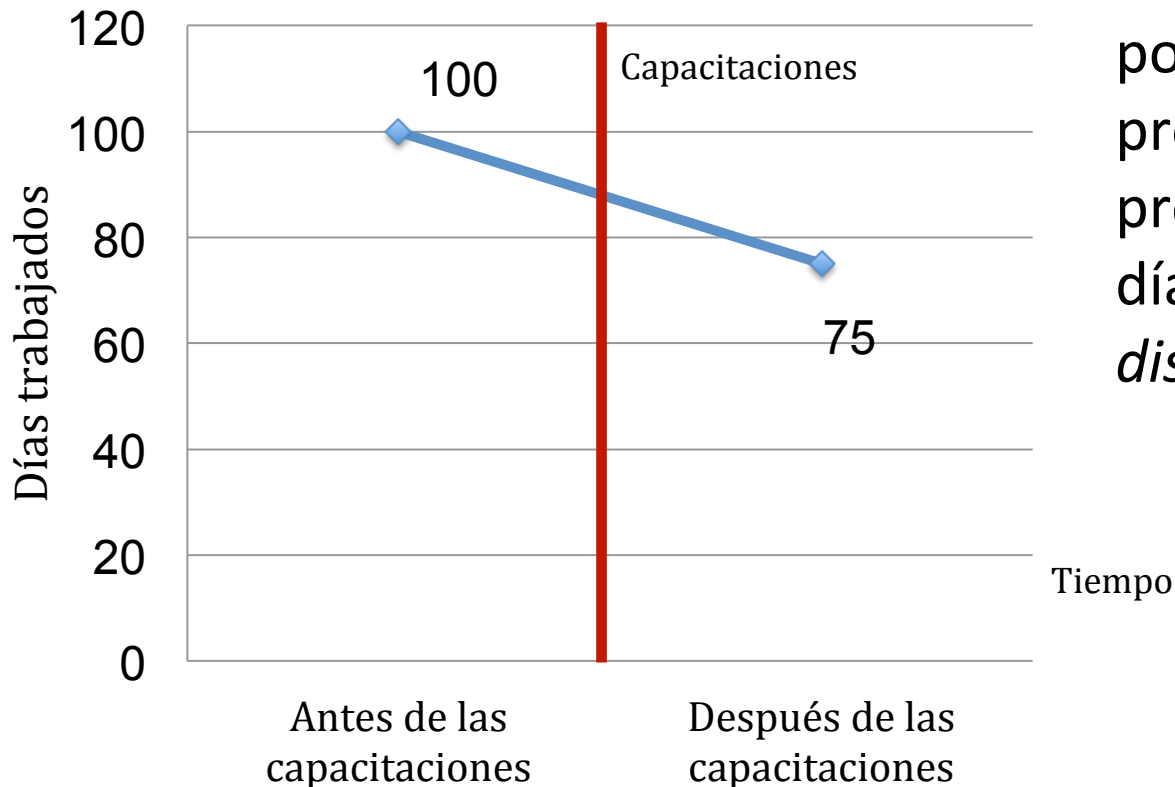
- Se implementa un programa de capacitación técnica para jóvenes que trabajan temporalmente.
- Queremos ver el **efecto de este programa** de capacitación **sobre la cantidad días** que los jóvenes trabajan en un año.



CONTRAFCTUAL FALSO 1: ANTES Y DESPUÉS

Observo los días trabajados:

Antes de las capacitaciones	Después de las capacitaciones
100	75



Con datos antes y después podemos concluir que el programa **causó** que en promedio el número de días trabajados *disminuyera* en 25.

¿Es correcto?



CONTRAFACTUAL FALSO 1: ANTES Y DESPUÉS

El gran problema es que **no se puede controlar por cambios en el tiempo.**

- Puede ser que durante el mismo periodo de evaluación, la economía entrara en recesión, agravando la situación laboral de los encuestados.
- Esto pudo haberles limitado la posibilidad de encontrar trabajo.



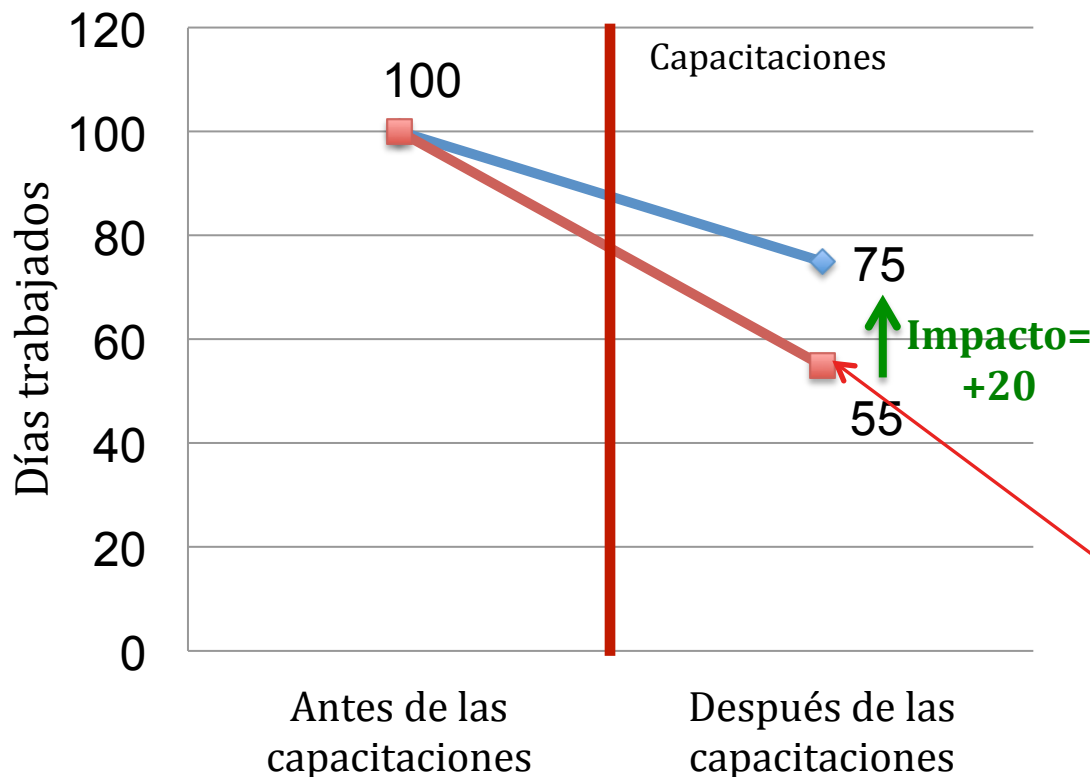
CONTRAFCTUAL FALSO 1: ANTES Y DESPUÉS

Observo los días trabajados:

Antes de las capacitaciones	Después de las capacitaciones
100	75

Lo que hubiera pasado:

Antes de las capacitaciones	Después de las capacitaciones
100	55



- En realidad el programa tuvo un impacto positivo.
- Pero sin considerar los cambios en el tiempo, parece que tuvo un impacto negativo.

Grupo de Comparación



CONTRAFACTUAL FALSO 1: ANTES Y DESPUÉS

Intuición:

- Observar una tendencia positiva no significa que el impacto haya sido positivo.
- Observar una tendencia negativa no significa que el impacto haya sido negativo.




MÉTODOS DE EVALUACIÓN DE IMPACTO

Métodos de evaluación experimentales

- Selección aleatoria del grupo de control
- Muy confiables si son bien implementados

Métodos no experimentales y cuasi experimentales

- Antes y después 
- Diferencia simple
- Diferencias en diferencias
- “Matching”
- Regresión discontinua



CONTRAFCTUAL FALSO 2: DIFERENCIA SIMPLE

Supongamos que se quiere ver el impacto de un programa de involucramiento parental sobre comportamientos antisociales de sus hijos.



- Se tiene información sobre **los que están inscritos y los que eran elegibles pero no se inscribieron** en el programa.
- Después de implementarse el programa, se observa que los niños del grupo de tratamiento tiene menos problemas de conducta que el grupo de comparación.



CONTRAFACTUAL FALSO 2: DIFERENCIA SIMPLE

	Con programa(T)	Sin programa(C)
Niños con problemas de conducta	125	200

¡El programa funciona!
¿O no?

- Con base en esta información, se concluiría que el programa tuvo impacto.
- Sin embargo, cuando se observan los datos, aparece que los grupos no son muy similares.



CONTRAFACTUAL FALSO 2: DIFERENCIA SIMPLE

	Con programa(T)	Sin programa(C)
Niños con problemas de conducta	125	200

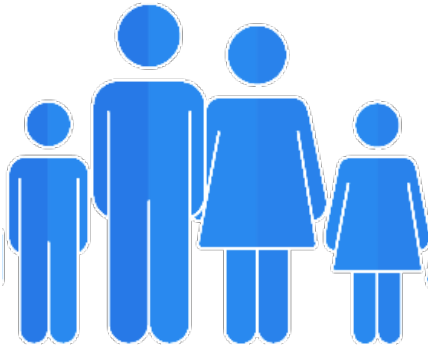
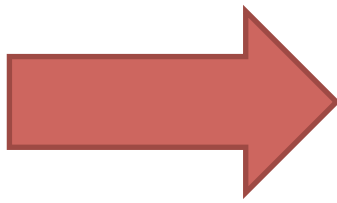
Características de los grupos:

	Con programa(T)	Sin programa(C)
Porcentaje madres solteras	20%	40%
Número promedio de hermanos	2.4	5.2

 **CONTRAFACTUAL FALSO 2: DIFERENCIA SIMPLE**



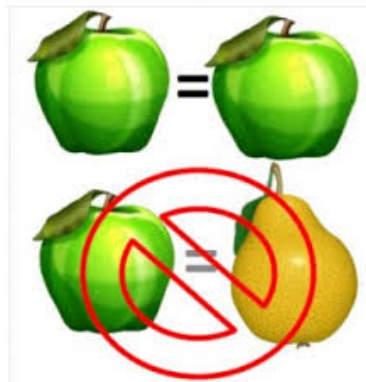
5 faltas de conducta al año



10 faltas de conducta al año

**Ya no
podemos
identificar el
impacto**

¡Estamos comparando peras con manzanas!





CONTRAFACTUAL FALSO 2: DIFERENCIA SIMPLE

- Al comparar inscritos al programa con no inscritos, es posible que sobre-estimemos el impacto del programa.
- Parece que algunas personas con características que mejoran su involucramiento parental se seleccionaron para entrar al programa.
 - Ej.: Los padres que no se inscriben en el programa pueden ser los que tienen menos motivación.



¿Qué sucede cuando las diferencias son **variables no observables**?



Métodos de evaluación experimentales

- Selección aleatoria del grupo de control
- Muy confiables si son bien implementados

Métodos no experimentales y cuasi experimentales

- Antes y después 
- Diferencia simple 
- Diferencias en diferencias
- “Matching”
- Regresión discontinua

3. ¿Por qué aleatorizar?



EL MÉTODO EXPERIMENTAL



POBLACIÓN
ELEGIBLE



GRUPO DE
TRATAMIENTO



GRUPO DE
COMPARACIÓN

¿POR QUÉ ALEATORIZAR?

Al tener una asignación aleatoria, los grupos no difieren de ninguna manera sistemática.

Estadísticamente **los grupos son iguales en promedio.**

GRUPO DE
TRATAMIENTO



GRUPO DE
COMPARACIÓN

Cualquier diferencia observable al finalizar el programa puede ser **atribuida** al programa.

¿POR QUÉ ALEATORIZAR?

- En la práctica, que los grupos sean estadísticamente idénticos significa que, **en promedio, no hay diferencias significativas** en sus características.
- Esperamos encontrar diferencias pequeñas, que no sean estadísticamente significativas a lo largo de la muestra.



EJEMPLO: PROGRAMA DE AUTOCONTROL JUVENIL*

Balance en características iniciales:

Variables	Grupo tratamiento	Grupo control	Diferencia
Edad	17.75	14.81	0.642
Nivel educativo	9.41	9.43	0.715
Mayor para su grado	34.64%	34.84%	0.654
Alguna vez arrestado	22.61%	22.93%	0.699

Esta tabla compara las características observables de los jóvenes que participaron en el programa (1,048) con los jóvenes que no participaron (1,016). Las diferencias no son estadísticamente significativas.

*(Become a Man Program)



VENTAJAS

Si son correctamente diseñadas e implementadas, las evaluaciones experimentales:

- **Eliminan cualquier sesgo** que puede presentarse en una evaluación no experimental.
- La aleatorización hace que los grupos sean **comparables en variables observables y no observables**.
- No es necesario hacer demasiados supuestos, como en las evaluaciones no experimentales.
- Los resultados son fáciles de interpretar.

¡Importante para evaluación de políticas públicas!





MÉTODOS DE EVALUACIÓN DE IMPACTO

Métodos de evaluación experimentales

- Selección aleatoria del grupo de control
- Muy confiables si son bien implementados

Métodos no experimentales y cuasi experimentales

- Antes y después 
- Diferencia simple 
- Diferencias en diferencias
- “Matching”
- Regresión discontinua



MÉTODOS CUASI EXPERIMENTALES

Para definir el grupo de control los métodos cuasi experimentales hacen supuestos.

¡Supuestos que muchas veces no se pueden verificar!

Diferencia-en-diferencias

- **Idea:** Combina el método de Antes y después con el método de Diferencia simple. Compara la variación en resultados a lo largo del tiempo de los participantes con la de los que no participaron del programa.
- **Grupo de Control:** Los individuos que no participaron en el programa, pero de quienes se recolectó información tanto antes como después del programa.
- **Supuesto:** Si el programa no existiera, los dos grupos tendrían idénticas trayectorias a lo largo de este periodo.



COMPARACIÓN DE ESTIMACIONES EN DIVERSOS PROGRAMAS

Conclusión: ¡En realidad no se sabe!

Diferentes métodos producen diferentes resultados

MÉTODO	ESTIMACIÓN DE IMPACTO		
	PROGRESA (México)	Balsakhi (India)	Get out to Vote (USA)
(1) Pre-Post	34.28 %***	26.42 %*	-7.2 pp
(2) Diferencia en diferencias	25.53 %**	6.82 %*	3.8 pp*
(3) Asignación aleatoria	29.75 %***	5.87 %*	0.4 pp



ENTONCES

- ¡El método importa!
- Según cada método se pueden tener distintos resultados.
- Algunos métodos nos permiten aproximarnos al verdadero impacto causal de un programa cuando no es posible aleatorizar.
- Lo importante es tener claro que detrás de cada método cuasi experimental, existen supuestos que pueden ser fuertes.






MÉTODOS DE EVALUACIÓN DE IMPACTO

Métodos de evaluación experimentales

- Selección aleatoria del grupo de control
- Muy confiables si son bien implementados

Métodos no experimentales y cuasi experimentales

- Antes y después 
- Diferencia simple 
- Diferencias en diferencias 
- Matching
- Variables instrumentales
- Regresión discontinua

4. Críticas a las evaluaciones aleatorias





PREGUNTA

¿Qué crítica has escuchado con más frecuencia acerca de las evaluaciones aleatorizadas?

- A. No son éticas
- B. Son demasiado caras
- C. Son demasiado largas
- D. Otra crítica
- E. No he escuchado críticas



RESPUESTAS A CRÍTICAS FRECUENTES

- Los experimentos aleatorizados son técnicamente superiores a los métodos no y cuasi-experimentales, y son más fáciles de entender.
- Pero ¿son **éticos**?
 - ¿son **viables**?
 - ¿son **generalizables**?



¿SON ÉTICOS?

- Hay muy pocos programas que pueden atender a todos a la vez.
 - La asignación aleatoria puede ser el modo más justo de asignar un recurso escaso.
- Se puede hacer una intervención en fases, donde todos reciben finalmente el programa.
 - No es necesario que el grupo de control no obtenga nada.
- También, podría ser que el impacto del programa sea negativo.
 - No es ético desperdiciar el dinero en programas que no funcionan.
- No es ético no aprender.

¿CUÁNDO ALEATORIZAR?

Existen múltiples situaciones en las cuales es oportuno y ético realizar evaluaciones aleatorias:

1. Cuando hay exceso o déficit de demanda.
2. Cuando un programa se está probando con un piloto.
3. Cuando un programa incorpora nuevos servicios, nuevos destinatarios o nuevos lugares.
4. Cuando un programa debe implementarse por fases.
5. Cuando un programa tiene un criterio de admisión numérico.

Conclusión: ¡existen numerosas oportunidades!



¿SON VIABLES POR SUS COSTOS?

- Lo caro es un buen levantamiento de datos y su rigor (tamaño muestral).
- Se pueden hacer evaluaciones aleatorizadas baratas.
 - Ventaja de disponer de buenos datos administrativos.
- ¿Caro comparado con qué?
 - Los programas no probados pueden resultar mucho más caros.
- Es mejor tener unos pocos estudios de buena calidad que muchos estudios baratos de mala calidad.
- Las lecciones extraídas son un bien público.
- Las evaluaciones experimentales pueden ayudarnos a identificar soluciones costo-efectivas.



¿SON VIABLES POR SU DURACIÓN?

- Depende de lo que queramos medir y no del tipo de evaluación.
- Ocurre lo mismo con otros métodos de evaluación de impacto: es necesario esperar que la intervención produzca sus resultados.
- En muchas evaluaciones, se calculan los resultados demasiado temprano, antes de que se haya producido el impacto. En consecuencia, se atribuye un impacto más bajo de lo que es realmente.



¿SON GENERALIZABLES?

- Son tan válidos externamente como otros métodos de evaluación de impacto.
- Es posible (y se debe) hacer réplicas.
- No es posible tener validez externa sin validez interna.

5. Pasos clave para su diseño





PASOS CLAVES

- 1. Definir** las preguntas de evaluación.
 - Definir el problema, revisar la literatura, identificar las brechas de conocimiento tanto desde el punto de vista académico como de la política pública.
- 2. Diseñar** la evaluación cuidadosamente.
 - Definir la población y sub-grupos de interés, el tratamiento, las variables a medir, etc.
- 3. Asignar aleatoriamente** las unidades a tratamiento o control.
- 4. Recolectar datos** de línea de base.
 - No es estrictamente necesario pero sí recomendable, ayuda a verificar balance, tamaño muestra, interacciones.
- 5. Verificar** que la asignación generó grupos similares.
 - Revisar balance



PASOS CLAVES

6. **Monitorear** el proceso para garantizar la integridad del experimento.
 - Que no se les dé el programa a los de control y sí se les dé a los de tratamiento.
7. **Recolectar datos de seguimiento** después de la intervención tanto para el grupo de tratamiento como para el grupo control.
8. **Estimar impacto** del programa, comparando los resultados promedio del grupo de tratamiento vs. los resultados promedio del grupo de control.
9. Determinar si el impacto (la diferencia entre los grupos) es **estadísticamente significativo**.
10. **Difundir los resultados** a los actores relevantes.



CONCLUSIONES GENERALES

- Existen **muchas formas** de estimar el impacto de un programa.
- En este taller, queremos transmitir que un método es superior a los demás:

los experimentos aleatorios

- Argumento conceptual: *Si son correctamente diseñados y realizados*, los experimentos aleatorios constituyen el método más confiable para estimar el impacto de un programa.
- Argumento empírico: Diferentes métodos pueden generar diferentes estimaciones del impacto.

Gracias!

Vivian Bronsoler

Gerente de Políticas

Abdul Latif Jameel Poverty Action Lab

TALLER PARA LA EVALUACIÓN DE PROYECTOS DE PREVENCIÓN SOCIAL DE LA VIOLENCIA Y LA DELINCUENCIA



SEGOB
SECRETARÍA DE GOBERNACIÓN



PROGRAMA PARA LA
CONVIVENCIA CIUDADANA

ABDUL LATIF JAMEEL
Poverty Action Lab
TRADUCIENDO LA INVESTIGACIÓN EN ACCIÓN