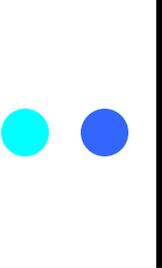




Evaluación de Impacto: Racionalidad y Métodos

Felipe Barrera-Osorio





Porque es importante la evaluación de impacto?

1. La evaluación de impacto puede producir estimadores confiables de los efectos causales de programas
2. La evaluación de impacto puede ayudar a mejorar el diseño e implementación de programas
3. La evaluación de impacto puede aumentar el apoyo político de programas
4. La evaluación de impacto puede aumentar la sostenibilidad de programas exitosos y terminar los no exitosos
5. La evaluación de impacto puede ayudar a entender como funcionan internamente los programas sociales

● ● | Que es la evaluación de impacto?

1. La evaluación de impacto es un grupo de métodos para identificar y medir el efecto *causal* de programas. Por ejemplo
 - A. Cual es el efecto de un programa de becas en el rendimiento de estudiantes en exámenes?
 - B. Cual es el efecto de programas de entrenamiento para el empleo en la probabilidad de empleo y salarios?
 - C. Cual es el efecto de la provisión de computadores en la escuela sobre la asistencia estudiantil?

● ● | Por qué esto es difícil de medir?

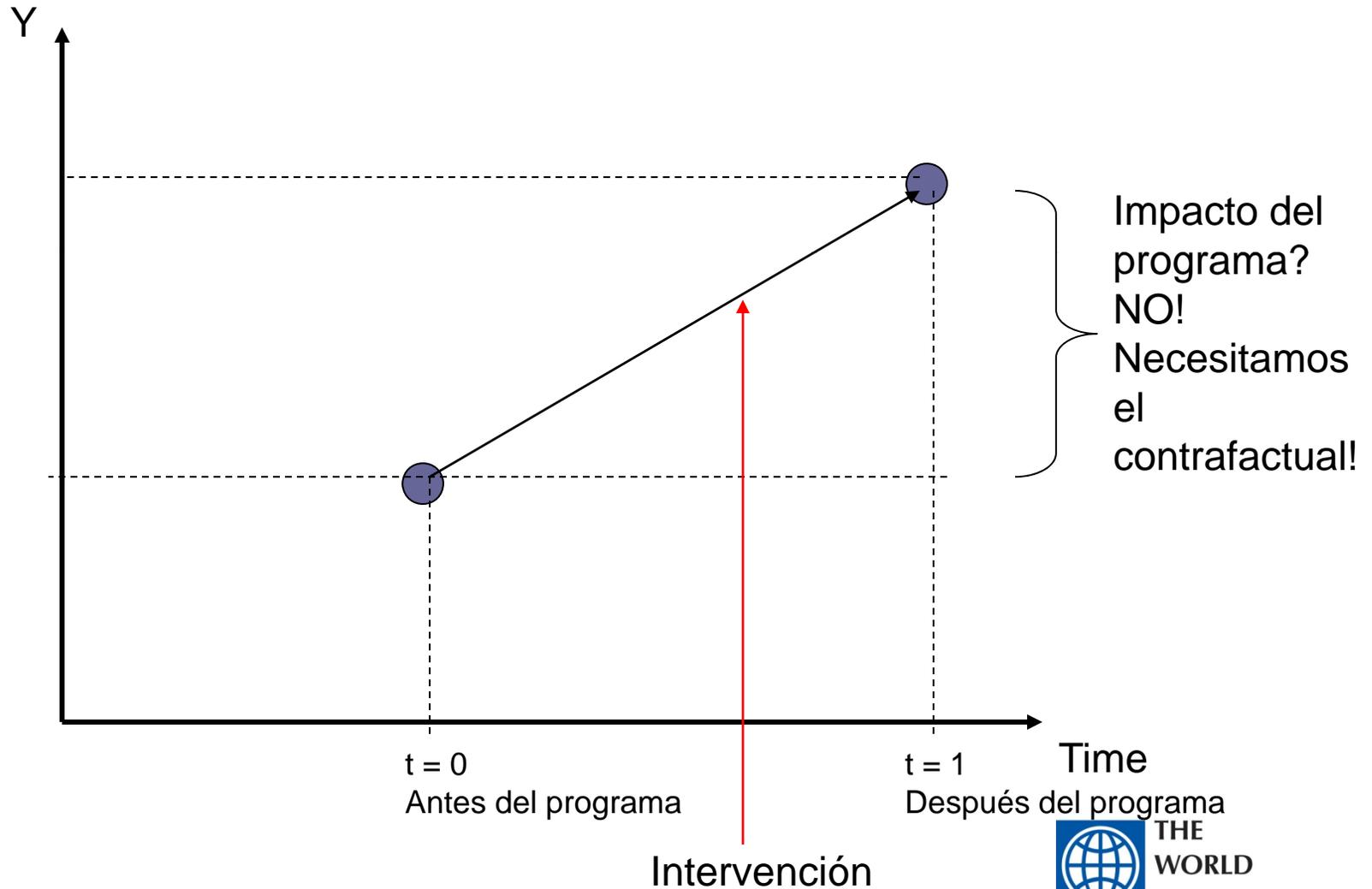
1. La pregunta básica es la siguiente: *Que hubiera pasado en ausencia del programa* (en términos de los economistas, cual es el contrafactual?)
2. Necesitaríamos observar al *mismo* individuo, sin y con el programa
3. ...pero es imposible observar al mismo individuo en los dos estados (sin y con programa...)
4. La solución es construir el contrafactual correcto
Encontrar individuos que no se han beneficiado del programa, pero que tienen características muy similares a los que sí han tenido acceso al programa

● ● |Cuál es el problema?

1. El problema es encontrar el correcto grupo de comparación:
 - A. El grupo de comparación y los beneficiarios tienen que tener las mismas características (observables y no observables) antes de comenzar el programa
 - B. Los factores externos tienen que afectar de la misma forma al grupo de comparación y el del programa
 - C. Los efectos del programa deberían ser los mismos para ambos grupos
2. El problema crítico: las personas se auto-seleccionan cuando aplican y reciben un programas, y por consiguiente, con una probabilidad alta son diferentes a las que no aplicaron o no recibieron el programa

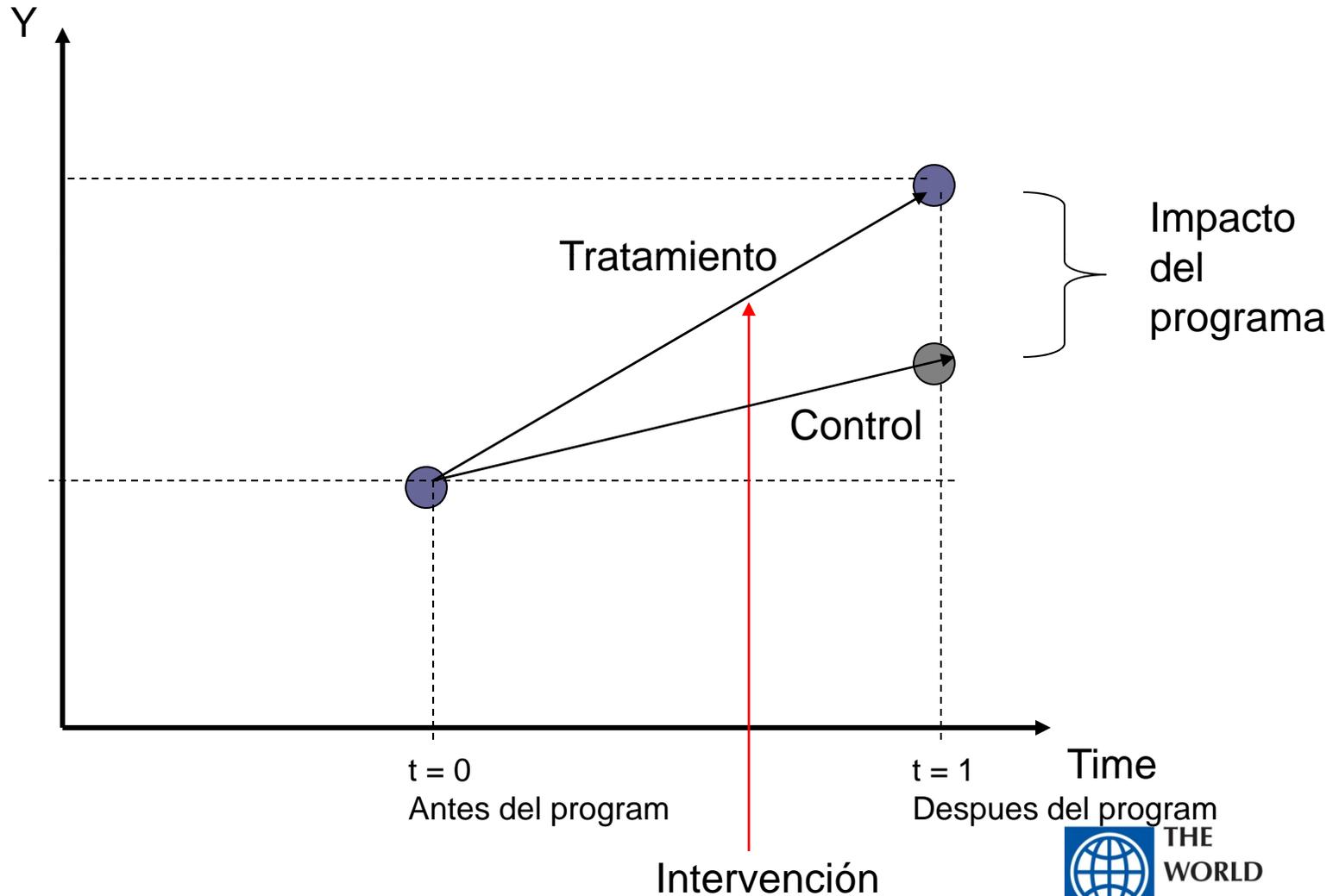
La intuición básica.

Primer caso: solo datos para los tratados, antes y después del programa



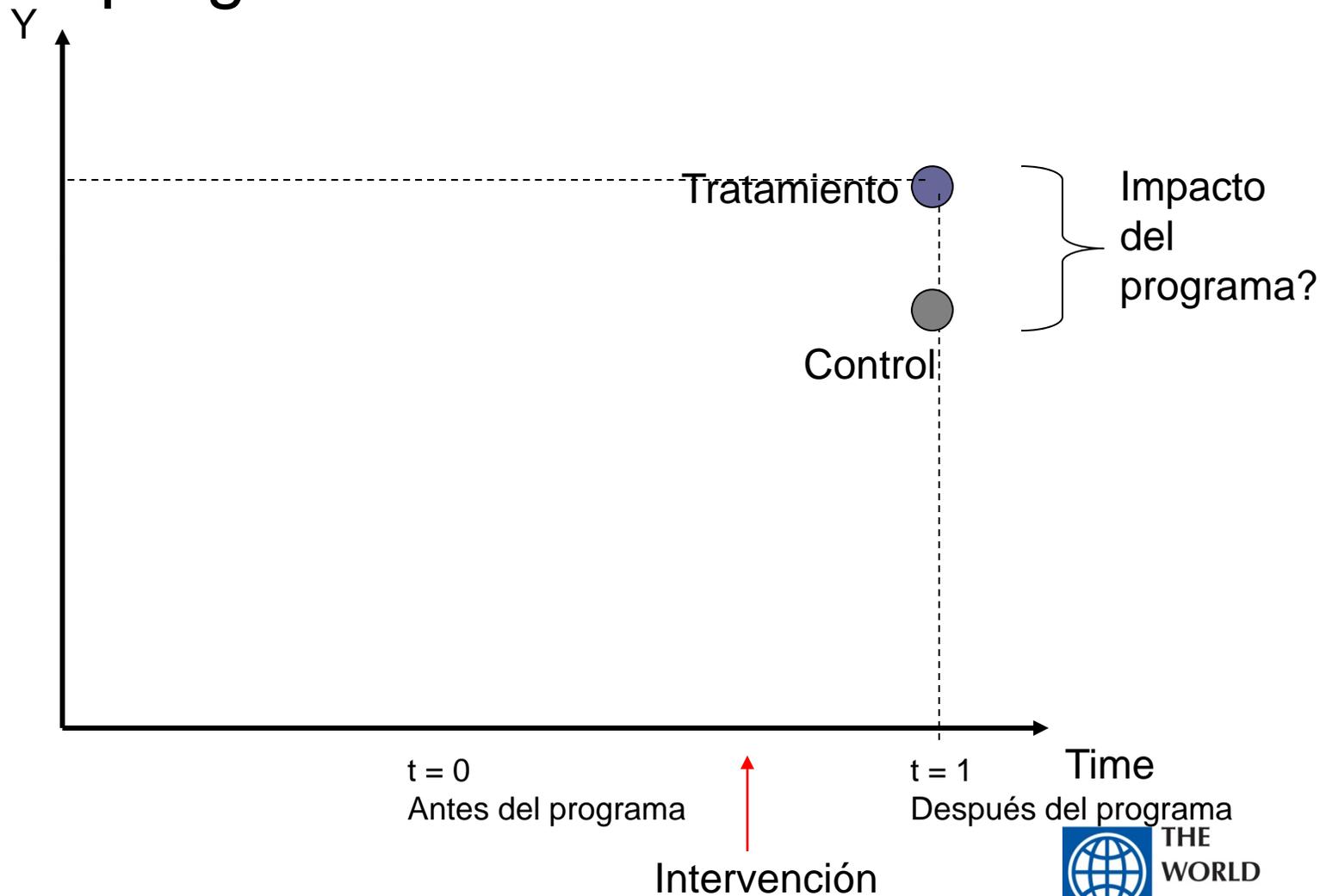


Es necesario tener el grupo de comparación



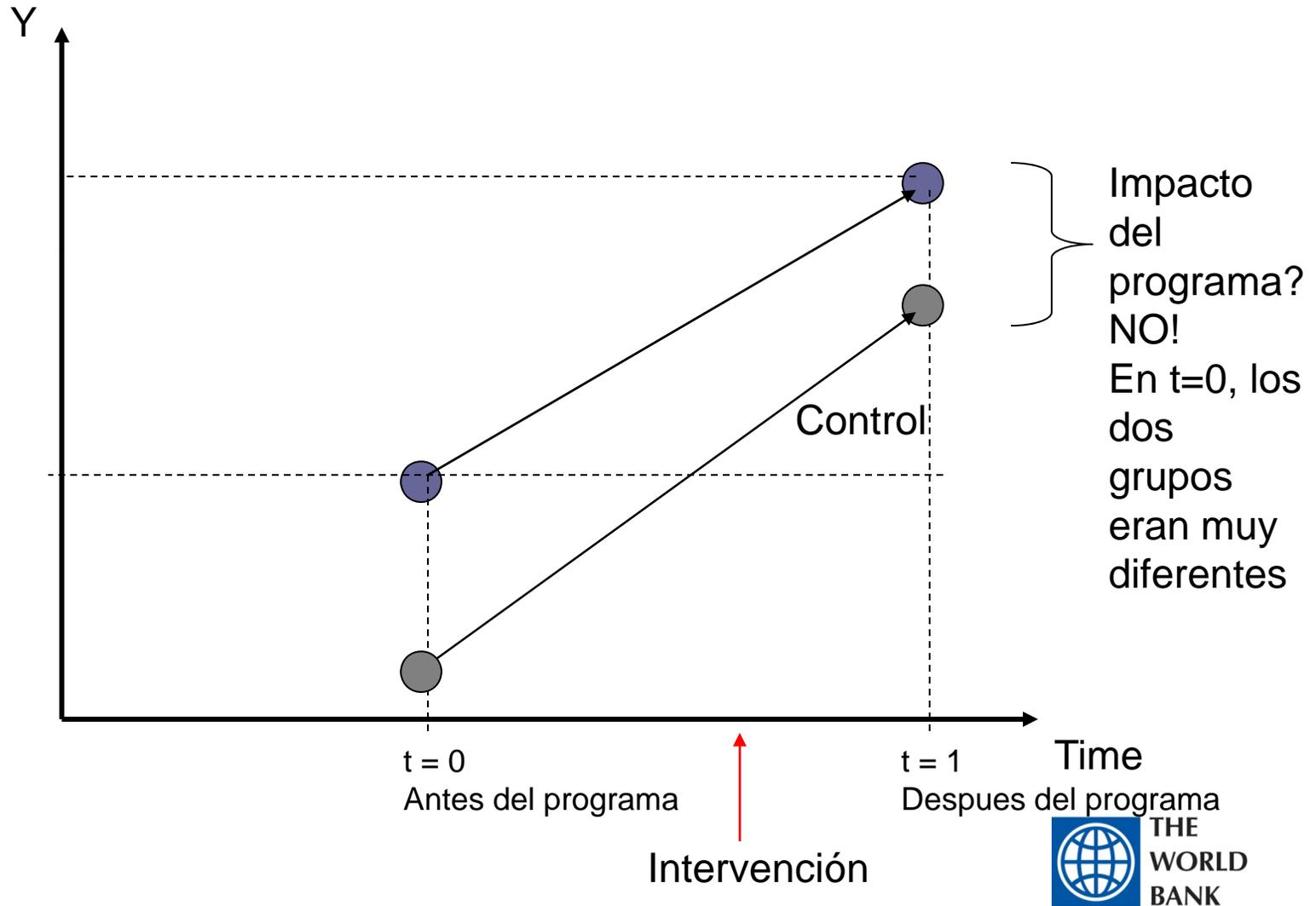
La intuición básica.

Segundo caso: solo datos después del programa





Necesitamos el grupo de comparación correcto!





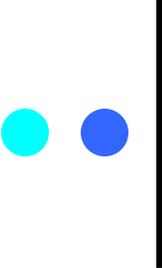
Posibilidades para encontrar el grupo de comparación

1. Evaluaciones prospectivas
 - A. Aleatorización de beneficiarios
 - B. Aleatorización de entrada
 - C. Aleatorización de información
2. Evaluaciones retrospectivas
 - A. Regresión de discontinuidad
 - B. Variables instrumentales
 - C. Diferencias en diferencias
 - D. Estimación de pareo



Aleatorización

1. Una lotería entre individuos separa la población entre beneficiarios y no beneficiarios
 - A. Modelo de sobre-suscripción al programa
2. La lotería crearía grupos homogéneos entre beneficiarios y no beneficiarios: tendrían las mismas características, observables y no observables
3. La “unidad” de aleatorización es importante:
 - A. Unidades geográficas (estados, distritos, municipios, etc) versus unidades individuales (personas, hogares, colegios)
 - B. El numero de observaciones es importante: si es muy pequeño, la muestra puede no estar balanceada.



Ejemplo de aleatorización: el programa PACES en Colombia

1. Programa otorgó 125,000 vouchers entre 1992 y 1997 a estudiantes al inicio de secundaria
2. Los vouchers fueron ofrecidos a familias de bajos ingresos que atendían colegios públicos y que habían sido aceptadas en colegios privados
3. Los vouchers eran renovables cuando el rendimiento académico había sido satisfactorio
4. El voucher tenía un valor de \$US190 – mitad del costo de colegios privados
5. Los colegios recibieron los fondos directamente en el banco
6. Los colegios podían pedir dinero complementario al voucher
7. El voucher fue ofrecido por medio de loterías cuando existió sobre-subscripción al programa

Impacto del programa PACES

	Promedio	Diferencia	
	Perdedores	con Beneficiarios	
Ultimo grado aprobado	7.5	0.167	
	(0.965)	(0.053)	*
En colegio	0.831	0.021	
	(0.375)	(0.021)	
Repitió un grado desde lotería	0.254	-0.077	
	(0.508)	(0.029)	*
Finalizo grado 8	0.632	0.114	
	(0.483)	(0.028)	*
Resultados en pruebas	-0.099	0.379	
	(1)	(0.111)	*
Casado o viviendo con compañera(o)	0.016	-0.009	
	(0.126)	(0.006)	***

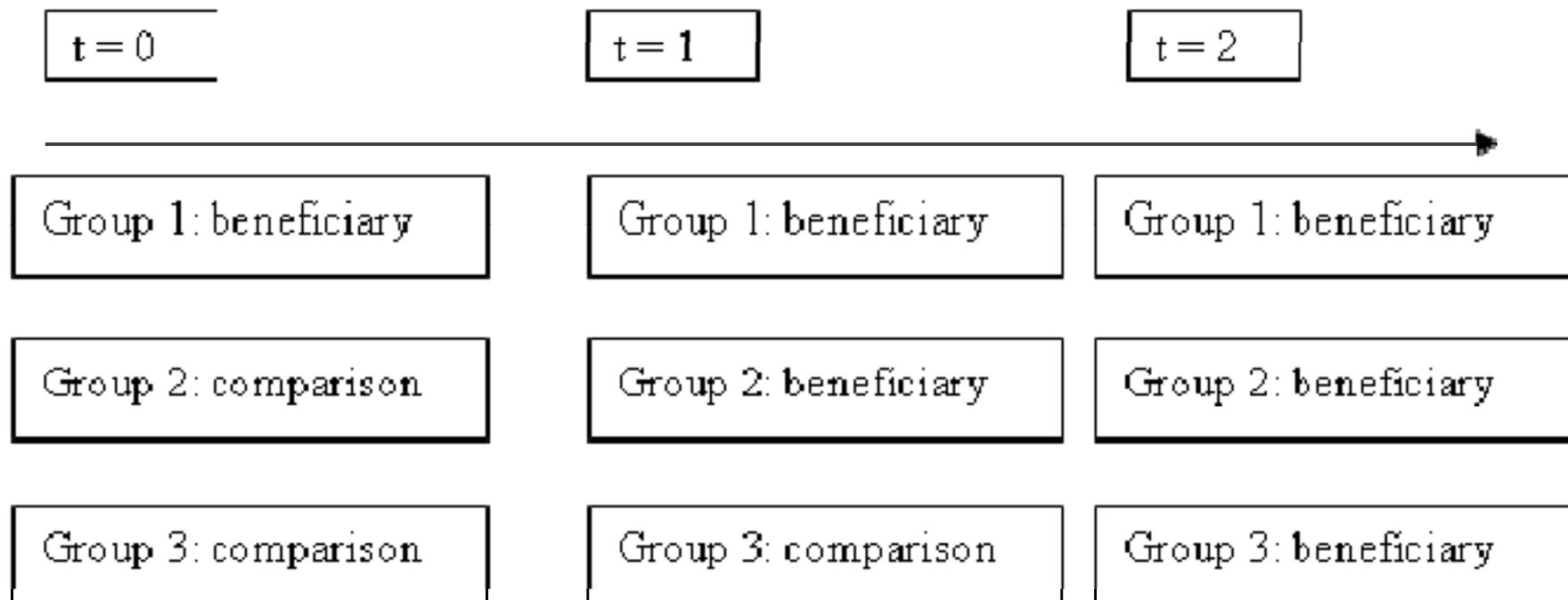
● ● | Aleatorización de entrada al programa

1. Cuando se aplica este método?
 - A. Cuando el gobierno / instituciones está planeando expansiones de cobertura en el tiempo
 - B. Es posible aleatorizar el orden de entrada
 - C. Además, es posible medir diferente intensidad de tratamiento: los que entran primero están “expuestos” por más tiempo al programa que los que entran más tarde
 - D. Usualmente, la aleatorización se realiza a un nivel geográfico



Diseño de aleatorización de entrada

Time:



Time: at t=0, first group receive benefits; at t=1 second group enters...and so on



Aleatorización de información

1. Es posible aleatorizar la información sobre el programa
 - A. Un grupo recibe información; el otro no
 - B. Cualquier individuo puede aplicar al programa (los que recibieron la información y los que no la recibieron)
2. La idea básica es crear una variable (la campaña de información), correlacionada con recibir los beneficios del programa, pero que no está correlacionada con características de los individuos
 - A. Esto es, crear una variable instrumental correcta
3. Lo que se asume es que el grupo que recibe la información tiene una mayor probabilidad de recibir beneficios

Regresión de discontinuidad

1. Cuando se usa este método?
 - A. Los beneficiarios y no beneficiarios de un programa pueden ser ordenados usando una variable cuantificable (por ejemplo, el SISBEN)
 - B. El programa determina que un individuo / hogar / etc es elegible usando un punto de corte en el índice
 - C. El valor del índice es lo que define la asignación a los beneficios del programa

- ● | Índices son utilizados regularmente para focalizar programas sociales

Ejemplos:

1. Programas anti-pobreza

→ focalizados a hogares debajo de un índice de pobreza

2. Programas pensionales

→ focalizados a poblaciones por encima de cierta edad

3. Becas

→ focalizados a estudiantes con altos resultados en pruebas académicas

● ● Intuición: un ejemplo hipotético

1. Método:

- A. Construir un índice de pobreza de 1 a 100 basado en las características de los hogares antes de la intervención
- B. Los hogares con un puntaje menor a 50 puntos son pobres
- C. Los hogares con un puntaje por encima de 50 puntos no son pobres

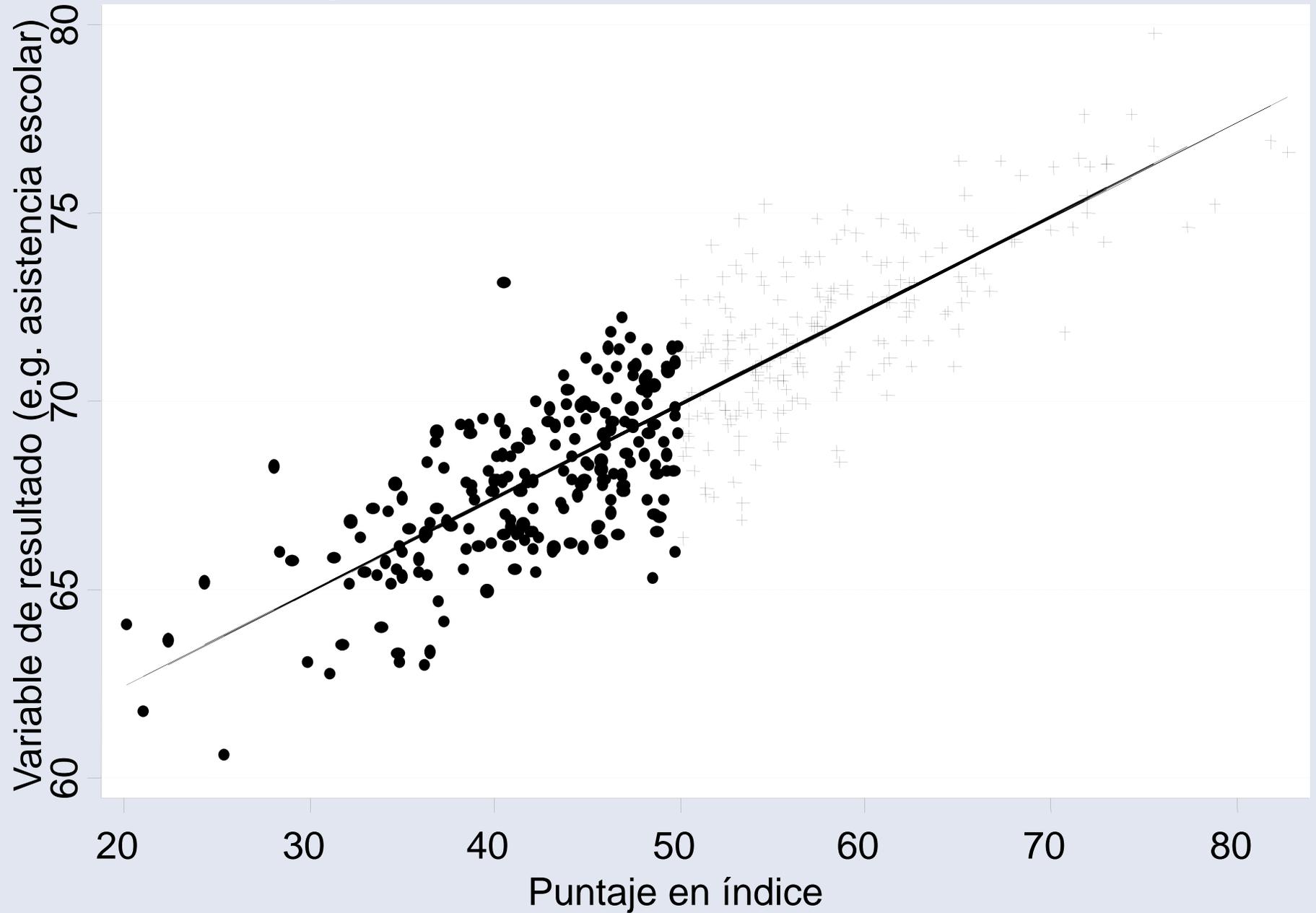
2. Implementación de un programa:

- A. Transferencia de efectivo a los pobres

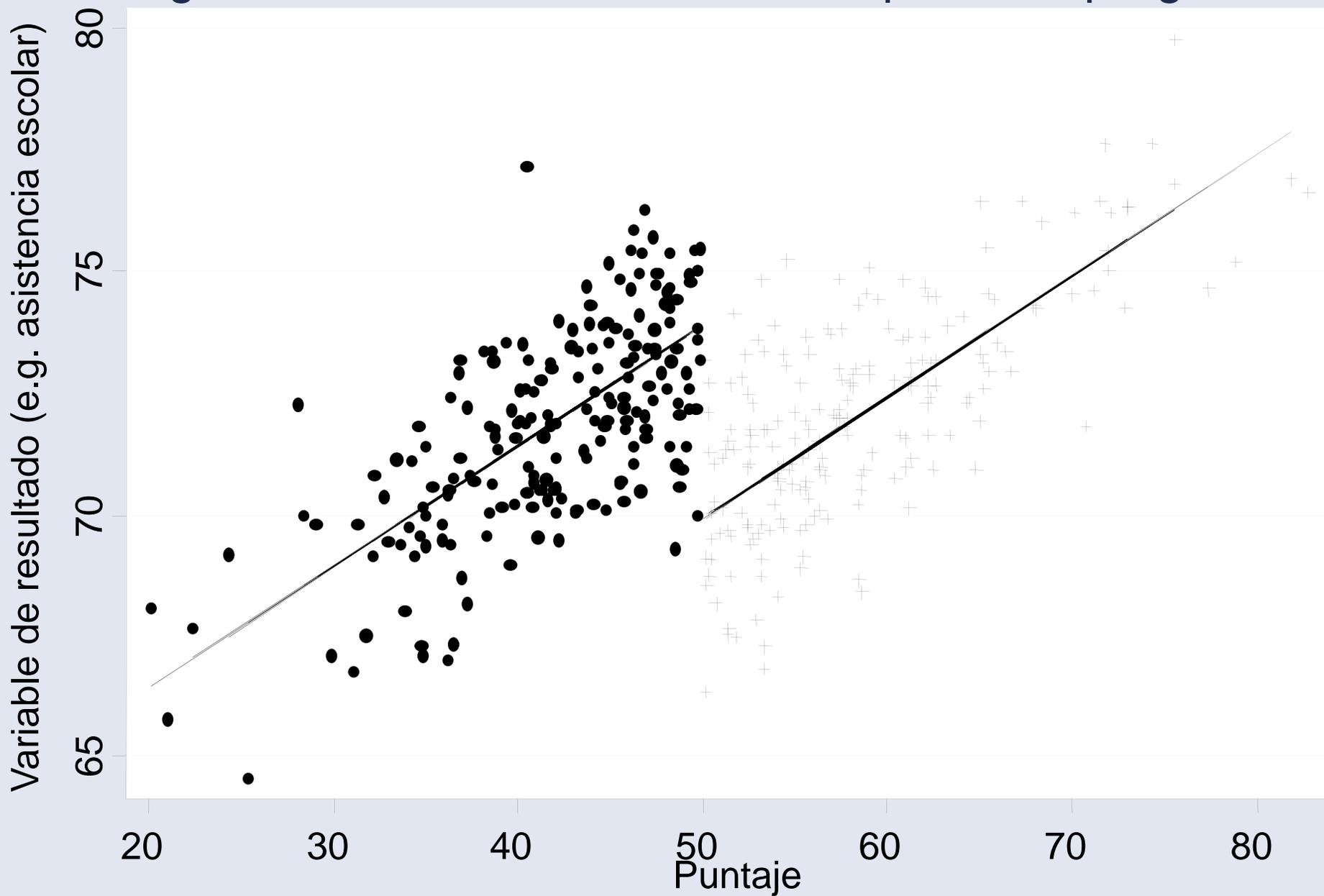
3. Evaluación:

- A. Medir resultados (por ejemplo, en consumo o asistencia escolar) después del programa en hogares pobres cercanos a los 50 puntos contra hogares no pobres cercanos a los 50 puntos

Regresión de Discontinuidad – Línea de Base



Regresión de Discontinuidad – Después del programa



● ● | Intuición

Los beneficiarios (los cuales están por encima del punto de corte) cercanos al punto de corte son muy similares a los no beneficiarios (los cuales están por debajo del punto de corte) que se encuentran cercanos al punto de corte



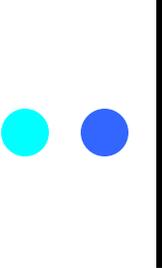
Problemas con Regresión de Discontinuidad

1. La estimación es “local”
 - A. No es un problema necesariamente: algunas veces el individuo que más nos interesa es el marginal
2. Es necesario tener un número “alto” de observaciones
 - A. De lo contrario, la estimación es muy imprecisa



Regresión de Discontinuidad: un ejemplo para Colombia

1. Los efectos de Reducciones en Costos Educativos sobre asistencia escolar (Barrera-Osorio, Linden y Urquiola, 2006)



El programa de Gratuidad en la ciudad de Bogotá en el 2005

1. Cada años los gobiernos locales establecen una resolución que estipula el listado de cobros y la respectiva tarifa máxima
2. Estos costos pueden representar entre 7 y 29 dólares mensuales—aproximadamente entre 6 y 25 por ciento del salario mínimo.
3. El programa de Gratuidad reduce algunos de estos cobros
4. El programa esta focalizado por medio del SISBEN (el cual identifica población vulnerable).
 - A. Basado en encuestas de hogares, cada hogar recibe un puntaje entre 0 y 100 puntos.
 - B. Basado en este puntaje, los hogares son clasificados en seis niveles (nivel 1 es el mas pobre, nivel 6 el mas rico)
 - C. Hogares con un puntaje por debajo de 11 puntos están en Nivel 1, los que tiene puntaje entre 11 y 22 en Nivel 2, entre 22 y 43 en Nivel 3, etc

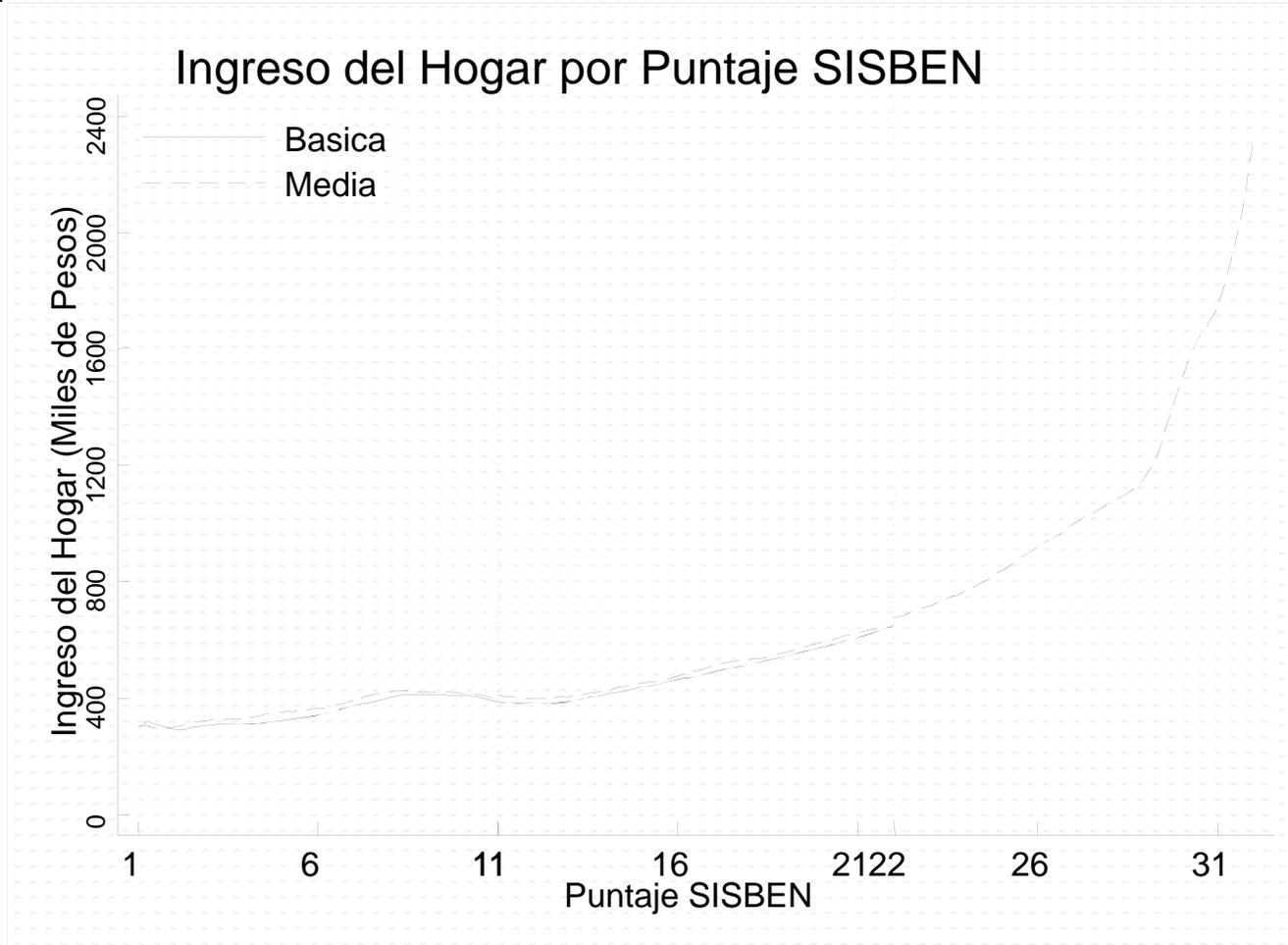


El programa (continuación)

1. Para educación básica (grados 1-9), Nivel 1 de Sisben reciben una reducción del 100% en servicios complementarios, mientras que Sisben 2 y encima no reciben reducciones
2. Para grados 10 y 11, Sisben 1 reciben la eliminación de cobros de servicios académicos y complementarios, mientras que Sisben 2 recibe aproximadamente una reducción del 50%
3. Hogares en Niveles 3 y mayores, no reciben beneficios.

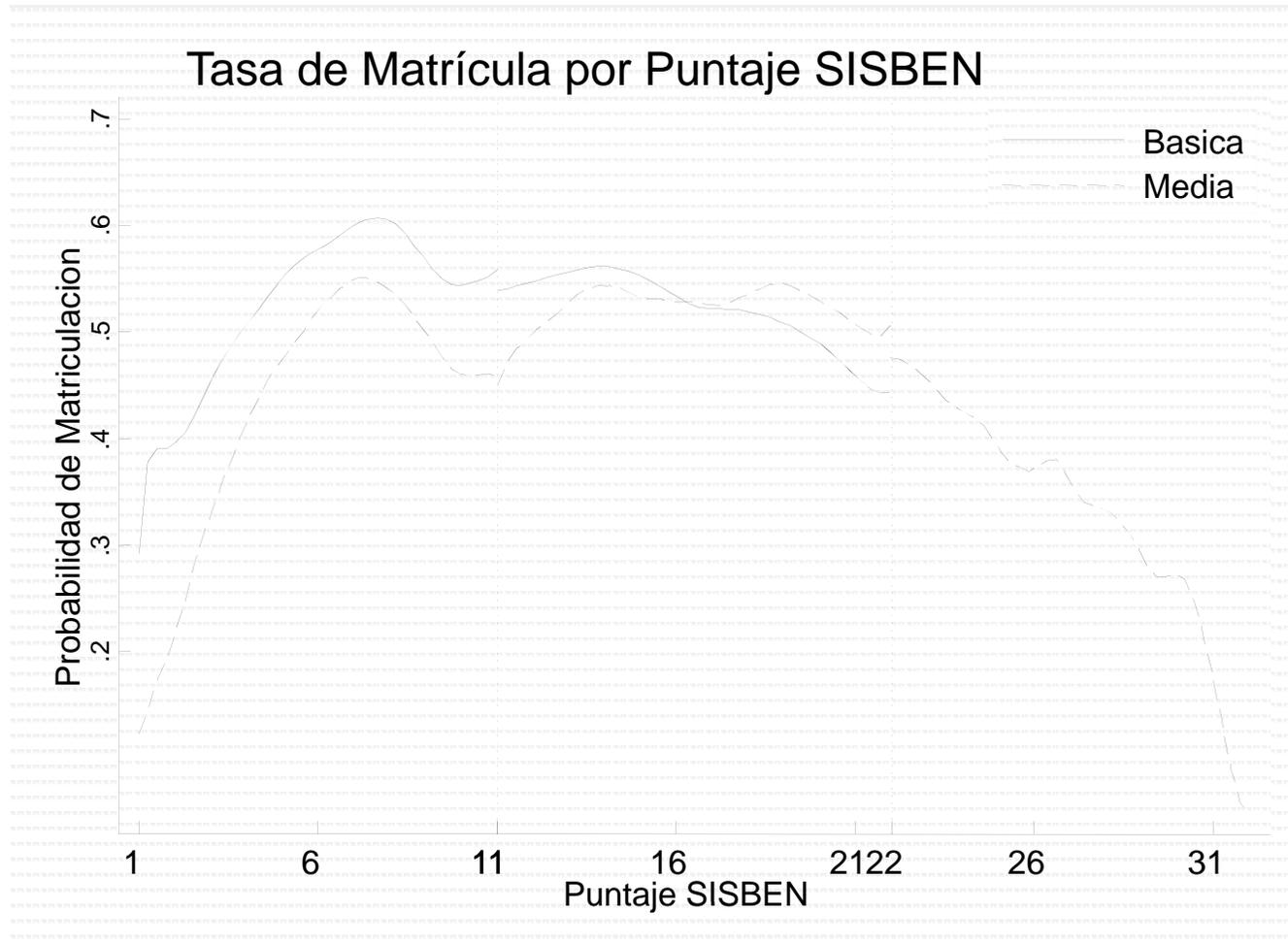


Son similares los grupos beneficiarios y de comparacion? Ingreso





Resultado principal





Diferencias en Diferencias

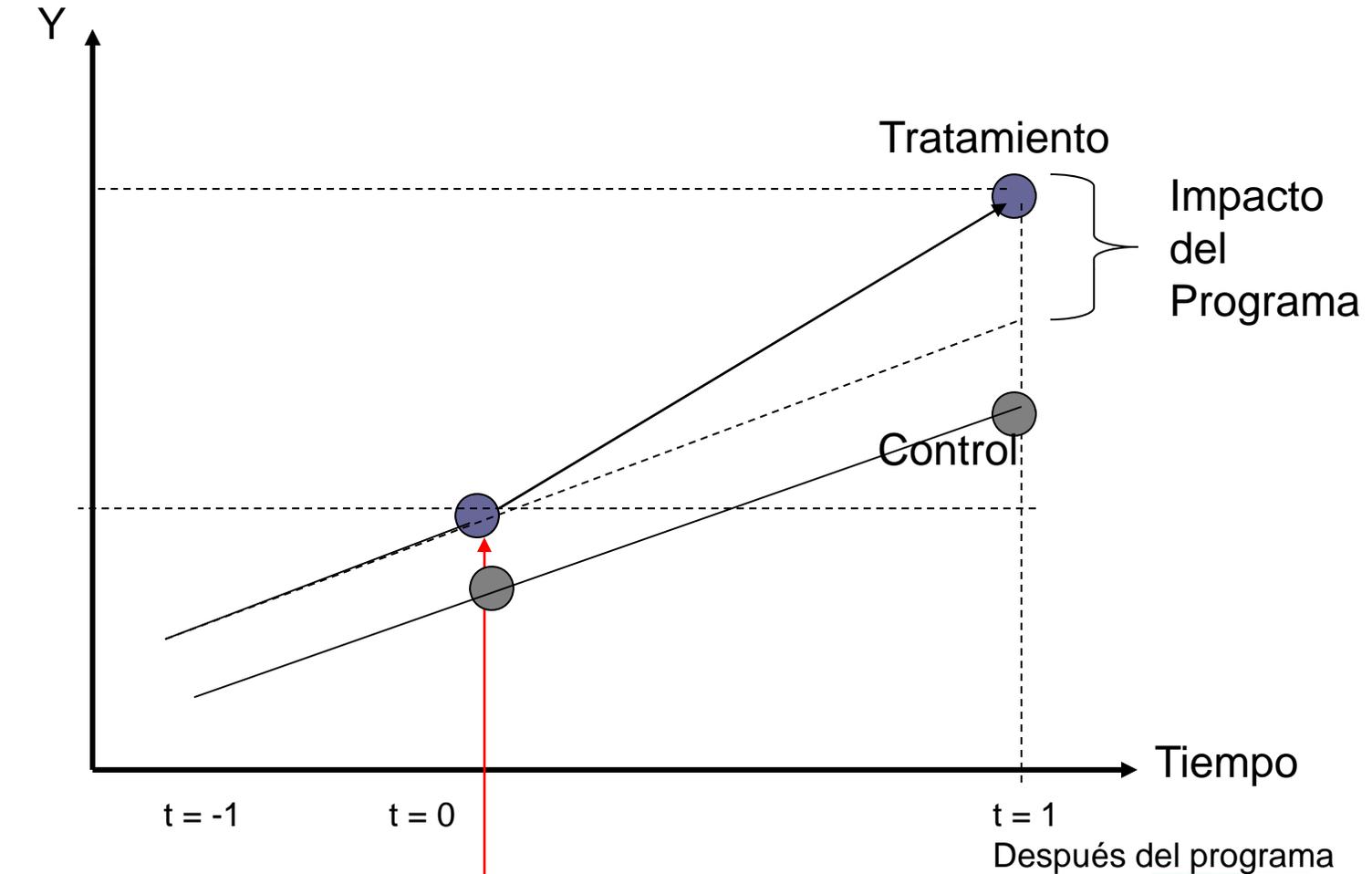
1. Es un tipo de evaluación retrospectiva
2. Es posible encontrar un “experimento natural” que ayude a identificar el impacto de una política
 - Por ejemplo, un cambio inesperado de una política determinada
 - Por ejemplo, una política que afecte solo a las personas de 16 años y no a los de 15
3. La parte crítica es identificar el grupo de comparación que es similar al tratado
4. Las características del este grupo deben ser lo mas similares al tratado para que la evaluación sea de buena calidad.

● ● | Intuición

1. Encontrar un grupo de comparación con el mismo patrón de crecimiento en la variable de interés *antes* del programa
 - A. Los dos grupos (beneficiarios y de comparación) deben tener el mismo perfil antes de comenzar el programa



Intuición: cuando es correcto usar DD



Las pendientes antes del programa son iguales

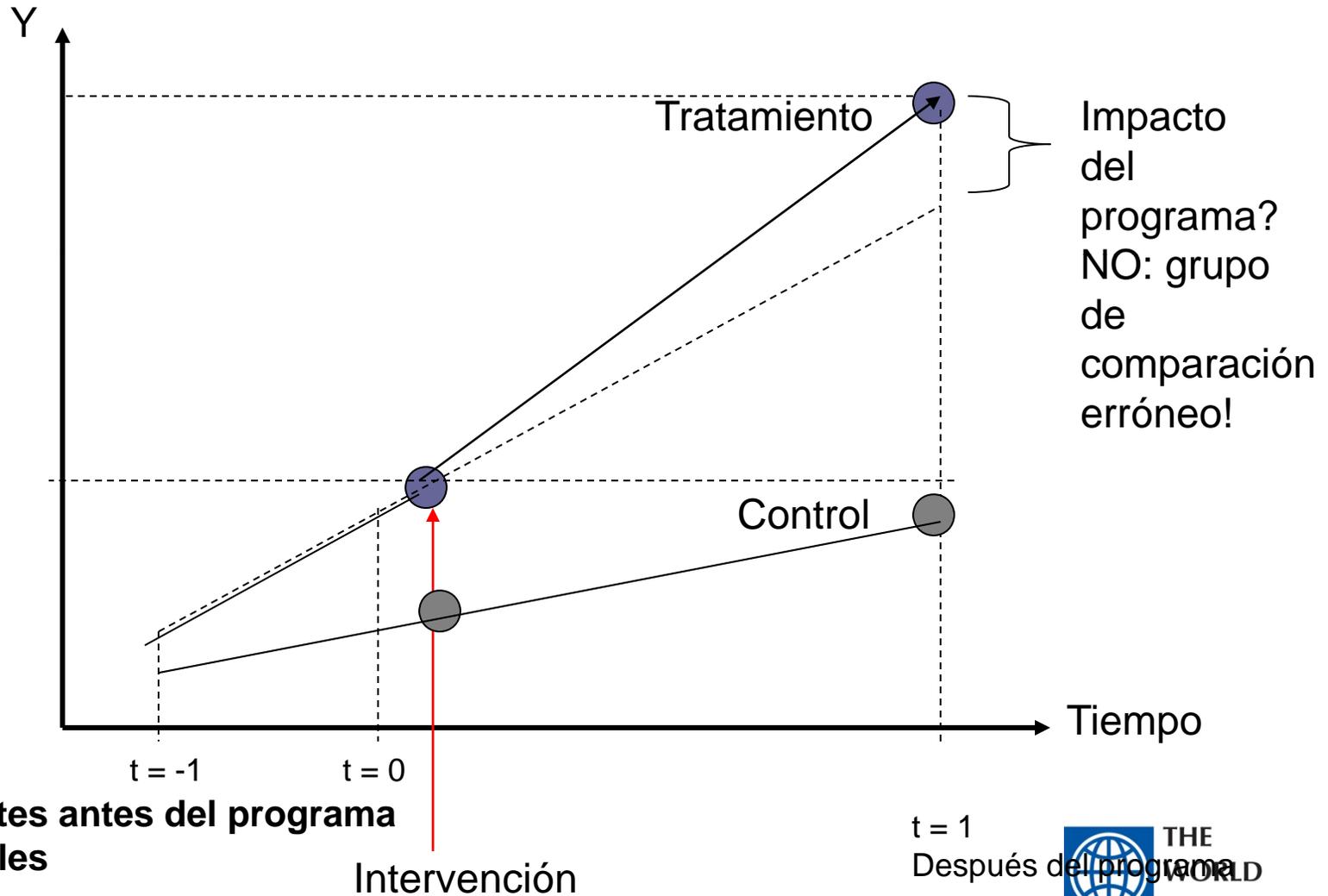
Intervención

Después del programa





Intuición: cuando es erróneo usar DD



Las pendientes antes del programa
NO son iguales

t = 1
Después del programa

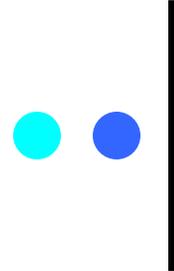


THE
WORLD
BANK



DD: Potenciales problemas

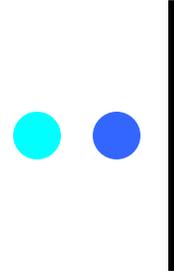
1. El método controla por efectos “fijos” en el tiempo
 1. Efectos que son fijos en un momento del tiempo
 - A. Si los factores externos afectan de forma diferente a los grupos de comparación y control, entonces la estimación es sesgada
 - B. Si existen características no-observables que cambian en el tiempo, la estimación va a ser sesgada



DD: un ejemplo

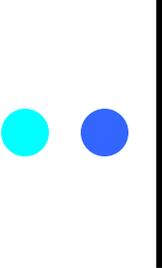
1. Evidencia de impacto de una construcción masiva de colegios en Indonesia sobre mercado laboral

(Articulo: Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment (E. Duflo) American Economic Review Sept 2001)



Descripción del programa

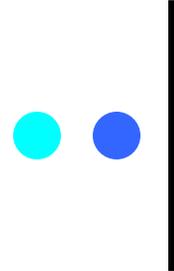
- Entre 1973 y 1978 el gobierno de Indonesia construyó 61,000 colegios (equivalente a un colegio por cada 500 niños entre los 5 y los 14 años de edad)
- La tasa de matrícula (asistencia) incrementó de 69% a 85% entre 1973 y 1978
- El número de colegios construidos dependió del número de niños por fuera del sistema en 1972



Identificación de los efectos

Existen dos fuentes de variación

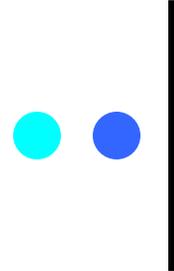
- A. Por región: cada región recibió un número diferente de colegios
- B. Por edad:
 - a. Niños que eran mayores de 12 años en 1972 no se beneficiaron del programa
 - b. Niños que eran menores de 12 años se beneficiaron del programa
 - Entre más joven, más beneficios recibieron



Estimación del impacto

1. Dos simplificaciones

- A. Intensidad por región: alta o baja
- B. Grupos de niños afectados: la cohorte joven (beneficiarios) y no joven (comparación)



Resultados

Efectos del programa: Años de Educación

		Intensidad del programa		
		Alta	Baja	Diferencia
Tratados:	Edad 2-6 en 1974	8.49 (0.043)	9.76 (0.037)	-1.27 (0.057)
Control:	Edad 12-17 in 1974	8.02 (0.053)	9.4 (0.042)	-1.39 (0.067)
Diferencia		0.47 (0.07)	0.36 (0.038)	0.12 (0.089)



Placebo: No deberíamos encontrar efectos

Efectos del programa: Años de Educación

		Intensidad del programa		
		Alta	Baja	Diferencia
Control 1:	Edad 12-17 in 1974	8.02 (0.053)	9.4 (0.042)	-1.39 (0.067)
Control 2:	Edad 18-24 en 1974	7.7 (0.059)	9.12 (0.044)	-1.42 (0.072)
Diferencia		0.32 (0.08)	0.28 (0.061)	0.034 (0.098)

● ● | Metodos de Pareo (Propensity and matching estimation)

1. Toma el grupo de comparación de una encuesta amplia
2. Cada beneficiario del programa tiene individuos de comparación con características observables muy similares a los tratados
3. El método asume que no existe auto-selección basado en características no observables, y que la selección se realiza por medio de características observables

Como se hace este procedimiento?

Dos pasos

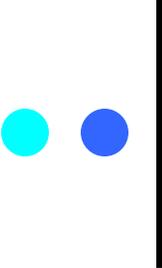
1. Estimación del Puntaje de ser “propenso” a ser tratado (Puntaje de “Propensidad”—“Propensity score”): El puntaje de “propensidad” es la probabilidad de recibir el programa dado las variables observadas antes de del programa
 - A. Estimar la probabilidad de ser tratado basado en las características medidas en la encuesta
2. Estimación del efecto promedio dado el puntaje de “propensidad”
 - A. Para cada tratado identificar controle(s) que tengan el mismo (o cercano) puntaje de “propensidad”;
 - B. Computar el efecto de tratamiento para cada unidad
 - C. Obtener el promedio para todos los tratados

● ● Problemas con Métodos de Pareo

1. El problema mas importante es asumir que las variables no observables no juegan un papel importante en ser beneficiario de un programa

● ● | Comentarios finales

1. Qué datos se necesitan para cada uno de estos métodos?
2. Cuando aplicar cada método?



Qué datos se necesitan para cada uno de estos métodos?

1. Aleatorización: idealmente, tener una línea de base (pero no es necesario); tener datos de seguimiento, y construir los datos de forma simultánea a la implementación del programa
2. RD: datos de la variable de asignación y datos de resultados; idealmente tener una línea de base
3. IV: tener un instrumento válido
4. DD: datos en el tiempo *antes* y después del programa; series largas de tiempo
5. Estimación de Pareo: datos detallados de línea de base y de seguimiento
 - A. Los requerimientos de datos aumentan entre 1 a 5; DD y Métodos de Pareo requieren muchos datos de alta calidad

● ● Cuando aplicar cada método?

1. Idealmente, aleatorización es la mejor opción
 - A. Aleatorización individual o geográfica; hacer programas pilotos antes de expandir
 - B. Aleatorización de entrada: programa es universal y el programa se hace a pasos
 - ayuda en la administración del programa
2. RD
 - A. El programa es focalizado utilizando un índice
3. DD y Pareo:
 - A. Programa no es universal, y se cuentan con datos de alta calidad antes y después del programa