

# Taller II

RCT

---



Germán C. Rodríguez <sup>1</sup>

25 de septiembre de 2020

<sup>1</sup>Facultad de ciencias económicas  
Evaluación de Impacto

# RCT y datos disponibles

Un Randomized controlled trial (RCT) es un *diseño de investigación* muy empleado en la *evaluación de impacto causal*. En un RCT el investigador tiene *control* sobre el mecanismo de asignación y la asignación a tratamiento está basada en un mecanismo de *aleatorización*[1].

Ventajas de la aleatorización en un RCT:

- El sesgo de selección desaparece  $E[Y_i^0 | D_i = 1] = E[Y_i^0 | D_i = 0]$
- Los parámetros causales:  $ATE = ATT = ATE$
- La simple diferencia de medias se puede interpretar causalmente:  $SDO = ATE$

Un modelo base para la realización de un RCT es el de la *switching regression*:  $Y_i = \alpha + \beta D_i + \varepsilon_i$ <sup>1</sup>

Datos disponibles: **tamano\_hogar**: para el tamaño del hogar, **rural**: zona rural o casco urbano, **escuela**: escuela que frecuentan los hijos, **mujer\_cabeza**: sobre la cabeza de hogar, **ingreso\_hogar**: ingreso mensual, **treatment**: si los niños fueron tratados mediante píldoras, **outcome**: índice de estado de salud, **nota\_hijos**: promedio del desempeño escolar y **hogar**: denota el hogar.

---

<sup>1</sup>Para que dicho modelo sea adecuado se necesita que se cumpla los supuestos del SUTVA

## Cuadro 1: Estadística descriptiva que muestra posible asignación al tratamiento de carácter aleatoria

treatment	rural		mujer_cabeza		tamaño_hogar					
	0	1	0	1	4	5	6	7	8	9
0	47	45	40	52	8	12	24	17	15	16
1	50	58	56	52	20	20	16	19	23	10

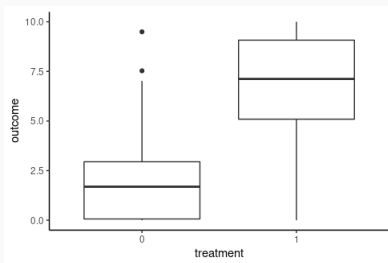
La primera tabla de contingencia es la formada entre las variables **treatment** y **rural** donde claramente se ve una asignación del tratamiento alrededor del 50 % de la población tanto en la zona rural como en el casco urbano.

La segunda tabla de contingencia es la formada entre las variables **treatment** y **mujer\_cabeza** donde nuevamente se observa casi una asignación del tratamiento alrededor del 50 % de la población tanto en los hogares donde la mujer es la cabeza de hogar como en los que no lo es.

Finalmente, la última tabla de contingencia entre las variables **treatment** y **tamaño\_hogar** muestra una distribución bastante aleatoria en la asignación del tratamiento en las familias de diferente tamaño. La distribución del tratamiento parece uniforme tanto entre hogares del mismo tamaño como entre los diferentes grupos de familias de tamaño distinto.

# Impacto causal del tratamiento sobre salud

1. Si se calcula como una diferencia de media, se tiene que:  
 $SDO = E[Y_i^1 | D_i = 1] - E[Y_i^0 | D_i = 0] = 4,873795$ .
2. Si se calcula mediante una estimación por **OLS** de una **switching regression** se pueden obtener dos resultados. Si no se incluyen variables de control se tiene que  $SDO = ATE = 4,874$  y si se decide incluir variables control en la regresión se obtiene  $ATE = 4,603$ .



**Figura 1:** Diferencia de media entre los tratados vs los no tratados

$SDO = ATE$ , es decir  $SB = 0$  (sesgo de selección desaparece) por el proceso de aleatorización del **RCT**

Controles para incluir:

- **escuela:** No todas las escuelas tienen las mismas condiciones de salubridad
- **tamano\_hogar:** Afecta las condiciones de salubridad en un hogar
- **rural:** Salubridad es distinta en el capo que la ciudad
- **mujer\_cabeza:** Puede afectar la salubridad
- **ingreso\_hogar:** El ingreso puede ser un predictor de la salubridad

Variables para no incluir:

- hogar

Cuadro 2:

	outcome
treatment1	4.603*** (0.560)
ingreso_hogar	0.000 (0.001)
rural1	-0.491 (0.334)
mujer_cabeza1	-1.367*** (0.385)
Constant	3.131** (1.243)
Observations	200
R <sup>2</sup>	0.594
F Statistic	14.707*** (df = 18; 181)

Notes:

\*\*\*Significant at the 1 percent level.

\*\*Significant at the 5 percent level.

\*Significant at the 10 percent level.

# SUTVA y violación del mismo

El supuesto de Stable across all units treatment value assumptions (**SUTVA**) es un supuesto clave para el RCT por medio de una *switching regresión*. Se basa en la definición de *Outcome alternativo potencial* sencilla y consiste en :

- La dosis del tratamiento es la misma para todas las unidades
- Lo que pasa a la unidad  $j$  no afecta el outcome de la unidad  $i$

El supuesto de **SUTVA** deja de ser válido cuando hay:

1. **efectos spillover**: El efecto en los tratados se esparce al grupo control
2. **efectos de red**: Se esparce la información entre los tratados al grupo control
3. **efectos de equilibrio económico general**: Hay mecanismos que funcionan a nivel agregado en la economía que hay que tener en consideración a la hora de estimar impactos causales

En este RCT podría existir la sospecha de que se está violando el supuesto de **SUTVA** dado que es muy probable que existan *efectos spillovers* en donde el tratamiento al grupo tratado pueda tener un efecto indirecto también sobre el grupo control mediante la interacción entre estos dos grupos.

# Submuestra y no violación del SUTVA

Para evitar la crítica de la diapositiva anterior una forma de controlar por esos *efectos spillovers* es por medio de un aislamiento del grupo de tratamiento frente al grupo control mediante lo que se conoce en la literatura como un **RCT con clustered design**.

Se observa que una submuestra que contenga solo los hogares cuyos niños vayan solo a las escuelas 5, 6, 7, 8, 9 y 10 se puede hacer un **RCT con clustered design** dado que para las escuelas 6, 8 y 10 ningún niño es tratado mientras que para las escuelas 5, 7 y 9 todos los niños son tratados lo que evita el efecto lo que evita los *efectos spillovers* al separar las dos grupos y así si se evita que haya **efectos de red** en la difusión de información el supuesto de **SUTVA** no se violaría, frente a los efectos de equilibrio general el riesgo es mas bien bajo al ser una muestra pequeña de niños en una región.

Realizando el filtramiento de la base de datos, seleccionando solo las escuelas mencionadas, se obtiene  $ATE = 4,652$  usando las mis variables control que se usaron para el punto 5. Se observa que el  $ATE$  calculado con la submuestra es un poco mayor al calculado por toda la muestra lo que demuestra que los efectos spillover estaban *subestimando* el impacto causal.

# Otros impactos positivos del programa

De igual forma, se observa que el programa no solo tiene un impacto positivo en el nivel de salud de los niños, variable **outcome** si no también en el nivel de aprendizaje/notas de los niños. Realizando una **switching regresión** de **notas\_hijos** respecto a la variable de asignación **treatment** se observa que existe un impacto causal positivo entre proveer el tratamiento y las notas de los niños como lo demuestra:

Teóricamente, esto se debe a que a mayores niveles de salud se esperaría que los niños puedan destinar más tiempo a estudiar además de tener una mayor concentración gracias a la mejor nutrición. La identificación del impacto causal estaría garantizado por el proceso de aleatorización.

Cuadro 3:

	nota_hijos
treatment1	0.801*** (0.196)
Constant	3.441*** (0.435)
Observations	200
R <sup>2</sup>	0.300
F Statistic	4.311*** (df = 18; 181)

Notes: \*\*\*Significant at the 1 percent level.  
 \*\*Significant at the 5 percent level.  
 \*Significant at the 10 percent level.



# Fortalezas y debilidades del estudio

## Fortalezas:

- El **RCT**, gracias a la *aleatorización*, puede estimar de manera correcta el impacto causal. Aún mejor, los *efectos spillovers* se pueden mitigar si se conduce un **RCT con clustered design**
- Si la muestra es significativa en tamaño y además también es de carácter aleatorio se puede generalizar los resultados obtenidos del **RCT** en la muestra a la población de niños de interés.

## Debilidades:

Es muy probable que no se pueda generalizar los resultados del **RCT** de esta muestra a poblaciones mucho mayores a la población en consideración. Por lo anterior, se puede decir que hay una limitación frente a la generalización de los resultados dado que puede que la muestra no sea significativa de poblaciones mucho mayores.

Este **RCT** no se puede generalizar a poblaciones distintas a la considerada en el estudio. Si se quisiera estimar el impacto causal sobre otra población del tratamiento en lugar de usar los resultados obtenidos en éste estudio se tendría que realizar un nuevo **RCT** sobre una muestra significativa de la nueva población de interés.

# ¿El impacto justifica el costo del programa?



En primer lugar, lo primero que se necesitaría sería el costo monetario del programa, dado que sin éste no podría hacer una comparación del costo del programa frente al impacto y dicho costo aún no se ha proveído. Además, se necesitaría una forma de evaluar el beneficio monetario del aumento general de salud de los niños. Por ejemplo, se podría estudiar la reducción de los gastos en salud y atención médica que implica el impacto causal de un  $ATE = 4,873795$  para los niños. Si dicha reducción en los gastos en salud y atención médica debido al mejoramiento en la salud de los niños gracias al programa es mayor al costo del programa, entonces presupuestariamente hablando valdría la pena conducir el programa.

No obstante, el programa no solo se debería medir por la reducción de gastos en salud atención médica si no que para hacer una verdadera medición económica se debería tener en cuenta los efectos sobre el bienestar de los niños que no se pueden cuantificar monetariamente fácilmente. Por ejemplo, no es fácil saber calcular monetariamente el aumento de bienestar porque ahora tienen mejores niveles de salud o porque ahora tienen mayores notas escolares. Además, existe externalidades que no son fácilmente cuantificables monetariamente como el beneficio en la comunidad por un aumento general de la salud de los miembros de ésta o el impacto a largo plazo que tendrá un mejor aprendizaje de los niños tanto en la comunidad como en una población más grande.



F. Bogliacino.

*Evaluación de Impacto: RCT.*

Universidad Nacional de Colombia, 2020.