



**Universidad**  
Zaragoza

# Trabajo Fin de Grado

## Ensayos controlados aleatorizados en el desarrollo económico

Autor/es

Alberto Palacios Abad

Director/es

Vicente Pinilla Navarro

Grado en Economía

Facultad de Economía y Empresa  
Zaragoza, Septiembre 2016

### Resumen:

La irrupción de un nuevo enfoque experimental en el desarrollo económico ha provocado un gran cambio en la manera de pensar las políticas de desarrollo. Los ensayos controlados aleatorizados son el principal método experimental de este nuevo enfoque. Este trabajo pretende analizar su funcionamiento y comprender por qué tienen tanta importancia en este campo. También intenta indagar en el papel que juegan los ensayos controlados aleatorizados en la actualidad en el desarrollo económico y sus posibilidades en el futuro.

### Abstract:

The irruption of a new experimental approach in development economics has induced a mayor change towards the way of thinking the development policies. The randomized controlled trials are the main experimental method of this brand new approach. This dissertation intends to analyze their functioning and understand why is so important in this field. It also pretends to examine the role of the randomized controlled trials in the present of the development economics and their possibilities in the future.

# Índice:

1. Introducción
2. Ensayo controlado aleatorizado
  - 2.1 Historia
  - 2.2 Evolución del desarrollo económico
  - 2.3 Economía de la pobreza
3. Metodología
  - 3.1 Principios básicos
  - 3.2 Problemas de diseño
  - 3.3 Un estudio de caso: Impacto en la educación y en la salud de la desparasitación intestinal.
4. Ventajas
5. Críticas
  - 5.1 Validez externa
  - 5.2 Heterogeneidad en los resultados
  - 5.3 Relación con la teoría
6. Conclusión
7. Bibliografía

## **CAPÍTULO I: INTRODUCCIÓN**

Este trabajo tiene como principal objetivo analizar el papel de los ensayos controlados aleatorizados en el campo del desarrollo económico. Durante los últimos años se han producido cambios en el enfoque de esta área que han revolucionado la manera de entenderla.

Tradicionalmente el desarrollo económico estaba dominado por una visión deductivista (general-particular), es decir, los académicos se dedicaban a formular teorías generales que se pudieran aplicar a todas las situaciones concretas. Sin embargo, desde hace unos años esa visión está cambiando hacia una más inductivista y experimental. Estos nuevos economistas le dan una importancia crucial al método empírico como fuente de conocimiento y su objetivo es extraer conclusiones de los experimentos que llevan a cabo una vez que queda demostrado su funcionamiento.

Mi intención con este trabajo es explicar cuáles son los factores que hacen que los ensayos controlados aleatorizados sean una herramienta tan adecuada según algunos científicos en la investigación en el campo del desarrollo económico desde un punto de vista experimental. Para ello mostraré la evolución que han tenido estos experimentos a lo largo de la historia y en qué campos han tenido más éxito con el objetivo de ser capaz de comprender qué papel juegan en el proceso científico. También pondré en contexto la situación académica del desarrollo económico hasta la llegada de este nuevo enfoque y analizaré cómo ha cambiado a causa de esto. Con el fin de que se entiendan los detalles de estos experimentos intentaré explicar de una manera sencilla la metodología que siguen y las dificultades a las que se enfrentan los profesionales a la hora de diseñar un programa de estas características. El último objetivo es hacer un análisis de las conclusiones a las que he llegado, señalando el papel que juegan los experimentos aleatorizados en el desarrollo económico.

Es clave una disquisición sobre este método experimental ya que está cobrando cada vez más importancia en el diseño de nuevas políticas para luchar contra la pobreza y ayudar a los países subdesarrollados. Más adelante en el trabajo se verá cómo organismos multilaterales como el Banco Mundial o el Banco Iberoamericano de

Desarrollo, o incluso gobiernos de los propios países están apostando por este enfoque debido a sus resultados.

El trabajo está estructurado en cuatro capítulos de desarrollo del tema más la conclusión. En el capítulo II explico qué es un ensayo controlado aleatorizado así como su desarrollo a lo largo de la historia. También incluyo una breve explicación de las teorías principales en el desarrollo económico hasta la llegada de este nuevo enfoque y la base teórica en la que se apoya. En el capítulo III muestro los anclajes metodológicos de estos experimentos y el proceso de toma de decisiones en el diseño de los mismos ilustrándolos con un sencillo ejemplo. En los capítulos IV y V pongo en contraste las principales virtudes que tiene este método experimental y las críticas más comunes que recibe. En la conclusión llevaré a cabo un balance final del papel que tienen en la actualidad y el que podrían tener en un futuro.

## **CAPÍTULO II: ENSAYO CONTROLADO ALEATORIZADO**

Un ensayo controlado aleatorizado es un tipo de experimento científico diseñado para probar la eficacia de una intervención (ya sea un fármaco, nuevas técnicas o políticas públicas) en un individuo o población. A priori se puede pensar que para ello bastaría observar las consecuencias tras haber implementado la intervención en un sujeto, pero debido a la complejidad de la realidad esto no es posible.

Hay numerosos factores o variables que pueden afectar al sujeto y falsear el resultado observado de la intervención. Siempre hay que tener en cuenta el contexto en el que se lleva a cabo la intervención para intentar aislar al experimento de todos los posibles elementos externos que puedan afectar al resultado. Esta no es una tarea fácil, sobre todo dependiendo del tipo de intervención que sea. Por ejemplo imaginemos una reforma del mercado laboral de un país determinado en un periodo de crisis. Al observar un año más tarde los resultados puede ocurrir que la tasa de paro no haya mejorado, pero no podemos saber hasta qué punto es culpa de la reforma o de la crisis, es decir, en sistemas complejos con muchas variables que no se comportan de forma lineal como sucede en la realidad es muy complicado dirimir a simple vista la causalidad directa de una variable sobre otra.

Para intentar solventar algunos de estos problemas se utilizan los ensayos controlados aleatorizados, donde los sujetos son aleatoriamente seleccionados para el grupo que va a recibir el tratamiento propuesto en el experimento o bien en el grupo que no lo recibe y que ejerce como control. El objetivo es que al final de la selección, los grupos sean estadísticamente equivalentes, por lo que los individuos deben tener características similares.

La finalidad de este tipo de experimentos es evaluar los resultados en el grupo que ha recibido el tratamiento y compararlo con el grupo de control que no lo ha recibido para comprobar el efecto que tiene y si es acorde a lo esperado. Al hacer esto, mantenemos los factores externos constantes, por lo que podemos aislar el efecto concreto del tratamiento y evaluar el resultado.

## 2.1 Historia

El primer experimento aleatorizado del que se tiene constancia lo llevó a cabo un médico escocés de la armada inglesa llamado James Lind en 1747. Encontrándose navegando a bordo del *Salisbury*, varios marineros enfermaron de escorbuto, Lind para intentar encontrar la cura prescribió a cada pareja de enfermos un tratamiento alimenticio diferente para después observar los efectos de cada uno. A los pocos días se dió cuenta de que la pareja de marineros a los que les había asignado una dieta a base de cítricos como naranjas y limones había mejorado rápidamente. Así se dió cuenta de que el escorbuto era producido por deficiencia de vitamina C y la gente que no consumía frutas y verduras eran los que sufrían esta enfermedad.

Posteriormente estos experimentos aparecieron en distintos campos como en la psicología, la educación y la agricultura hacia finales del S. XIX. A principios del siguiente siglo este método ya empezó a cobrar importancia en la investigación científica.

Ronald Fisher, biólogo y estadístico, publicó en 1925 el libro “*Statistical Methods for Research Workers*” y posteriormente en 1935 “*Design of Experiments*”. Estos libros tuvieron gran influencia a la hora de promover el uso de experimentos con métodos estadísticos en distintos campos de estudio.

En 1948 se publica el primer estudio de medicina realizado con este método que se titula “*Streptomycin treatment of pulmonary tuberculosis*”. Uno de los investigadores que lo llevaron a cabo fué Austin Bradford Hill, pionero en el diseño de los ensayos controlados aleatorizados modernos en el campo de la medicina. En la segunda mitad del S. XX los Ensayos Controlados Aleatorizados se convirtieron en el método estándar para los ensayos clínicos en medicina.

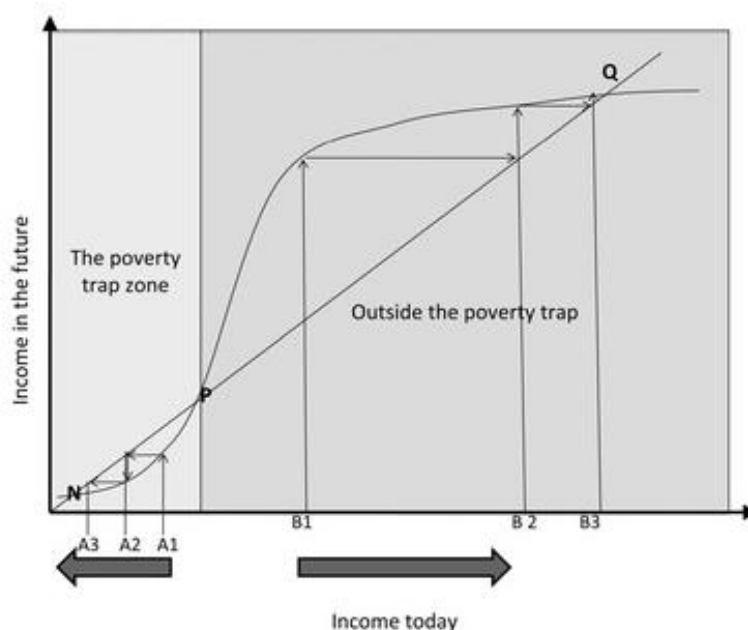
En la economía estos experimentos nunca han sido muy populares debido a la imposibilidad de observar un grupo control, sobre todo en temas de macroeconomía o política económica. La gran cantidad de variables interrelacionadas constituyen un obstáculo con el que hasta ahora no se ha podido lidiar para llevar a cabo este tipo de experimentos en esos campos.

## 2.2 Evolución del desarrollo económico

El campo del desarrollo económico, ha estado dominado históricamente hasta hace pocos años por un enfoque teórico y deductivista. El debate bipolar ha estado copado por los gurús de la oferta como Jeffrey Sachs que creen que el principal problema de los países pobres son las trampas de la pobreza “Africa’s governance is poor because Africa is poor” (Sachs, 2005) y los gurús de la demanda como William Easterly que piensan que la ayuda al desarrollo es más perjudicial que beneficiosa para los países pobres.

Según Sachs, estos países están estancados porque no tienen recursos suficientes para hacer las inversiones necesarias que les permitan iniciar el camino del progreso, sin embargo, si los países ricos destinaran suficientes fondos para ayuda al desarrollo podrían llevar a cabo las inversiones de capital necesarias para superar las trampas de la pobreza en varios campos, fertilizantes para aumentar la productividad en la agricultura, establecer un mercado de crédito y de seguros, mejorar la educación y la sanidad, etc..

Gráfico 2.2.1: Trampa de la pobreza



Fuente: Duflo y Banerjee (2011)

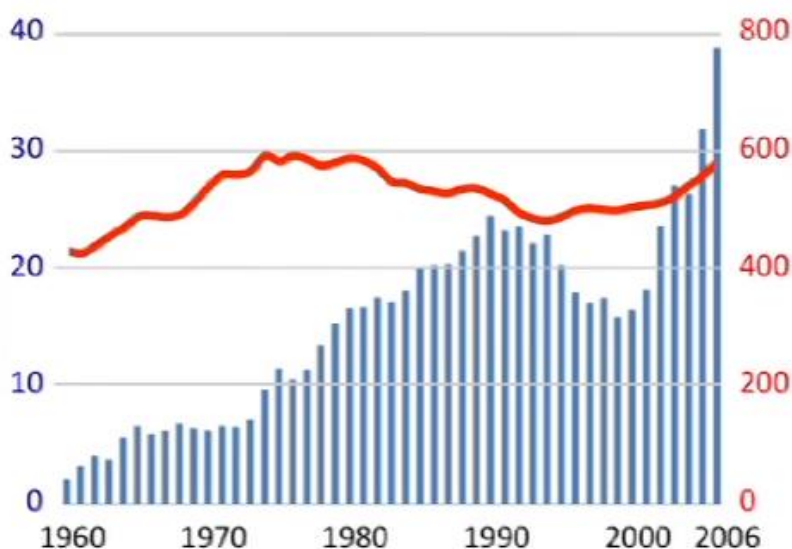


Como se puede observar en el gráfico, las personas que están en la zona de la trampa de la pobreza, suponiendo unos ingresos A1, en el futuro los verán disminuidos debido a que las inversiones que puedan hacer están por debajo del umbral de rendimientos crecientes (P). Por lo que sucesivamente llegan a N, perpetuándose la pobreza. Sin embargo, si se consigue llegar a un nivel superior al umbral de rendimientos crecientes, con los ingresos B1 se pueden llevar a cabo inversiones que permiten en un futuro aumentar de los ingresos y así crecerán repetidamente hasta llegar a Q. Esa inversión que permite superar la barrera de P es a la que se refiere Sachs como fundamental para eliminar la pobreza.

En contraposición están los gurús de la demanda como William Easterly, escépticos con la ayuda al desarrollo porque creen que su única consecuencia es perpetuar los gobiernos corruptos y paternalistas. No ayuda en modo alguno a solucionar el problema clave que es la falta de derechos políticos y civiles que necesita una sociedad para prosperar.

En *The White Man's Burden*, Easterly hace una feroz crítica de los gurús de la oferta y en especial de su compañero Sachs. Cree que estos planificadores con grandes ideas son la reencarnación moderna del infame engaño colonial (Easterly, 2006), estas políticas que se han venido llevando a cabo desde la Segunda Guerra Mundial han hecho más mal que bien a los países pobres y que no hay evidencias concluyentes que apoyen la teoría de que la ayuda al desarrollo elimine la pobreza.

Gráfico 2.2.2: Ayuda al desarrollo e ingresos en África.



Fuente: Easterly "Can the West save Africa?", World Trade Institute (WTI)

Las barras en el gráfico (eje izquierda) representan la ayuda al desarrollo recibida por África en billones de dólares; la línea (eje derecha) representa el PIB/ cápita de África a precios constantes.

El gran problema es que es un debate muy ideologizado, en el que aún no sabemos a ciencia cierta qué es lo que funciona. Como observamos en el gráfico anterior, es difícil apreciar una correlación entre el dinero destinado a ayuda al desarrollo y el aumento del PIB per cápita, pero aunque la hubiera, habría que ser cauteloso a la hora de sacar conclusiones en contextos tan complejos y con tantas variables. Probablemente las dos partes tengan algo de razón y las causas de la pobreza en estos países sean varias y de distinta solución pero hasta ahora no hay datos concluyentes.

Sin embargo, desde las últimas décadas del S. XX, algunos académicos decidieron cambiar el enfoque de estudio del desarrollo económico. En vez de centrarse en explicaciones teóricas y deductivas como se venía haciendo hasta ahora, probaron una aproximación más experimental e inductiva. Estos autores llevan a cabo experimentos de políticas públicas concretas a pequeña escala, dado que permite evaluar los resultados y extraer conclusiones de qué funciona y cómo lo hace. De esta manera se puede conocer a ciencia cierta la efectividad de un programa en unas condiciones determinadas.

Un buen ejemplo es el programa PROGRESA, llevado a cabo por el gobierno de Ernesto Zedillo en México en 1997, con el objetivo de promover la educación y la salud infantil en las zonas rurales más pobres del país. El programa consiste en una transferencia monetaria condicionada entregada a las madres de cada familia con la condición de que sus hijos asistan a la escuela y al centro sanitario. Esta transferencia está pensada para suplir lo que el niño podría ganar trabajando y aumentar el poder adquisitivo de la familia, sobre todo para gasto en alimentación (suelen dar cheques para comprar alimentos).

El gobierno mexicano dadas las restricciones presupuestarias, decidió diseñar primero un programa piloto en 6.344 aldeas para evaluar los efectos. Algunas de ellas fueron elegidas aleatoriamente para ser beneficiarias del programa (grupo tratamiento) y

posteriormente se comparaban con aldeas donde no se había puesto en marcha (grupo control). Se vieron beneficiadas 300.000 familias con pocos recursos.

El International Food Policy Research Institute realizó un estudio durante los 15 primeros meses en aquellas aldeas en las que se llevó a cabo el programa y los resultados fueron bastante positivos. La tasa de chicos matriculados en la escuela secundaria aumentó en un 10%, y la de chicas en un 20%. También se comprobó que los niños tenían un 12% menos de incidencias de enfermedades y se estimó que el programa tenía un retorno del 8% (Schultz, 2000).

Gracias al éxito del programa el gobierno posterior de Vicente Fox amplió el programa (cambiándolo de nombre a Oportunidades) hasta llegar a las 5 millones de familias, casi la totalidad de las que viven en extrema pobreza, situadas en 86.000 localidades. También otros países de América Latina han puesto en marcha programas parecidos, como Bolsa Escola en Brasil, Red de Protección Social en Nicaragua o Programa de Asignación Familiar en Honduras.

Este tipo de evaluaciones se han popularizado desde comienzos del S.XXI, en parte gracias a Esther Duflo y Abhijit Banerjee, dos jóvenes economistas del MIT (Massachusetts Institute of Technology) que en 2003 decidieron invertir 100.000\$ que les había proporcionado la universidad para destinarlos en investigación y fundar la organización Abdul Latif Jameel Poverty Action Lab (J-PAL). El objetivo era evaluar políticas llevadas a cabo en el ámbito del desarrollo económico y la lucha contra la pobreza usando para ello ensayos controlados aleatorizados. A través de estos experimentos, Duflo y Banerjee intentan comprender los problemas específicos de la pobreza y cómo se comporta el ser humano en estas situaciones.

Desde entonces, han llevado a cabo más de 500 ensayos aleatorizados en 56 países distintos del mundo y han entrenado a 1.500 personas para que lleven a cabo estos experimentos. Organizaciones como el Banco Mundial, el Banco Interamericano de Desarrollo, ONGs y gobiernos de distintos países han tomado buena nota del éxito de estos métodos y ya colaboran con el J-PAL en la evaluación de programas en distintos campos como sanidad, educación, agricultura o microcréditos.

## 2.3 Economía de la pobreza

En este apartado voy a tratar de explicar la base teórica de la pobreza en las que se sustenta el enfoque experimental de los ensayos controlados aleatorizados. También intentaré poner de relieve los problemas a los que se enfrentan las personas que intentan diseñar políticas para ayudar a personas que se encuentran en la pobreza y cuáles son los causantes de estos problemas.

La pobreza afecta a la toma de decisiones de los individuos, aun suponiendo un comportamiento neoclásico, un *homo economicus* perfectamente racional y maximizador de su bienestar no se comportará igual en una situación de pobreza que de riqueza. La gente que vive en la pobreza no tiene la misma cantidad de información o posibilidad de ella que la gente que vive en una situación acomodada, lo que hace que tomar decisiones sea mucho más complicado. Además, este grupo de gente acomodada no tiene que tomar muchas de estas importantes decisiones como la cobertura sanitaria o el dinero de la jubilación, sin embargo, gente con muchos menos recursos y poco margen de error tienen que decidir sobre temas muy complejos.

Ted Schultz, ganador del Premio Nobel de Economía en 1979 fué de los primeros en darse cuenta de ello, postulando que los pobres son lo más eficientes que pueden dadas sus condiciones de vida. En su discurso de aceptación del Nobel lo explicó de una manera muy sencilla:

“La gente que es rica encuentra difícil de entender el comportamiento de la gente pobre. Los economistas no son una excepción, encuentran difícil comprender las preferencias y las restricciones de la escasez que determina las elecciones que la gente pobre hace. Todos sabemos que la mayoría de la población del mundo es pobre, que ganan una miseria trabajando, que más de la mitad de su escaso sueldo es gastado en alimentación, que viven predominantemente en países de bajos ingresos y que la mayoría viven de la agricultura. Lo que muchos economistas no logran entender es que los pobres no están menos preocupados en mejorar sus vidas y las de sus hijos que la gente rica.”

Este nuevo enfoque de observar la pobreza permitió nuevas aproximaciones para intentar comprender y solucionar los problemas inherentes de esta. Varios autores

iniciaron investigaciones para averiguar las ineficiencias de los mercados que producían los incentivos perversos que afectaban a los pobres y posibles políticas públicas que ayudaran a mejorar su situación.

A continuación voy a explicar con dos ejemplos el funcionamiento de los incentivos perversos en dos de las áreas donde mejor se puede distinguir la diferencia entre ricos y pobres: los mercados de seguros y la agricultura.

A priori podemos pensar que los seguros deberían ser muy atractivos para las personas que viven en una situación de pobreza, pues viven expuestos a innumerables riesgos todos los días que ponen en jaque su supervivencia, ya sean por motivos de salud, inclemencias del tiempo u otras circunstancias ajenas a ellos. Seguros que en los países desarrollados son muy comunes como los de salud, seguros contra las malas cosechas o el ganado, en los países poco desarrollados son casi inexistentes, ¿a qué se puede deber esto?

En primer lugar, las empresas oferentes se enfrentan a los mismos problemas de riesgo moral y selección adversa, solo que en los países poco desarrollados, tienen menos herramientas para combatirlos. Mientras en los países desarrollados pueden conseguir bastante información de un posible cliente y tienen herramientas para coaccionarlos a que cumplan el contrato, en los países pobres es bastante más complicado.

Una de las posibles soluciones que se han dado ha sido asegurar a un grupo suficientemente grande para evitar esta selección adversa. En 2007 SKS Microfinance, una gran institución de microcréditos en la India llevó a cabo un programa piloto de seguros sanitarios con varias condiciones. En primer lugar, el seguro tenía que ser contratado por el grupo entero al que se le ofrecía (que ya eran clientes suyos de microcréditos) para combatir la selección adversa, y sólo se podían utilizar los hospitales con los que SKS tenía un acuerdo.

En un principio ya se encontraron con el problema de que varios clientes protestaron porque fuera obligatorio, con lo cual solo lo hicieron obligatorio con la primera renovación del crédito. Posteriormente se encontraron con que la tasa de renovación las siguientes veces del mismo crédito caía del 60% al 50% en las zonas que ofrecían el

seguro. Decidieron eliminar el programa porque estaban perdiendo dinero y clientes, y además al hacerlo voluntario, volvía el problema de la selección adversa.

Este es un ejemplo que permite ilustrar la baja demanda de seguros en las regiones más desfavorecidas. Uno de los motivos de esta baja demanda puede ser la falta de confianza en las instituciones que los ofrecen y en la ayuda que reciben de los seguros, pues normalmente son para situaciones muy concretas. Por ejemplo, los seguros sanitarios son para accidentes o enfermedades graves, no sirven para una infección de estómago. O los seguros para la agricultura son para cuando llueve menos de una cantidad indicada, pero puede haber llovido y que la cosecha sea mala, con lo cual los potenciales beneficios de los seguros se ven diluidos.

Otro motivo es la inconsistencia temporal que sufrimos todas las personas, que tendemos a valorar más el presente que el futuro. En los países poco desarrollados aún se ve más acentuado dado que sus ingresos no les permiten planificar muchos más allá que lo que van a consumir cada día y no saben si pueden necesitar ese dinero que van a invertir en el seguro de un año para comer la semana que viene. Además tampoco ayuda que el beneficio del seguro solo se reciba en situaciones adversas que no invitan a pensar en ellas.

Todo esto complica el establecimiento de un mercado de seguros de manera espontánea, lo cual no quiere decir que sea la única manera. En algunos países se está experimentando con subsidios a las pólizas de seguros para que los posibles clientes cojan confianza con las empresas que los ofrecen y comprueben por sí mismos los potenciales beneficios aunque por el momento no hay conclusiones sólidas.

La agricultura es otra área donde pueden existir situaciones o incentivos perversos que afectan negativamente a la toma de decisiones o comportamiento de los agentes haciendo que el resultado no sea del todo eficiente.

Uno de los fenómenos que más negativamente contribuyen a la productividad es el régimen de aparcería. Shaban (1987) llevó a cabo un experimento en la India para ver cómo estos acuerdos entre terratenientes y trabajadores afectaban a la producción.

Descubrió que el trabajador trabaja un 40% más y usa un 20% más de fertilizante si la tierra es de su propiedad y además la productividad de esta será un 30% mayor. Al tener que repartir los beneficios de igual manera sea cual sea el esfuerzo, es comprensible que el trabajador no ponga tanto énfasis en el trabajo que si fueran para él solo, por lo tanto una redistribución de tierras entre pequeños agricultores podría ayudar a mejorar la productividad y la inversión.

Pero este no es el único problema, dado que aunque cada trabajador tuviera su parcela en posesión, seguiría habiendo mejoras potenciales de la productividad ya que la inversión no será la óptima. Aunque hay numerosas evidencias de los beneficios de los fertilizantes u otras semillas variadas, sigue habiendo bastante reticencia a usarlos en las regiones menos desarrolladas.

El principal problema para establecer la inversión en mejoras de productividad es el ahorro, a la gente pobre le cuesta mantener un dinero ahorrado en casa sin darle otro uso como comprar comida u otra emergencia. Por eso en el tiempo que pasa desde la cosecha hasta la siembra se consume la mayor parte de las ganancias y a la hora de comprar el fertilizante muchos se encuentran que no tienen el dinero suficiente (Kremer, Duflo, Robinson 2010). Otras causas también son que no saben utilizarlo correctamente o no están seguros de si la mejora del rendimiento va a merecer la pena.

Algunas políticas pueden incentivar a realizar las inversiones necesarias para mejorar el rendimiento de sus tierras como darles muestras gratuitas de fertilizante y enseñarles a utilizarlas para que comprueben por ellos mismos los beneficios (Sachs 2005) o poner a su disponibilidad la compra de fertilizantes cuando acabe la cosecha y guardarlos hasta la siembra o similares.

En resumen, estas dos áreas sirven de ejemplo para ilustrar la importancia de ser capaces de comprender el proceso de toma de decisiones de la gente pobre y de seguir investigando para descubrir cuáles son las políticas más efectivas para ayudar en el desarrollo de las regiones menos favorecidas. En este contexto es donde los ensayos aleatorizados demuestran ser una gran herramienta de gran ayuda para alcanzar estos objetivos.

## CAPÍTULO III: METODOLOGÍA

En este capítulo voy a explicar los principios básicos metodológicos de los ensayos controlados aleatorizados con el objetivo de proveer de las herramientas necesarias para comprender los diferentes diseños posibles de este tipo de experimentos, así como las cuestiones y suposiciones que se tienen que plantear los autores. Finalmente intentaré ilustrar este proceso con un ejemplo, un artículo de Michael Kremer y Edward Miguel sobre el impacto en la educación y en la salud de la desparasitación intestinal en los niños.

### 3.1 Principios básicos

La forma funcional básica de estos ensayos es una regresión que puede ser lineal o no lineal (aquí pongo el ejemplo de una lineal para simplificar) que describe el comportamiento de la variable dependiente con respecto de la(s) independientes.

$$Y_i = \alpha + \beta X + \gamma Z_i + \epsilon_i \quad (1)$$

La variable  $Y_i$  es el resultado de la regresión para el grupo  $i$ ,  $\alpha$  es la constante de la ecuación,  $X$  es la variable independiente (puede haber más de una variable independiente),  $\beta$  es el parámetro que mide la influencia de la variable independiente en la dependiente.  $Z_i$  es el conjunto de variables control del grupo  $i$  siendo  $\gamma$  su coeficiente, y por último,  $\epsilon_i$  es el error del cálculo.

En el caso de la ecuación 1, asumimos que solo hay un tratamiento y que una proporción  $P$  de la muestra  $S$  es tratada. Las variables tienen que cumplir la condición de ser independientes e idénticamente distribuidas (i.i.d). Debemos asumir también que cada individuo ha sido elegido de forma aleatoria de la muestra total y que los dos grupos resultantes tratamiento y control son estadísticamente similares con varianza  $\sigma^2$ .

Para comprobar si el tratamiento propuesto tiene efecto, se testea la hipótesis nula  $H_0$  que indica que el efecto del tratamiento es igual a 0 contra la alternativa de que es



distinto de 0. Establecemos un tipo de contraste bilateral, aunque en otros casos puede ser unilateral:

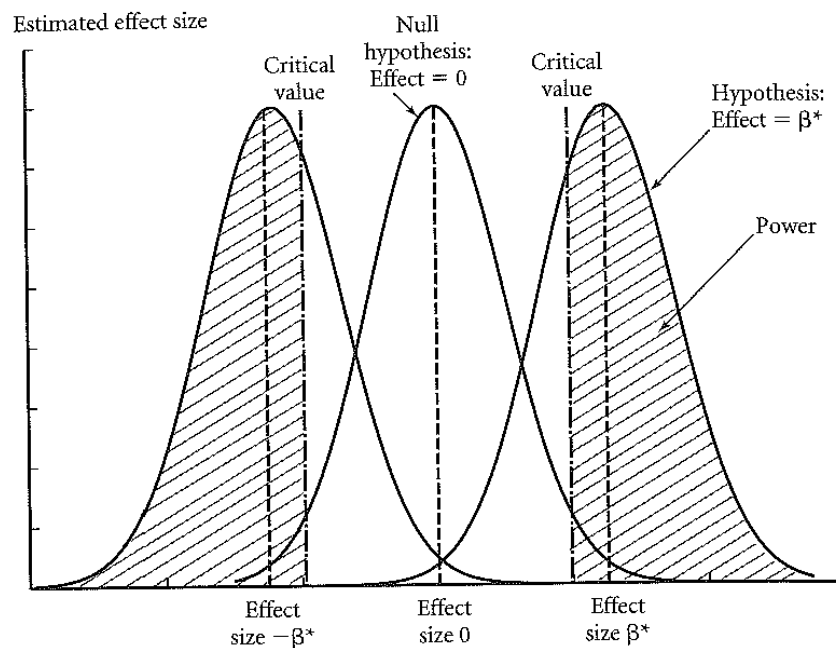
Hipótesis nula  $H_0: \beta = 0$

Hipótesis alternativa  $H_1: \beta \neq 0$

El valor crítico  $t_{\alpha/2}$  viene dado por el nivel de significatividad que queramos comprobar (normalmente es el 5%), y dado que es bilateral, se rechaza  $H_0$  si  $\hat{\beta}$  queda por encima o por debajo del valor crítico; esto es:

$$\frac{|\hat{\beta}|}{SE(\hat{\beta})} > t_{\alpha/2}$$

Siendo  $SE(\hat{\beta})$  la desviación típica de  $\hat{\beta}$ . Dado que es bilateral, puede ser significativa tanto si el valor de la estimación es positiva como si es negativa. En el siguiente gráfico se puede observar la distribución de los efectos estimados con el valor crítico para un test bilateral.



Fuente: Glennerster y Takavarasha

Otra manera alternativa de comprobar la significatividad del resultado, es mediante el p valor. Cuando se lleva a cabo la regresión, se compara el p-valor obtenido con 0,05 (para el 95%, también puede ser 0,01 para el 99% dependiendo de la significatividad que se quiera comprobar en cada caso). Si este es menor o igual que 0,05, la hipótesis nula  $H_0$  se rechaza y el resultado es significativo.

Hay que recordar que este es solo un ejemplo de lo que podría ser una regresión y la metodología que se usa para comprobar los resultados, pero las regresiones pueden tomar muchas formas diferentes y las suposiciones que se hagan también podrán variar dependiendo de las características del experimento, esto puede llevar a utilizar distintos contrastes para testear las hipótesis que dispongamos.

### 3.2 Problemas de diseño

En este apartado explicaré las dificultades y disyuntivas que se presentan a la hora de diseñar un ensayo controlado aleatorio.

Una de las principales decisiones que hay que tomar es a qué nivel aleatorizas, es decir a nivel individual o por grupos. Dependiendo de las características del experimento, será más conveniente uno u otro. Hay situaciones en las que la propia naturaleza del ensayo ya te obliga a hacerlo de una determinada manera, por ejemplo Chattopadhyay y Duflo (2004) estudiaron las consecuencias de que una mujer estuviera en una posición importante en los consejos de algunas aldeas en la India, en este caso está claro que hay que llevar a cabo una aleatorización a nivel de aldea para observar los efectos.

El problema surge cuando la decisión no es tan fácil y puedes elegir. Para ello tienes que tener en cuenta varios factores; cuanto más grandes sean los grupos que se aleatorizan, más grande será la muestra y esto repercute directamente en el coste del experimento haciendo que si no se dispone de mucho capital sea más conveniente el nivel individual.

Otro factor a tener en cuenta son las posibles externalidades positivas que puedan afectar al grupo control si no están claramente separados. Por ejemplo en casos de ensayos sobre políticas sanitarias que fácilmente tengan efectos positivos sobre otras personas por evitar el contagio. En casos así puede ser más adecuado una aleatorización a nivel de grupo para evitar un resultado erróneo.

Además suele ser más eficiente ya que los costes de implementación de la política son fijos cuando los llevas a cabo en un sitio concreto por lo que sale mejor si todos los individuos lo reciben que si solo los reciben algunos. Así también se evitan comportamientos no deseados, como envidias o resentimientos entre miembros del mismo grupo.

Con el mismo objetivo de maximizar la eficiencia, los autores se pueden decantar por un diseño “cross-cutting”. Si el programa que quieren implementar consta de diferentes intervenciones, este tipo de diseño permite comparar cada intervención por separado

para comprobar su efecto. Por ejemplo, si un experimento consiste en implementar las intervenciones A+B, tendremos el grupo de control, el grupo de tratamiento 1 (se le implementará A+B), el grupo de tratamiento 2 (se le implementará A) y el grupo de tratamiento 3 (se le implementará B), de esta manera podemos averiguar qué intervención tiene mejor resultado, o incluso si es la combinación de las dos lo que hace que tengan efecto.

También es importante señalar que esto también permite una mejor comprensión de las consecuencias de las intervenciones llevadas a cabo, lo que nos dará ventaja para en un futuro aprender a diseñar mejor este tipo de intervenciones.

La recolección inicial de datos es fundamental para la correcta planificación de la intervención, para ello hay dos opciones, llevar a cabo una encuesta de referencia en los sitios donde se va a implementar la política o usar datos de la administración si están disponibles, ya sea a través del gobierno del país o de organizaciones multilaterales como el Banco Mundial.

Las ventajas de una encuesta de referencia son varias, en primer lugar, llevar a cabo esa encuesta puede proporcionar variables que no se habían tenido en cuenta que ayuden a mejorar el modelo y a su vez simplificar la intervención lo que puede llevar a reducir costes. También puede mejorar la comprensión del funcionamiento de los incentivos de la intervención en esa zona, útil para posteriormente hacer un análisis sobre la validez externa del experimento.

En muchas ocasiones, los datos disponibles de la administración u otras organizaciones son suficientes para diseñar el experimento, aunque la disponibilidad de estos datos es mayor cuando se trata de países desarrollados que en vías de desarrollo, en los últimos años ha habido un gran esfuerzo por recolectar datos de estos últimos países que ayude al diseño de políticas.

Hay que tener cuidado con los datos recolectados, puede ser que no sean compatibles entre sí, porque puede que los hayan calculado con diferentes metodologías o supuestos que a la hora de obtener nuestros resultados lleven a errores o confusiones. La gran

ventaja de este método es que reduce el coste del ensayo en gran medida en tiempo y en dinero.

Teniendo en cuenta las ventajas y desventajas de cada método, hay que hacer un balance que tenga en cuenta las características del experimento y decidir qué diseño se adecua más al experimento que se desea llevar a cabo.

### **3.2 Un estudio de caso: Impacto en la educación y en la salud de la desparasitación intestinal.**

Para ayudar a comprender el proceso de diseño y análisis de resultados de los ensayos controlados aleatorizados vamos a estudiar un caso concreto. He elegido un experimento muy exitoso que ha ayudado al diseño de políticas para combatir los parásitos intestinales en los niños de países poco desarrollados. Edward Miguel y Michael Kremer (2004) decidieron diseñar un experimento para comprobar cómo afectaba la desparasitación de los intestinos en los niños con pocos recursos en la asistencia a la escuela, sus resultados académicos y su salud general. Aunque había bastante literatura existente creyeron que no se capturaba bien el impacto, pues la mayoría de los anteriores ensayos habían aleatorizado individualmente a los niños y esto no tenía en cuenta las posibles externalidades positivas que podía tener el programa y por lo tanto la diferencia entre control y tratamiento era menor de la real. También confiaban en la comparación pre-post sin prestar atención a elementos externos que pueden afectar en buena medida a la propagación de estas infecciones intestinales como las estaciones, las temporadas de lluvias y otros factores.

Para combatir estos problemas, decidieron llevar a cabo un enfoque distinto. En vez de aleatorizar individualmente, lo hicieron a nivel de la escuela porque así podían medir mejor el efecto entre escuelas aunque siguiera existiendo externalidades positivas debido a la menor probabilidad de contagio de los niños en barrios que hubiera una escuela con el tratamiento. Pero esto no solo no dificultaba la medición del efecto, si no que permitía medir las externalidades positivas al hacer un análisis del efecto teniendo en cuenta la densidad de escuelas con el tratamiento en las zonas determinadas.

Decidieron llevar a cabo el proyecto en una región de Kenia, una zona rural pobre densamente poblada y principalmente agrícola que tenía la tasa de infecciones de estómago más alta de la región. Lo llevó a cabo una ONG holandesa llamada ICS con la colaboración de las autoridades de la región. Para ello escogieron 75 escuelas de educación primaria que en total tenían alrededor de 30000 alumnos de entre 6 y 18 años. Crearon 3 grupos con 25 escuelas cada uno de manera aleatoria, debido a las restricciones presupuestarias del proyecto, las intervenciones médicas se llevaron a cabo

durante varios años. En 1998 el Grupo 1 de escuelas empezaron a recibir tratamiento gratuito para desparasitar a los niños, por lo que hacían de grupo tratamiento mientras que los grupos 2 y 3 actuaban como control. En 1999 el Grupo 2 también empezó a recibir tratamiento y el Grupo 3 sirvió como control, finalmente en 2001 el Grupo 3 también recibió el tratamiento.

Los miembros de ICS realizaron encuestas entre los alumnos de las escuelas en 1998 y principios de 1999 previas a la puesta en marcha del tratamiento para comprobar las características demográficas, socio-económicas y nutricionales de cada grupo que a pesar de la selección aleatoria resultaron tener diferencias en algunos aspectos. Por ejemplo, el Grupo 1 tenía más síntomas de infecciones estomacales y estaban menos limpios que los del Grupo 2 y 3 aunque solo eran dos características de entre todas las que constaban las encuestas.

Cuando llevaron a cabo la intervención médica, un 78% de los niños a los que se les había seleccionado para recibir el tratamiento lo recibieron, la mayoría de los que faltan no lo recibieron debido a la no asistencia a la escuela los días que tocaba.

Para calcular el impacto del tratamiento y de las posibles externalidades positivas del mismo diseñaron una regresión lineal teniendo en cuenta las posibles variables que pudieran afectar al resultado del tratamiento.

$$Y_{ijt} = \alpha + \beta_1 \cdot T_{1it} + \beta_2 \cdot T_{2it} + X_{ijt}\delta + \sum_d(\gamma_d \cdot N_{dit}^T) + \sum_d(\theta \cdot N_{dit}) + u_i + e_{ijt} \quad (2)$$

Siendo  $Y_{ijt}$  el posible resultado ya sea de salud o de educación, donde  $i$  es la escuela,  $j$  el estudiante y  $t$  el año del programa (1 o 2).  $T_{1it}$  y  $T_{2it}$  son indicadores variables del primer y segundo año del tratamiento en la asignación de las escuelas.  $\beta_1$  y  $\beta_2$  son el efecto directo del tratamiento durante el primer y segundo año respectivamente.  $X_{ijt}$  son las variables características de la escuela y del alumno, siendo  $\delta$  su coeficiente.  $N_{dit}$  es el número de alumnos de una escuela primaria situados a una distancia  $d$  de una escuela  $i$  en el año  $t$  y  $N_{dit}^T$  es el número de estos alumnos aleatoriamente seleccionados para el tratamiento sanitario, siendo  $\theta$  y  $\gamma_d$  sus correspondientes coeficientes.

Ahora voy a comentar algunos de los resultados que obtuvieron después de llevar a cabo el tratamiento sanitario durante los dos primeros años en el Grupo 1 y Grupo 2.

Tabla 1: Tasa de asistencia escolar

	Group 1 (25 schools)	Group 2 (25 schools)	Group 3 (25 schools)		
<b>Panel A:</b>					
<i>First year post-treatment (May 1998 to March 1999)</i>	<i>1st Year Treatment</i>	<i>Comparison</i>	<i>Comparison</i>	<i>Group 1 – Group 3</i>	<i>Group 2 – Group 3</i>
Girls <13 years, and all boys	0.841	0.731	0.767	0.093*** (0.031)	-0.037 (0.036)
Girls ≥13 years	0.864	0.803	0.811	0.057** (0.029)	-0.008 (0.034)
Preschool, Grade 1, Grade 2 in early 1998	0.795	0.688	0.703	0.100*** (0.037)	-0.018 (0.043)
Grade 3, Grade 4, Grade 5 in early 1998	0.880	0.789	0.831	0.070*** (0.024)	-0.043 (0.029)
Grade 6, Grade 7, Grade 8 in early 1998	0.934	0.858	0.892	0.059*** (0.021)	-0.034 (0.026)
Recorded as “dropped out” in early 1998	0.064	0.050	0.030	0.022 (0.018)	0.020 (0.017)
Females <sup>b</sup>	0.855	0.771	0.789	0.076*** (0.027)	-0.018 (0.032)
Males	0.844	0.736	0.780	0.088*** (0.031)	-0.044 (0.037)
<b>Panel B:</b>					
<i>Second year post-treatment (March to November 1999)</i>	<i>2nd Year Treatment</i>	<i>1st Year Treatment</i>	<i>Comparison</i>	<i>Group 1 – Group 3</i>	<i>Group 2 – Group 3</i>
Girls <13 years, and all boys	0.713	0.717	0.663	0.050* (0.028)	0.055* (0.028)
Girls ≥14 years <sup>c</sup>	0.627	0.649	0.588	0.039 (0.035)	0.061* (0.035)
Preschool, Grade 1, Grade 2 in early 1998	0.692	0.726	0.641	0.051 (0.034)	0.085** (0.034)
Grade 3, Grade 4, Grade 5 in early 1998	0.750	0.774	0.725	0.025 (0.023)	0.049** (0.023)
Grade 6, Grade 7, Grade 8 in early 1998	0.770	0.777	0.751	0.020 (0.027)	0.026 (0.028)
Recorded as “dropped out” in early 1998	0.176	0.129	0.056	0.120* (0.063)	0.073 (0.053)
Females <sup>b</sup>	0.716	0.746	0.648	0.067** (0.027)	0.098*** (0.027)
Males	0.698	0.695	0.655	0.043 (0.028)	0.041 (0.029)

Los resultados son significativos al 99 (\*\*\*), 95 (\*\*) y 90 (\*) por ciento de confianza.

Fuente: Miguel y Kremer (2004)

Como se puede observar la tabla contiene dos paneles en los cuales se muestran los resultados de asistencia a la escuela de varias variables en distintos años, el panel A es



correspondiente al primer año de tratamiento donde solo el Grupo 1 lo recibe y el panel B corresponde al segundo año, donde los Grupos 1 y 2 lo reciben.

Tenemos 8 variables observables, la primera son niñas menores de 13 años y todos los niños en los que se muestra una mejora de un 9,3% en el primer año de tratamiento del Grupo 1 con respecto a los grupos control y en el segundo año una diferencia del 5% entre los grupos tratamiento (1 y 2) y el control (3). La llamativa separación de las niñas mayores de 13 años es debida a que se encuentran en una edad fértil, y los protocolos sanitarios dictan que no hay que tratarlas por posibles amenazas de teratogenicidad de los medicamentos usados en el tratamiento. Aun así alrededor del 9% de estas niñas recibieron el tratamiento porque a veces era complicado determinar la edad o justo cumplían los años cuando lo estaban recibiendo. Pero la mejora de la asistencia de las niñas en el grupo tratamiento del 5,7% en el primer año (en el segundo apenas son significativas) es debida en mayor parte a las externalidades positivas de encontrarse en una escuela en la que los alumnos eran tratados.

También tenemos tres variables creadas para observar la diferencia de mejora de la asistencia escolar dependiendo del curso que el alumno cursara en la escuela, la primera corresponde a infantil, Grado 1 y 2, la segunda a Grados 3, 4 y 5 y la tercera a Grados 6, 7 y 8. En el Panel A se observa claramente que los cursos más bajos son en los que más diferencia existe entre el Grupo 1 y los Grupos 2 y 3, con un 10%, mientras que en los cursos superiores esa diferencia es del 5,9%. En el Panel B a pesar de que algunos resultados no son significativos los que sí lo son apoyan esta tendencia, así que podemos concluir que el tratamiento produce un mayor efecto en los niños que van a los primeros cursos que a los más mayores.

Otras variables son niños que se tenía constancia que habían abandonado la escuela, pero los resultados que se muestran no son significativos y el sexo de los niños. Hay que señalar que en el primer año, el efecto de mejora sobre los niños es mayor que el de las niñas, 8,8% y 7,6% respectivamente, sin embargo, en el segundo año es al revés, si bien los resultados de los niños no se advierten como significativos.

A rasgos generales se puede decir que el primer año de tratamiento tiene un mayor efecto que el segundo como se puede advertir en el Panel B aunque no sea una

diferencia muy significativa, y también parece claro el mayor efecto positivo sobre los niños más jóvenes. Hay que tener en cuenta que este solo es uno de los múltiples escenarios que los autores analizaron con los resultados del ensayo.

Gracias a este programa de desparasitación intestinal, se sacaron varias conclusiones significativas, en primer lugar, el tratamiento resultó en una mejora de la asistencia a la escuela del 7,5% de media, reduciendo el absentismo escolar en una cuarta parte. Este tratamiento también tuvo efectos positivos sobre alumnos que no recibieron el tratamiento, tanto a nivel educativo como sanitario, y estas externalidades justifican según cálculos de los autores no solo el tratamiento gratuito de desparasitación, sino incluso el pago a los individuos por recibir el tratamiento. Este tratamiento se demostró el más coste-efectivo de varios tratamientos que llevó a cabo ICS en Kenia, y pone de manifiesto la importancia de las enfermedades tropicales que afectan a ámbitos tan importantes como la salud, la educación y que finalmente repercuten en los ingresos.

## **CAPÍTULO IV: VENTAJAS**

La dificultad de establecer causalidad entre dos hechos siempre ha sido uno de los principales quebraderos de cabeza de la ciencia, pues no siempre basta con la simple observación de los hechos, en entornos complejos hay multitud de variables que no se pueden apreciar a simple vista y que pueden influir en el proceso. Para resolver este problema, los científicos han desarrollado diferentes métodos experimentales para intentar probar dicha causalidad, por ejemplo, diferencias en diferencias, cuasiexperimentos, regresiones multivariantes, regresiones discontinuas y otros.

Uno de los principales retos a los que se enfrentan estos experimentos, es el sesgo de selección, es decir, cómo seleccionar a los participantes en dichos experimentos, ya sea para el grupo control o tratamiento, y asegurarse de que no haya ningún sesgo o variable no controlada que pueda afectar a la validez interna, puesto que resultaría en un falseamiento de los resultados. Es en esta encrucijada, en la cual los distintos tipos de experimentos intentan solventar este obstáculo de diferentes formas, donde los ensayos controlados aleatorizados tienen su mayor ventaja frente a otros métodos.

La selección aleatoria de los individuos pertenecientes a la muestra escogida para formar los diferentes grupos de control y tratamiento permite asegurar que no haya diferencias sistemáticas entre los que van a recibir el tratamiento y los que no. Si esta selección aleatoria se lleva de forma adecuada, el resultado serán varios grupos estadísticamente iguales antes de llevar a cabo el tratamiento, lo que permite que tanto variables observables que podemos controlar como variables que no podemos observar o controlar (shocks externos, motivación u otros) se distribuyan de igual manera en los grupos. Este es uno de los métodos más seguros de acabar con el sesgo de selección, lo que resulta en que estos experimentos tienen una validez interna muy alta.

Otra gran ventaja de los ensayos controlados aleatorizados es la flexibilidad de estos, los conductores del experimento pueden diseñarlo con mucha libertad, ya que debido a su estructura metodológica permiten amoldarse a las hipótesis que estos quieran probar o preguntas que quieren responder. Se puede aleatorizar diferentes componentes de una intervención para observar el efecto de cada uno por separado y ver cuál es el más

efectivo. Esa posibilidad de diseñar un experimento para responder a una pregunta concreta es la que marca la diferencia.

Por ejemplo, había un debate muy controvertido entre los gurús de la oferta y de la demanda sobre si repartir mosquiteras para cama en países africanos para combatir la malaria. Los gurús de la demanda pensaban que si se repartían de forma gratuita estas mosquiteras se desperdiciarían recursos, la gente no las valoraría y no las utilizaría o la usaría para otras cosas como redes de pesca, además de que perturban al mercado y si se acostumbran a recibirlas gratis luego no estarían dispuestos a pagar por ellas. Esto se podría solucionar si se les hiciera pagar una pequeña parte del precio, para ahorrar recursos y que le den más valor, acostumbrándose a pagar por ello abriendo una puerta para un posible mercado de mosquiteras.

Los gurús de la oferta defendían lo contrario, al poner un precio a las mosquiteras la demanda de estas disminuiría notablemente, perdiendo la amplia cobertura que asegura el reparto gratuito y las externalidades positivas que ello conlleva, la disminución de la transmisión de enfermedades. Además, no hay pruebas concluyentes de que la utilización se vea afectado por el precio de compra de la red antimosquitos.

Para resolver la cuestión de si es más efectiva la distribución gratuita o el reparto de costes, Jessica Cohen y Pascaline Dupas (2008) diseñaron un experimento controlado aleatorizado en Kenia. Trabajaron con 20 clínicas sanitarias rurales en distintos distritos de Kenia especialmente afectados por la malaria, además del grupo control, crearon varios grupos de tratamiento en los cuales las mosquiteras se distribuían a diferentes precios a mujeres embarazadas que acudían a los centros, así podían observar la elasticidad precio de la demanda de las mismas. En una segunda etapa, se seleccionaron aleatoriamente 246 mujeres para ir a visitarlas y comprobar si habían utilizado apropiadamente las mosquiteras y si tenía alguna relación con el precio de compra.

Los resultados fueron claros. No se encontraron evidencias de que compartir el coste fomentara la mejor utilización de las mosquiteras, las mujeres que lo recibieron gratis les daban el mismo uso que las que pagaron un pequeño precio. Sin embargo sí que se encontraron evidencias de que aumentar el precio hacía caer la demanda de mosquiteras de manera importante, un 60% disminuye si pasa de ser gratis a costar 0.6\$, aunque

sigue estando por debajo del precio al que se vende de normal. Llevaron a cabo un análisis de coste-efectividad del impacto de estas mosquiteras en la mortalidad infantil que concluyó que la distribución gratuita es más efectiva dadas las externalidades positivas que se dan, y al fin y al cabo el coste de cada vida salvada sería menor que si se distribuyen con un pequeño precio.

Este ejemplo muestra en gran medida los puntos más positivos de este tipo de experimentos. Además normalmente no hay que hacer muchas suposiciones como en otros experimentos sobre la evolución de los grupos tratamiento o control, ya que al aleatorizar controlas hasta las variables no observables a simple vista. Aunque puede ser que dependiendo del diseño a veces haya que hacerlas, sobre todo en cuestiones como utilidad y otros valores difícilmente tangibles.

La menor necesidad de hacer suposiciones unido a la transparencia de los resultados (los resultados que se obtienen son poco manipulables), hacen de los experimentos controlados aleatorizados una herramienta muy útil para los responsables de llevar a cabo las políticas.

Por estos motivos, este tipo de ensayos han disfrutado de un gran aumento de su popularidad durante los últimos años. Se ha observado un importante incremento de artículos con datos microeconómicos que se llevaban a cabo con experimentos controlados aleatorizados, por ejemplo en 2008 de los seis artículos publicados usando datos microeconómicos en el *Quarterly Journal of Economics*, tres usaban la aleatorización. Y no solo en el mundo académico se ha extendido este entusiasmo, en 2009 de los 89 programas de evaluación que el Banco Mundial estaba llevando a cabo en África, 67 eran ensayos aleatorizados.

Un buen ejemplo de un artículo que haya causado un gran impacto no sólo en el mundo académico, sino que haya influido en la forma en que se diseñan las políticas públicas es el explicado en el punto 3.2 Impacto en la educación y en la salud de la desparasitación intestinal. El gran coste-eficiencia mostrado en ese artículo convenció a políticos de varios países para llevar a cabo este tipo de programas a gran escala. En 2009 el primer ministro de Kenia anunció un programa de desparasitación en la escuela que beneficiaría a 3 millones de niños, en 2014 el mismo programa llegó a los 6,4 millones

de niños, sobrepasando todas las metas. En algunas regiones de la India como Delhi, Bihar y Rajasthan ya han sido tratados más de 29 millones de niños gracias a la cooperación del gobierno de estas regiones con la iniciativa “Deworm the world”, que ayuda en varios países del mundo como Vietnam, Indonesia o el Congo a los gobiernos a poner en marcha este tipo de programas.

## CAPÍTULO V: CRÍTICAS

A todo auge le sigue inevitablemente una ola de críticas que ayudan a ver con mejor perspectiva el objeto de discusión, pues a veces al calor de los elogios los aspectos negativos son difíciles de discernir. La mayoría de las críticas a los ensayos controlados aleatorizados se centran en las limitaciones que estos tienen, ya sean metodológicos o de un carácter más teórico. Hay que tener en cuenta de que la mayoría de los economistas que utilizan esta herramienta son conscientes de las limitaciones que tiene. El debate surge a la hora de interpretar los límites de los resultados que se obtienen y las conclusiones que se sacan. A continuación voy a enunciar y explicar las principales críticas que diferentes economistas hacen a este tipo de experimentos.

### 5.1 Validez externa

La respuesta a la pregunta ¿si se repitiera el experimento en otro lugar con otras características tendría un resultado diferente? es la clave de la generalización de las posibles conclusiones, también llamado validez externa. Este es el mayor problema al que se enfrentan los ensayos aleatorizados puesto que al llevarse a cabo en un lugar con unas características determinadas resultará complicado advertir si en otras ocasiones funcionarían de igual manera. Este factor depende también del tamaño de la muestra, pues cuanto mayor sea esta, mayor será la validez externa pero debido al conocido “trade-off” entre validez externa e interna provocará una merma de esta última. Esto ocurre porque cuanto mayor sea la muestra del experimento más difícil es probar una relación causal fuerte (piensen en un experimento conducido en un país entero), debido a que el resultado obtenido es una media de cada participante y puede ocurrir que sea un resultado muy heterogéneo (este problema se trata más adelante).

Rodrik (2008) lo ilustra con el artículo de Dupas-Cohen “*Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment*” que he explicado en el capítulo anterior. A priori parece que el resultado no deja lugar a dudas y que el reparto gratuito de redes antimosquitos es la mejor política, pero tras analizar un poco más a fondo las condiciones del experimento surgen algunas dudas. En primer lugar, en la zona de Kenia donde se condujo el experimento ya se habían distribuido anteriormente redes antimosquitos junto con campañas que concienciaban de su uso, lo

que pudo beneficiar al experimento porque la gente estaba más predispuesta y/o ya sabía cómo utilizarlas. En segundo lugar, el experimento estaba diseñado para que los participantes fueran mujeres embarazadas, que tienen un mayor riesgo de ser afectadas por la malaria. Esto podría causar que estuvieran más interesadas en obtener ayuda de los servicios sanitarios. Por último, es posible que la diferencia entre el precio subsidiado y cero sea demasiado pequeña como para activar la falacia del costo hundido y no se haya podido comprobar adecuadamente. Estas cuestiones pueden poner en duda la generalización del resultado del experimento, pero corresponde al lector decidir cómo de importantes son en el experimento para haber podido afectar al resultado.

Banerjee y Duflo (2009) sostienen que para solucionar este problema se tienen que conducir experimentos que repliquen el original en diferentes lugares y con diferentes equipos para comparar los resultados y examinar si es posible generalizarlo. Hay ejemplos que han obtenido resultados positivos, como el Programa PROGRESA que expliqué anteriormente, pues se han llevado a cabo programas similares en otros países como Colombia, Ecuador y Honduras con parecidos resultados. Por otro lado también existen experimentos que no han podido replicar los resultados de los experimentos originales. En general Duflo y Banerjee asumen el problema de la generalización y defienden que es un problema inherente a los enfoques experimentales o cuasiexperimentales. Por ello hay que tenerlo en cuenta en cada ocasión e intentar probar y convencer de la validez externa.



## 5.2 Heterogeneidad en los resultados

Una de las características de este método experimental es que el efecto del tratamiento se indica con la media de los resultados. En parte es positivo porque simplifica mucho la interpretación de los resultados ya que comparas las medias y observas la diferencia y además no requiere suposiciones adicionales. Pero también conlleva una importante desventaja, a veces esa información no es suficiente para el autor del experimento ya que no informa por ejemplo de la distribución de los resultados. Al estimar la media de los resultados se supone que estos son más o menos homogéneos entre todos los individuos que han recibido el tratamiento, pero puede ocurrir que esto no sea así. Se pueden dar unos resultados con una desigual distribución entre la muestra que ha recibido el tratamiento. Banerjee et al. (2007) en un experimento en la India sobre cómo afectaba al rendimiento de los niños en las notas un tutor que les ayudara con las materias descubrieron que la mayoría de los beneficios los recibían los niños más retrasados, mientras que los que mejor iban tenían un beneficio considerablemente menor. Fueron capaces de observar este efecto porque contaban con resultados de exámenes pre-tratamiento y compararon las diferencias a través del tiempo.

Para solventar este problema hay que establecer unas hipótesis claras sobre cómo los efectos del tratamiento varían conforme la covarianza y recoger información relevante antes de llevar a cabo el experimento para después analizar los resultados y ver si se cumplen.

### 5.3 Relación con la teoría

El enfoque micro y experimental de los ensayos controlados aleatorizados también es objeto de crítica por numerosos autores. Para Acemoglu (2008) la teoría económica juega un papel fundamental en el desarrollo económico, principalmente porque ayuda a comprender los mecanismos económicos y además provee de una gran validez externa a diferencia de los enfoques más experimentales. Según el autor se debe desarrollar teoría económica para formular e intentar estimar los parámetros estructurales ya que estos pueden ayudar a estimar los efectos que distintas políticas o shocks tendrían en diferentes circunstancias y condiciones. Los enfoques experimentales no pueden estimar estos parámetros estructurales debido a las condiciones excepcionales en las que se conducen y por lo tanto sus resultados no se pueden generalizar sirviendo sólo para las condiciones concretas en las que se encuentran.

Otro factor fundamental según Acemoglu a favor de la teoría son los efectos del equilibrio general, se refiere a que los efectos de un programa conducido a pequeña escala pueden no ser generalizables cuando se escala a un nivel mayor. Esto se debe a que hay factores que se convierten en relevantes cuando se pasa de una escala pequeña a una grande debido a que en el análisis del equilibrio parcial se suponen constantes, pero al pasar al equilibrio general pueden cambiar. Dependiendo de la magnitud de estos efectos, las interacciones del equilibrio general pueden provocar que los resultados de trabajos experimentales de equilibrios parciales se diluyan o incluso sean contrarios.

Por ejemplo, consideramos un programa en el que niñas con problemas económicos reciben cupones para asistir a escuelas privadas, esto mejora su educación y sus ingresos futuros. Al escalar este programa a nivel nacional pueden aparecer dos problemas: masificación de las escuelas privadas (seguido del correspondiente colapso de las públicas) y un descenso en los retornos de la educación debido al incremento de la oferta. A priori no se conoce con seguridad los efectos de estos problemas, pero es posible que aparezcan y no se pueden ignorar al diseñar programas a una escala grande.

Banerjee y Duflo (2009) defienden que sí es posible llegar a estimar parámetros estructurales a partir de métodos experimentales. Gracias a la flexibilidad de estos en la recogida de datos y diseño para la implementación de tratamientos se puede diseñar un

experimento que tenga en cuenta la variación de algunos factores y asumir algunas suposiciones que ayuden a los investigadores a identificar los parámetros estructurales. Si bien esto puede ayudar a aumentar la validez externa, también tendría un efecto negativo en la validez interna como he explicado en el punto 5.1.

También reconocen el problema del equilibrio general, que aunque no se manifieste siempre, es un factor a tener en cuenta al escalar programas. Creen que en parte puede solucionarse si los experimentos se han preocupado por estimar los parámetros estructurales, a partir de ellos combinándolos con modelos de equilibrio microfundados se puede hacer una idea de los efectos del equilibrio general. Aunque reconocen que queda mucho por hacer para reducir la brecha entre el nivel micro y macro.

Donde sí pueden jugar un papel importante los ensayos controlados aleatorizados de la mano de la teoría es evaluando su funcionamiento. Gracias a la independencia de los resultados que se observan en los experimentos con respecto de la teoría, Banerjee y Duflo consideran que es la única manera de probar evidencia fuerte (junto con algunos experimentos naturales) “When we talk of hard evidence, we will therefore have in mind evidence from a randomized experiment, or, failing that, evidence from a true natural experiment, in which an accident of history creates a setting that mimics a randomized trial” (Banerjee 2007, 12). Gracias a esta evidencia fuerte hay un importante progreso en cuestiones de diseño de políticas y que es la mejor manera de ayudar a los países en vías de desarrollo, teniendo evidencia de que funcionarán.. También creen que la teoría podría ayudar al enfoque experimental en la medida que provea un marco de referencia para interpretar los resultados de estos experimentos que ayuden a sacar conclusiones más generales y mejore la validez externa.

Dani Rodrik (2008) es bastante crítico con estas ideas. En el punto 5.1 ya he explicado que las conclusiones de la distribución gratuita vs reparto de costes no son del todo definitivas. Cree que la idea de que sólo los experimentos aleatorizados y naturales pueden proveer evidencia fuerte es engañosa. La evidencia fuerte que corresponde a estos experimentos tiene muy limitado su alcance y aplicaciones, para que sea utilizable hay que complementarlo con muchas evidencias suaves.

Está clara la gran validez interna de este tipo de experimentos, pero al igual que un estudio con poca validez interna no es muy útil, uno con poca validez externa no lo es mucho más. Rodrik cree que los ensayos aleatorizados se podrían beneficiar en gran medida de la incorporación de teoría económica a la hora de establecer supuestos que ayuden a mejorar su validez externa.

## CAPÍTULO VI: CONCLUSIONES

En este trabajo he tratado de analizar el papel que tienen los ensayos controlados aleatorizados en el desarrollo económico. En el comienzo explico qué son y un breve resumen de su historia. Estos experimentos aleatorizados han tenido éxito en campos como la medicina debido a que permiten una fácil comparación para comprobar el efecto del tratamiento, pero en otros campos como la economía donde es muy difícil llevar a cabo los experimentos nunca han gozado de aceptación. La economía y en concreto el desarrollo económico estaban dominados por una visión teórica y generalista.

La incorporación de una visión diferente sobre la pobreza aportó las bases teóricas necesarias para abrir la puerta a otros enfoques en la lucha contra ella. El método experimental cobró importancia debido a su adecuación a la nueva perspectiva inductiva y de un carácter más microeconómico. La flexibilidad de diseño de los experimentos ayudaban a responder preguntas concretas sobre los problemas que causaba la pobreza. Además gracias a la sencillez de interpretación de los resultados no cabe lugar a trampas o prejuicios. La gran validez interna es otra de las ventajas de los experimentos aleatorizados, demuestran una causalidad con una evidencia fuerte.

Por otro lado, también tienen problemas inherentes a su diseño metodológico. He señalado la falta de validez externa que permita generalizar los resultados obtenidos a otras situaciones con características diferentes. Es un equilibrio delicado el de validez interna-externa. Algunos autores han propuesto medidas para compensar la falta de validez externa, como hacer suposiciones o disponer de un marco teórico que ayude a interpretar los resultados de una manera más genérica. También han aparecido los posibles problemas de la heterogeneidad en los resultados que pueden falsearlos si la muestra que recibe el tratamiento no es lo suficientemente homogénea.

Una cuestión crucial es el papel de los métodos más teóricos en contraposición a los experimentales en el desarrollo económico. En mi opinión creo que las dos concepciones se pueden beneficiar mutuamente. La teoría necesita llevar a cabo

experimentos para probar sus conclusiones y los experimentos necesitan un marco teórico para sustentar sus resultados.

Para concluir, en mi opinión la aparición de los ensayos aleatorizados en el campo del desarrollo económico ha sido muy positiva. En la actualidad este método experimental constituye una herramienta fundamental para los científicos sociales. Pero no sólo ha ayudado desde el punto de vista metodológico. Igual o incluso más importante ha sido su aportación al cambio de mentalidad a la hora de pensar en los problemas de la pobreza. La convergencia hacia una visión pragmática, experimental y contextual tanto desde una perspectiva macro como micro pueden ayudar en un futuro a no cometer los mismos errores que en el pasado. Ejemplos como las reformas propugnadas por el Consenso de Washington obedecen más a ídoles ideológicas que racionales. Un nuevo enfoque del desarrollo económico donde se tenga en cuenta las características propias de cada país y las políticas estén diseñadas con motivaciones funcionales y no ideológicas es necesario si de verdad queremos atajar los problemas de la pobreza y el desarrollo.

## CAPÍTULO VII: BIBLIOGRAFÍA

Acemoglu, D. (2010). “Theory, General Equilibrium and Political Economy in Development Economics”. MIT Press.

Banerjee, A. Duflo, E. (2011). *Repensar la pobreza. Un giro radical en la lucha contra la desigualdad global*. Taurus.

Banerjee, A. Duflo, E. (2009). “The experimental approach of development economics”. *Annu. Rev. Econ.* 2009. 1:151–78

Banerjee, A. et al. (2007). “Making aid work”. MIT Press.

Banerjee, A. et al. (2005). “Remedying education: evidence from two randomized experiments in India”

Banerjee, A. (2006). *Understanding poverty*. Oxford University Press

Blog Free exchange (2013). “Coming of an age”. *Semanario The Economist*. Edición digital. Publicado el 12 de Diciembre de 2013.

<http://www.economist.com/blogs/freeexchange/2013/12/randomised-control-trials>

Blog Free exchange (2013). “Random harvest”. *Semanario The Economist*. Edición digital. Publicado el 14 de Diciembre de 2013.

<http://www.economist.com/news/finance-and-economics/21591573-once-treated-scorn-randomised-control-trials-are-coming-age-random-harvest>

Chattopadhyay, R. Duflo, E. (2004). “ Women as policymakers: evidence from a randomized policy experiment in India”. *Econometrica*, Vol. 72, No. 5 (September, 2004), 1409–1443.

Cohen, J. Dupas, P. (2008). “Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment”. *The Quarterly Journal of Economics*, Vol. 125, (February 2010), Issue 1.

Deaton, A. (2009). “Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development”. National Bureau of Economic Research.

Duflo, E. Glennerster, R. Kremer, M. (2007). “Using randomization in development economics research: a toolkit”. Centre for Economic Policy Research.

Duflo, E. Kremer, M. Robinson, J. (2010). “ Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya”. *American Economic Review* 101(6): 2350-90

Easterly, W. (2006). *The white man's burden*. Penguin Books.

Easterly, W. (2008). "Can the West save Africa?". National Bureau of Economic Research.

Glennerster, R. Takavarasha, K. (2013). *Running randomized evaluations a practical guide*. Princeton University Press.

Kremer, M. Miguel, E. (2004). "Identifying impacts on education and health in presence of treatment externalities". *Econometrica*, Vol. 72, No. 1 (January, 2004), 159–217.

Kremer, M Zwane, A. (2014). "Deworming the world for better schools, brighter futures". World Economic Forum. Publicado el 18 de Diciembre de 2014.

<https://www.weforum.org/agenda/2014/12/deworm-the-world-schools-programme-kremer-zwane/>

Rodrik, D. (2008). "The new development economics: we shall experiment, but how shall we learn? Harvard University.

Sachs, J. (2005). *The end of poverty*. Penguin Books.

Schultz, T. P. (2000). "School subsidies for the poor: evaluating a mexican strategy for reducing poverty". International Food Policy Research Institute.

Shaban, R. (1987). "Testing between Competing Models of Sharecropping". *Journal of Political Economy*, Vol. 95, No. 5 (Oct., 1987), pp. 893-920