LA NUEVA ECONOMÍA DEL DESARROLLO: UNA APUESTA EXPERIMENTAL EN LA BATALLA CONTRA LA POBREZA

Juan David Correales Ceballos

Universidad del Valle
Facultad de Ciencias Sociales y Económicas
Departamento de Economía
Santiago de Cali
2016

LA NUEVA ECONOMÍA DEL DESARROLLO: UNA APUESTA EXPERIMENTAL EN LA BATALLA CONTRA LA POBREZA

Juan David Correales Ceballos

TRABAJO DE GRADO PARA OPTAR POR EL TÍTULO DE ECONOMÍSTA

Director**:**
Boris Salazar
Profesor titular del Departamento de Economía
Universidad del Valle

Universidad del Valle
Facultad de Ciencias Sociales y Económicas
Departamento de Economía
Santiago de Cali
2016

LA NUEVA ECONOMÍA DEL DESARROLLO: UNA APUESTA EXPERIMENTAL EN LA BATALLA CONTRA LA POBREZA

RESUMEN

Este trabajo tiene como principal objetivo realizar un análisis del enfoque J-PAL en la economía del desarrollo, efectuando un estudio riguroso de su herramienta insignia: los Ensayos Controlados Aleatorios (ECA). Cuestionando tanto la validez de los resultados como las inferencias que se desprenden de los experimentos. El trabajo se divide en tres capítulos, donde se trata el trasfondo histórico de los ECA, la utilización de la aleatorización como un elemento clave en la identificación causal, fundamentándose en el modelo de Neyman-Rubin-Holland, y su incursión en la economía del desarrollo como estándar de oro de los métodos de evaluación de impacto liderado por el Laboratorio de Acción Contra la Pobreza, J-PAL. A partir de este estudio se determinó que la metodología ECA tiene deficiencias tanto a la hora de hallar efectos causales como en la generalización de los resultados.

Palabras Clave: Economía del desarrollo, Ensayos Controlados Aleatorios, Causalidad, Validez Interna y Validez Externa.

Clasificación JEL: O21, O20, I28, I3, I32, I38

ABSTRACT

This paper's main objective is a J-PAL approach analysis in develoment economics, by carring out a rigorous study of its flagship tool: randomised controlled trials (RCTs). Questioning both the validity of the results and inferences derived from experiments. The paper is divided into three chapters, where the historical background of RCTs, is the use of randomization as a key element in identifying causal, based on the model of Neyman-Rubin-Holland, and his foray into the economics of development as a golden standard led by the laboratory of action against poverty impact assessment methods , J-PAL. From this study it was determined that the ECA methodology has weaknesses both when it comes to finding causal effects as in the generalizability of the results.

Key words: Develoment Economics, Randomised Controlled Trials, Causality, Internal Validity and External Validity

Contenido

TABLA DE ILUSTRACIONES

INDICE DE TABLAS

Introducción

"Esperamos convencer al lector de que nuestro enfoque, paciente y basado en el avance paso a paso, no solamente es una vía útil para luchar contra la pobreza, sino también un camino para hacer del mundo un lugar más interesante"[1]

En el año 2010, durante las reconocidas conferencias TED[2] (Tecnología, Entretenimiento y Diseño) que se realizan anualmente, se presentó una innovadora idea que transformaría la visión que se tiene sobre la pobreza y sobre cuál es la mejor manera de enfrentarla. En la conferencia titulada *"Experimentos sociales para luchar contra la pobreza"*[3], se observó no sólo un giro radical en el estudio de la pobreza sino también un profundo cambio en la manera en que se investiga desde la ciencia económica.

El público de TED que tradicionalmente recibe buenas noticias de los avances por parte de la ciencia y la tecnología, quedó atónito ante las primeras palabras de una menuda mujer de origen francés al decir: "todos los días, 25.000 niños mueren por causas totalmente prevenibles. Eso es un terremoto en Haití cada ocho días"[4]. Con estas estremecedoras palabras, la economista francesa Esther Duflo, les recordaba a los espectadores lo perturbador que es vivir en un mundo donde la pobreza y la miseria son causantes de la pérdida de miles de vidas año tras año.

Pero a pesar de esta realidad horripilante, lo cierto es que hoy la batalla contra la pobreza se maneja con cierto escepticismo. En primer lugar, porque el tema de la pobreza se percibe como un problema monumental e inabordable a tal punto de presentarse como una tarea de

[1] (Banerjee & Duflo, 2011, p.35)
[2] TED, bajo el emblema "ideas dignas de ser difundidas" es una organización sin ánimo de lucro que tiene como objetivo principal difundir excelentes ideas y potenciar el poder de las mismas para cambiar actitudes, vidas y en última instancia el mundo. Es un evento de una constante calidad e interés, TED realiza un congreso anual (TED Conference) y diferentes charlas (TED Talks) que cubren un amplio abanico de temas que incluyen ciencias, política, educación, cultura, negocios, asuntos globales y tecnología, entre otros.
[3] Duflo, E. (2010). *Experimentos sociales para luchar contra la pobreza.* [Video Web]. Congreso llevado a cabo en Long Beach, CA. Recuperado de: http://www.ted.com/talks
[4] Ibíd.

nunca acabar. Y en segundo lugar, porque no se tiene certeza de que se estén haciendo las cosas correctas en la lucha contra la pobreza.

Por esta razón, Duflo (2010) refiriéndose a los economistas y actores políticos, durante la conferencia dijo: "la cuestión es, que si no sabemos si estamos haciendo algún bien, no somos mejores que los médicos medievales y sus sanguijuelas. A veces el paciente mejora, a veces el paciente muere. ¿Son las sanguijuelas? ¿Es algo diferente? No lo sabemos". Desde luego, se refiere a la situación de no tener suficiente evidencia a favor de lo que funciona, o a la inversa, de lo que precisamente no funciona para promover el desarrollo económico.

Con el propósito de encontrar respuestas específicas a problemas concretos, el equipo de investigadores de J-PAL ha introducido a la economía del desarrollo un instrumento nuevo proveniente de la medicina: los ensayos controlados aleatorios (ECA). Básicamente, este método consiste en realizar experimentos de campo con la intención de evaluar el impacto de los diferentes programas del desarrollo, y con base en ello, hacer más efectiva la lucha contra la pobreza. El uso de este método ha generado un verdadero impacto entre los economistas del desarrollo, y en general, en la ciencia económica y el ámbito político (Baele, 2013). En los últimos años, el número de investigaciones, artículos y libros que incluyen el método experimental, crece de manera exponencial. Angrist y Pischke (2010) hablan incluso de una "*revolución empírica*".

Este enfoque incluso se ha institucionalizado, desde la creación en el año 2003 del "*Laboratorio de Acción Contra la Pobreza Abdul Latif Jameel*" (J-PAL) del Instituto Tecnológico de Massachusetts (MIT), donde se lideran cientos de proyectos en todo el mundo en contra de la pobreza, desde una mirada común que es por supuesto la economía experimental. Por lo tanto, el siglo XXI está siendo testigo de cómo nace una nueva economía, la cual Rodrik (2008) denominó la *"**Nueva economía del desarrollo**"*.

Sin embargo, a la luz de la filosofía de la ciencia han surgido muchos interrogantes con respecto a cómo se genera y valida el conocimiento producido por estos nuevos desarrollos. Este fenómeno también denominado "*revolución experimental*" se ha basado en una serie de premisas filosóficas que se han mantenido hasta ahora en su mayoría sin explorar, y que

por lo tanto, exigen una evaluación (Guala, 2005). Por consiguiente, este trabajo tiene como principal objetivo realizar un análisis del enfoque J-PAL en la economía del desarrollo, efectuando un estudio riguroso de su herramienta insignia: los ensayos controlados aleatorios. Cuestionando tanto la validez de los resultados como las inferencias que se desprenden de los experimentos.

El presente trabajo se ejecuta bajo una metodología argumentativa, y es de carácter monográfico o documental, desde una revisión de literatura de tres campos: la economía del desarrollo, la economía experimental y la metodología de la economía. Por lo tanto, el trabajo utiliza fuentes bibliográficas que orientan la investigación, y de esa manera, sustentan los argumentos y las conjeturas que en el documento se realicen.

Este trabajo se divide en tres capítulos. En el capítulo 1 se encuentra la historia de los Ensayos Controlados Aleatorios, investigando el porqué de la aleatorización en el diseño experimental y su incursión en la economía del desarrollo. En el capítulo 2 se encuentra el marco analítico que justifica la realización de los experimentos en la economía y su expansión a través de J-PAL. Por último, en el capítulo 3 se muestra el análisis de la validez interna y externa del método ECA en la práctica, mostrando tanto sus virtudes como limitaciones.

Problema de investigación

¿Cómo crecen y cambian las economías? Esta fue la pregunta principal que persiguieron los economistas clásicos, considerada como la primera escuela económica moderna, que derivó en los trabajos más influyentes de la disciplina económica con autores como Adam Smith, David Ricardo y Marx (Stern, 2002). No obstante, con la llegada de la "revolución marginalista" en los años 1870, el interés por los determinantes de la riqueza (lo que es hoy, el estudio del crecimiento económico) llegó a su fin. Después de la Segunda Guerra Mundial, los llamados "pioneros" de la economía del desarrollo reconocieron la herencia de la economía clásica y retomaron la senda de la investigación del crecimiento y el desarrollo económico.

Ya han pasado más de sesenta años, y dos generaciones de economistas del desarrollo han dejado un fuerte legado a esta rama de la economía. Sin embargo, el nuevo milenio

tiene preparado lo que es una tercera generación dispuesta a cambiar la práctica y doctrina de los economistas del desarrollo. Esther Duflo y Abhijit V. Banerjee, ambos profesores del Instituto de Tecnología de Massachusetts (MIT), junto a muchos colaboradores, han propiciado una verdadera explosión de la experimentación a través de la asignación aleatoria en el campo de la economía del desarrollo (Banerjee & Duflo, 2009).

Los Ensayos Controlados Aleatorios (ECA) es un método proveniente de las ciencias de la salud que fue "importado" hace aproximadamente quince años a la economía y constituye la herramienta principal de análisis del movimiento "Medicina Basada en la Evidencia", que en efecto es el precursor del movimiento "Política Basada en la Evidencia" dentro del cual se sitúa el enfoque J-PAL. La evaluación aleatoria es un tipo de evaluación de impacto que utiliza un proceso aleatorio con el propósito de evaluar programas de acción a favor del desarrollo.

Existen actualmente varios métodos para esta finalidad, pero las evaluaciones aleatorias son generalmente consideradas como las más rigurosas, ya que producen los resultados más precisos. Esto le ha brindado un estatus de "estándar de oro" en las evaluaciones de impacto y la denominación de "evidencia dura", proporcionando una cierta "hegemonía metodológica", como lo describe Salomone (2012), en este tipo de evaluaciones. Es ahí donde se concentra esta investigación, en determinar la validez de las anteriores afirmaciones y en encaminar un análisis exhaustivo acerca del impacto del enfoque J-PAL lanzado por Duflo y su equipo sobre la economía, la política económica del desarrollo, y en mayor importancia sobre la comprensión del fenómeno de la pobreza en el mundo entero.

Por lo tanto, es momento de empezar, y con especial énfasis en lo que se habrá de llamar el problema metodológico de J-PAL, que constituirá el primer eje de esta investigación. Con este propósito es pertinente empezar por definir qué es un ECA en su forma más básica; Banerjee & Duflo (2011) lo explican de la siguiente manera:

> En un ECA, como en los estudios sobre mosquiteros, las personas o las comunidades son asignadas de forma aleatoria a distintos tratamientos, es decir, a programas diferentes o a versiones distintas de un mismo programa. Puesto que los diferentes tratamientos se suministran a individuos que son exactamente comparables (ya que

> fueron elegidos al azar), cualquier diferencia posterior entre ellos constituye el efecto del tratamiento (p.32)

De este modo, los investigadores de J-PAL están interesados por el *efecto causal* promedio de una intervención, el cual es la diferencia de las medias observadas que se presentan al final de cada experimento. Lo anterior, los introduce al problema de la inferencia causal del cual son conscientes. Por lo cual, este tipo de investigación busca responder a este tipo de preguntas:

> ¿Cuál es el efecto causal de la educación sobre la fertilidad? o ¿Cuál es el efecto causal del tamaño de la clase sobre el aprendizaje? Lo que requiere contestar preguntas esencialmente contrafactuales: ¿Cómo les habría ido a las personas que participaron en un programa si hubiesen estado ausentes del mismo? ¿Cómo, aquellos que no fueron expuestos al programa les habría ido en presencia del mismo? La dificultad de estas preguntas es inmediata. (Duflo, Glennerster y Kremer, 2006, p.7)

Esta noción de causalidad se expresa en el modelo de Rubin (1974), que es la base y justificación de las evaluaciones aleatorias y de la pretensión de J-PAL en encontrar causalidades significativas. Y es aquí, donde entra en escena el concepto de identificabilidad, el cual hace referencia a que el grupo de tratamiento y control deben ser completamente idénticos, es decir, que haya una representación adecuada de la población que se desea estudiar, de lo contrario generaría un "sesgo de selección" y el efecto del impacto no sería plenamente identificado (Cacheiro, 2011).

Con la aleatorización -o lo que es lo mismo, la asignación al azar- J-PAL afirma que se elimina el problema del sesgo de selección, por lo cual se sitúa en un plano superior frente a los estudios observacionales los cuales no controlan el mecanismo de asignación del tratamiento y deben hacer uso de otros mecanismos que conllevan a un mayor número de supuestos. La presencia o ausencia del sesgo de selección compromete la validez o invalidez interna del estudio, y es por esta razón que los ECA sean calificados de: "Estándar de Oro" y "Evidencia dura y rigurosa".

Debido a lo anterior, los ECA han adquirido una relativa *supremacía* frente a otros métodos como los cuasi-experimentales[5] y los estudios econométricos tradicionales que se utilizan a su vez en ese mismo propósito (evaluaciones de impacto). Esta supremacía, ha sido alimentada por un general escepticismo acerca de la capacidad del análisis econométrico para hacer frente a la causalidad (Cohen & Easterly, 2009). Los viejos problemas de identificación y baja validez interna de los estudios observacionales, al parecer, han sido resueltos por el método ECA. En palabras de Deaton (2009):

> Ciertamente, los estudios econométricos que utilizan la evidencia internacional para examinar la eficacia de la ayuda tienen hoy bajo estatus profesional. (p.4)

En este sentido, el enfoque J-PAL ha desarrollado lo que podría denominarse una "Jerarquía de la evidencia" no sólo en las evaluaciones de impacto sino en todo el campo de la economía del desarrollo (Labrousse, 2010). Para Duflo y compañía, básicamente la evaluación aleatoria es el mejor método y se sitúa en la parte superior de la evidencia, seguido de los diseños cuasi-experimentales que simulan a su vez la aleatorización. Es por ello que Rodrik (2008) señala que:

> Los "randomistas" (como Deaton [2007] los ha llamado) tienden a pensar que la evidencia creíble puede ser generada únicamente con ensayos de campo aleatorios o cuando la naturaleza coopera proporcionando la oportunidad de un experimento natural. (p.3).

A lo cual Banerjee (2007) ratifica:

> Cuando hablamos de pruebas contundentes, por lo tanto, tendremos en cuenta la evidencia de un experimento aleatorizado o, en su defecto, la evidencia de un verdadero experimento natural en la que un accidente de la historia crea un entorno que imita un ensayo aleatorio. (p.12)

Duflo (2009) añade:

[5] Entre los métodos cuasi-experimentales los más usados son: Pre-Post, Simple Difference, Differences in Differences, Multivariate regression, Statistical Matching, Regression Discontinuity Design, Instrumental Variables.

> Simplemente no es posible aislar los mecanismos subyacentes de crecimiento económico como única guía de las experiencias de crecimiento de un centenar de países. Cualquier variable puede ser causa o efecto, o podría explicarse por una tercera variable correlacionada con las otros dos. [...] Tenemos que llegar a lo obvio, incluso con "dos millones de regresiones", no vamos a llegar a descubrir el secreto del crecimiento de una experiencia pasada sobre la base de datos de un centenar de países. (p.23)

No obstante, Deaton (2009) ha reprobado dicha postura:

> Los ensayos controlados aleatorios no ocupan ningún lugar especial en cierta jerarquía de la evidencia, tampoco tiene sentido referirse a ellos como evidencia "dura" mientras que otros métodos son de evidencia "blanda". Estos dispositivos retóricos son sólo eso; una metáfora no es un argumento. (p.6)

Por lo tanto, esta primera parte del problema metodológico de J-PAL es, inequívocamente, una pieza fundamental del rompecabezas de la presente investigación, y pretende responder a las siguientes preguntas: ¿Ha resuelto ECA los problemas estadísticos de la inferencia causal que aquejan a la econometría tradicional? ¿Los ECA generan información sobre la causalidad, en verdad lo logran? ¿Producen la mejor evidencia para la ayuda al desarrollo económico? ¿Existe una "jerarquía de la evidencia" en la economía del desarrollo que diferencia entre evidencia dura y blanda? ¿Cuál es la evidencia admisible en la economía del desarrollo?

Por otro lado, la segunda parte del problema metodológico de J-PAL, que también conforma el primer eje de esta investigación, se refiere a una contradicción destacada por Favereau (2014), la cual se había venido gestando con autores como Guala (2005), Cartwright (2007), Rodrik (2008), Deaton (2009), Ravallion (2009), Cohen & Easterly (2009) y Salomone (2012), entre otros. Para entender a plenitud dicha contradicción, es menester presentar, en primer lugar, la misión del enfoque J-PAL:

> Nuestra misión es reducir la pobreza y mejorar la calidad de vida de las personas, creando y difundiendo evidencia rigurosa sobre qué políticas públicas y programas

sociales realmente funcionan. Esto lo hacemos a través de investigación, incidencia en política pública y capacitación. [6]

De este modo, los dos objetivos principales de J-PAL se pueden resumir en: (1) Generar evidencia rigurosa de la efectividad de los programas; (2) Guiar las decisiones de política a partir de sus resultados experimentales. Es por ello que, el lema de J-PAL es "Traduciendo la investigación en acción", que va en concordancia con la afirmación de que la acción política debe estar basada en la evidencia. Ahora bien, antagónicamente a este discurso profesado por J-PAL, emerge un problema metodológico significativo que se expresa entre el concepto de validez interna y externa que no solamente debilita los objetivos anteriormente mencionados, sino que los convierte en contradictorios.

La validez interna, se refiere a la calidad de la identificación causal, es decir, que el estudio ha demostrado de manera admisible una relación causal entre el tratamiento en cuestión y el resultado de interés (Rodrik, 2008). Por ejemplo, que el impacto de la educación sobre la fertilidad sea plenamente identificado con el menor sesgo posible. Por tanto, la validez interna asegura que la variación estadística observada es causada por el programa y no por otros factores (Favereau, 2014). Mientras que, por validez externa se entiende como la capacidad de predecir o extrapolar la estimación de la muestra a una población, es decir, el alcance al que se pueden generalizar o extender los resultados obtenidos a otros lugares o personas.

En ese orden de ideas, los ECA son reconocidos por su fuerte validez interna gracias a la aleatorización, no obstante, su validez externa ha sido duramente cuestionada, pues la pregunta es si el resultado particular de un programa en particular, llevado a cabo en una población en particular, en un determinado país por un particular, se puede aplicar en otras circunstancias (Cohen & Easterly, 2009). Por el contrario, los enfoques econométricos son más débiles en validez interna, pero con menos problemas de validación externa. En vista de ello, Rodrik (2008) indica:

[6] Información extraída de la página oficial del Laboratorio de Acción Contra la Pobreza Abdul Latif Jameel: www.povertyactionlab.org

> La identificación por sí sola no es suficiente, a menos que se les dé una fuerte razón para creer que los efectos causales pueden generalizarse a la población más amplia de interés. Un estudio que carece de validez interna sin duda vale nada; pero un estudio que carece de validez externa es casi inútil también. (p.18)

Luego, existe un denominado "trade off" entre validez interna y externa que ha sido ampliamente detallado por el trabajo de Guala (2005), en que la promoción de la validez interna de una investigación compromete su validez externa y viceversa. Por lo tanto, como lo ha descrito Favereau (2014):

> Esta tensión entre la validez interna y externa socava los dos objetivos de J-PAL, haciéndolos contradictorios. Esta tensión o contradicción define lo que voy a llamar una falla epistemológica. Esta falla epistemológica, parece explicar la dificultad de J-PAL en producir recomendaciones de políticas claras y su dificultad en la producción de una teoría unificadora (p.18)

Cartwright (2007), ya lo habría expresado unos años antes al decir:

> El beneficio que las conclusiones siguen deductivamente en el caso ideal (de un ECA) viene con un gran costo: estrechez de alcance (P.5).

Llegado a este punto, la línea de investigación del presente trabajo, reconstruye los análisis críticos de los autores mencionados y los encamina hacia un examen metodológico y epistemológico, que como se observó son claves para determinar el futuro de esta nueva doctrina económica. A su vez, la pregunta esencial que habrá de resolver será la siguiente: ¿Cuáles son los alcances, los límites y las posibilidades del conocimiento de la investigación experimental de J-PAL, sopesando la idea de una falla epistemológica?

Siguiendo la línea de este análisis, es pertinente denotar el segundo eje que guiará el problema de investigación. En concordancia, esta nueva generación de economistas del desarrollo se diferencia de las anteriores de forma tajante. Una primera diferencia es la forma de abordar el problema del desarrollo económico. Para Banerjee & Duflo (2011) hay que cambiar, en primer lugar, las grandes preguntas por otras más sencillas y más pequeñas. En su libro, *Repensar la pobreza*, lo plasma de la siguiente manera:

> Este libro es una invitación a pensarlo bien (otra vez), a dejar de un lado la sensación de que la lucha contra la pobreza es demasiado abrumadora y a empezar a pensar en ella como un conjunto de problemas específicos que, una vez identificados y comprendidos, pueden ser resueltos de uno en uno. (p.19)

Así pues, el enfoque J-PAL aboga por un enfoque microeconómico de los problemas del desarrollo y la pobreza, pues en lugar de partir de los fenómenos macroeconómicos, a Banerjee & Duflo (2011) les es más prometedor comenzar con modelos y estimaciones microeconómicas y utilizarlos como piezas para construir un modelo macroeconómico. De esta forma, los investigadores de J-PAL parten del terreno, en situaciones reales y no simuladas, como en el caso de la experimentación de laboratorio, y evalúan empíricamente a escala microeconómica los efectos de las decisiones y de los programas contra la pobreza (Guillen, 2015).

Este cambio de perspectiva frente a las anteriores generaciones es desafiante, pues la primera generación formuló grandes modelos y teorías, que incluían estrategias de desarrollo mediante transformaciones estructurales con fuerte participación del gobierno, eran una clara programación del desarrollo (Meier, 2002). Entre los trabajos más prominentes se encuentran Rosenstein-Rodan (1943), Prebisch (1950), Chenery (1952), Rostow (1960), Nurkse (1961), entre muchos otros. Por otra parte, la segunda generación estuvo apoyada por los principios fundamentales de la economía neoclásica que tenía como premisa reducir la participación del gobierno, asimismo durante esta generación se multiplicaron los estudios econométricos sobre los determinantes del crecimiento económico (Harberger, 1993).

En cambio, el enfoque J-PAL no está fundamentado en grandes teorías o apoyadas en conceptualizaciones filosófico-matemáticas (Clavijo, 2012), pues para estos investigadores el principal interés está en *qué funciona o no* para combatir la pobreza basados en un enfoque experimental, dejando así de lado las discusiones teóricas. Está, en sí misma, es una primera crítica, Hurtado (2014) menciona que con este cambio de enfoque se pasó a "tener grandes teorías frente a ninguna teoría" (p.5). A esta discusión, Deaton (2009) expresa que "es verdad que no siempre es evidente como combinar la teoría con experimentos", razón por la cual también agrega:

> Los empiristas y teóricos parecen más separados que en cualquier período en el último cuarto de siglo. Sin embargo, la reintegración es apenas una opción porque sin ella no hay posibilidad de progreso científico a largo plazo. (p.45)

Lo anterior, le da importancia a la pregunta expuesta por Meier (2002) en este análisis: ¿Debe considerarse la economía del desarrollo simplemente como "economía aplicada", o hay la necesidad de una "teoría del desarrollo especial"? O en palabras de Rodrik (2008) "tendremos experimentos, pero ¿Cómo vamos a aprender?" (P.1). Esto ha encendido una acentuación de la "divergencia metodológica" entre los macro y micro economistas del desarrollo que amenaza con eclipsar la convergencia en políticas (Rodrik, 2008). De esta manera, este segundo eje de la investigación apunta a cuestionar la contribución realizada por el enfoque J-PAL a los debates teóricos, la pregunta es: ¿ECA ofrece una nueva comprensión de los fenómenos de la pobreza?

CAPÍTULO 1: Aleatorización, experimentos de campo y economía del desarrollo

"*La creación de una cultura en la cual rigurosas evaluaciones aleatorias son promovidas, alentadas y financiadas tiene el potencial de revolucionar la política social durante el siglo XXI, así como los ensayos aleatorios revolucionaron la medicina durante el siglo XX.*"
(Duflo, 2004 citado por Deaton, 2009, p.3)

1.1 Introducción

La evolución de las Evaluaciones Aleatorias atraviesa una extensa y fascinante odisea. Desde el primer ensayo clínico realizado por James Lind en el año de 1747 hasta los primeros experimentos de campo aleatorios en la economía del desarrollo, llevados a cabo por Edward Miguel y Michael Kremer en 1998, puede apreciarse la gran trayectoria realizada por el método ECA, la cual ha transitado y se ha desarrollado en campos como el de la psicofísica, la investigación agrícola, la medicina y la economía. Esta historia compartida, ha enfrentado, entre otras cosas, grandes retos y desafíos metodológicos.

Este primer capítulo, además de abordar el contexto histórico del método ECA, tiene como propósito exponer los conceptos básicos de aleatorización, experimentos de campo y evaluaciones de impacto que servirán como base para las discusiones que tendrán lugar más adelante. Es así, que las preguntas fundamentales que guiarán este primer capítulo se centran en: ¿Cuándo comenzaron las evaluaciones aleatorias? ¿Qué es la aleatorización? ¿Por qué aleatorizar en la economía?

Las respuestas a estas preguntas son la base del enfoque asumido por J-PAL, el cual ha institucionalizado el uso y la promoción de la asignación al azar. Según los investigadores de J-PAL, la aleatorización resuelve los problemas de identificación y de sesgo de selección, problemas que se habrán de detallar ampliamente en este capítulo y que son claves para el problema de investigación.

Este capítulo se ha organizado en cinco secciones. En la primera, que pretende poner en perspectiva la discusión de los temas tratados en esta investigación, se hace referencia al origen de los ensayos clínicos y de la asignación al azar, utilizando los trabajos de James Lind y Charles Sanders Peirce. El primer autor introdujo el concepto de grupo experimental y de control en un experimento, mientras que el segundo, realizó por primera vez un experimento utilizando la asignación al azar.

La segunda sección está dedicada al trabajo de Ronald Fisher quién condujo las primeras pruebas aleatorias en experimentos agrícolas. Los experimentos de Fisher culminaron en su trabajo más conocido, *el diseño de los experimentos*, que provocó en gran medida el crecimiento de las evaluaciones aleatorias. Por otra parte, en la tercera sección, se habla del primer ensayo clínico aleatorizado, que se realizó de la mano de Austin Bradford Hill, el cual buscaba establecer una relación causal entre la estreptomicina y la tuberculosis.

En la cuarta sección se discute sobre el movimiento de la "Medicina basada en la evidencia" antecesor respectivamente del movimiento "Política basada en la evidencia". Esta sección es fundamental para establecer las conexiones existentes entre ambos movimientos empíricos, que promueven una "jerarquía de la evidencia" donde los ensayos controlados aleatorios se configuran como el mejor método y, paralelamente, como la mejor evidencia.

Por último en la quinta sección, se expone como tuvo lugar la evolución y el ascenso del enfoque experimental de J-PAL, mencionando sus primeros experimentos y definiendo sus elementos principales como diseño de investigación. En este sentido, esta sección pretende descifrar como ha llegado el método ECA a un estatus de "estándar de oro" en las evaluaciones de impacto y la concedida denominación de evidencia "dura y rigurosa".

1.2 Origen de los ensayos clínicos y de la asignación al azar

El origen de los ensayos controlados aleatorios (ECA) se deriva principalmente del encuentro de dos historias. Por un lado, lo que sería la historia principal, que es la historia relacionada con el surgimiento de los ensayos clínicos en la medicina, y por el otro lado, la historia de la inserción de la aleatorización en el diseño experimental. El empalme entre ambas historias constituye la génesis del enfoque J-PAL.

James Lind es considerado hoy como el precursor de los ensayos clínicos de la era moderna (Dodgson, 2006). Lind, como cirujano naval vinculado a la Marina de Gran Bretaña durante la guerra de Sucesión de Austria en el año de 1747, realizó un estudio comparativo de seis tratamientos diferentes buscando la cura para la enfermedad del

escorbuto[7] que afectaba considerablemente a los marinos. Para Bhatt (2010) la descripción de dicha prueba cubre los elementos esenciales de un ensayo controlado. En su libro, *Un tratado del escorbuto*, Lind explica su experimento:

> El 20 de mayo de 1747, tomé doce pacientes con escorbuto al abordar el buque *Salisbury* en el mar. Sus casos eran tan similares como yo podría tenerlos. Todos en general tenían encías putrefactas, manchas, lasitud y debilidad en sus rodillas. Los puse juntos en un sólo lugar al ser un cuarto adecuado para los enfermos y tenía una dieta común a todos, […] A dos de ellos se les ordenó a cada uno beber 1,1 litros de Sidra al día. Otras dos personas tomaron veinticinco mililitros de vitriolo (ácido sulfúrico diluido) tres veces al día en ayunas, haciendo fuertes gárgaras en sus bocas. Otras dos personas tomaron 18 mililitros de vinagre tres veces durante el día antes de las comidas. Dos de los pacientes más graves, se les dio un cuarto de litro de agua de mar. Para otras dos personas, se les dio a beber zumo de naranja y limón durante seis días. Los dos pacientes restantes tomaron una pasta medicinal compuesta de ajo, mostaza, y rábano de raíz seca. (Lind, 1753, p.1)

La atención se fijaba pues en determinar el impacto de algún tratamiento médico para combatir la enfermedad del escorbuto, lo cual inmiscuía a la investigación médica hacia una pregunta de carácter estadístico: ¿Cómo cuantificar la magnitud de dicho impacto? Pues es sabido, por ejemplo, que en el caso del experimento de Lind cualquier precondición del paciente podía interferir en el resultado final de cada tratamiento, dejando sin saber a ciencia cierta cuál había sido el impacto real.

No obstante, el informe de James Lind, permite vislumbrar un aspecto esencial y es su conciencia de la necesidad de comparar grupos similares para permitir inferencias confiables. Hoy en día, se conoce como control al sesgo de selección. Y por supuesto, es evidente la forma en que trató de contener posibles factores externos que podrían alterar los

[7] La Real Academia Española define la enfermedad del escorbuto como una avitaminosis producida por la deficiencia de vitamina C, que es requerida para la síntesis de colágeno en los humanos. Causa anemia, debilidad, manchas en la piel y hemorragias. Era común en los marinos que subsistían con dietas en las que no figuraban fruta fresca ni hortaliza.

resultados, tales como: la condición clínica del paciente, el medio ambiente y la dieta básica.

El resultado del ensayo clínico realizado por Lind (1753) arrojó evidencia a favor del tratamiento con zumo de naranjas y de limón para combatir la enfermedad del escorbuto, pues los dos hombres dentro del grupo de doce que recibieron dicho tratamiento mostraron una mejor respuesta y una pronta recuperación. A pesar de que Lind no pudo responder a la pregunta de por qué funcionaba (las vitaminas aún no habían sido descubiertas), la Armada Real de Gran Bretaña incluyó las raciones de cítricos dentro de sus protocolos de navegación años más tarde (Milne, 2012). De esta manera, se gestaba dentro de la investigación una nueva forma de generación de conocimiento, en especial en el campo de la medicina. Sin embargo, muchas preguntas quedaron sin ser resueltas y había un profundo desconocimiento sobre cómo cancelar el sesgo de selección, a pesar de las precauciones tomadas por Lind.

Aunque el ensayo clínico sobre la enfermedad del escorbuto proporcionó una base fundamental para el posterior desarrollo de los ECA, aún faltaba un aspecto imprescindible en la investigación experimental: la aleatorización. Según Fienberg (1971) el principio es siempre el mismo: la suerte es una especie de fuerza poderosa, democrática e imparcial que hace que el dado o la moneda caiga de alguna manera determinada. Y es aquí donde viene la pregunta: ¿Por qué aleatorizar en la investigación científica?

Tiempo atrás, durante el invierno de 1883, el profesor Charles Sanders Peirce y un estudiante suyo, Joseph Jastrow, llevaron a cabo experimentos para poner a prueba la habilidad humana para detectar pequeñas diferencias en el peso, y detectaron por primera vez, la necesidad de la aleatorización en los estudios científicos, principalmente en la realización de experimentos.

Peirce tenía la pretensión de poner a prueba la hipótesis de Gustav Fechner, quién en 1860 en su trabajo titulado *Elemente der Psychophysik*, sostuvo la existencia de un límite inferior de la percepción, por debajo del cual un sujeto no es capaz de distinguir una ligera diferencia entre dos sensaciones de peso, Fechner (1860) lo nombró como “el umbral

diferencial". Para ello Peirce y Jastrow, elaboraron un experimento empleando la asignación al azar, Hall (2007) narra el experimento de la siguiente manera:

> Peirce creó un elaborado aparato mecánico; un sujeto experimenta por primera vez un peso de un kilogramo, y luego un segundo peso el cual puede ser un poco más pesado o un poco más ligero que el primero. El sujeto es requerido para ofrecer una opinión; es decir, no se le permitió decir que no podía diferenciar entre los dos pesos. Así que, después de haber dado una opinión, se le pidió evaluar su grado de confianza respecto a su opinión, en una escala de 0 a 3. Peirce usó la asignación al azar para determinar si el peso debía ser sumado o restado al primer kilogramo: él y Jastrow utilizan un juego ordinario de naipes donde hay cartas rojas y negras; el operador aumenta o disminuye el peso de acuerdo con el color de la siguiente carta. (p.304)

La aleatorización en el experimento de Peirce tenía, por lo tanto, dos propósitos: evitar que el sujeto que estaba haciendo parte del experimento adivinará el orden que se presentarían los pesos, y por otra parte, que el investigador no tuviera injerencia en el orden de las pruebas, pues si cree conocer cuál es el resultado de cada intervención, entonces es muy probable que la objetividad de su percepción se vea afectada y observe algo que no es, a esto se le denomina actualmente como "sesgo del observador" (Demirdjian, 2006). No obstante, Hall (2007) señala que a pesar de que Peirce describió su diseño experimental en detalle, incluyendo la asignación al azar, no salió nada de él; nadie más lo siguió. En la literatura posterior la aleatorización quedó totalmente abandonada. Fue hasta 1935, que Fisher establecería la aleatorización como una parte fundamental en el diseño experimental.

1.3 Sir Ronald A. Fisher y el diseño experimental

Fisher (1935) sistematizó la teoría matemática de la estadística y concibió muchas técnicas que siguen actualmente vigentes, entre ellas el análisis de la varianza y la distribución F (se llama así en su honor). A su vez, desarrolló el concepto de aleatorización y elaboró por primera vez un diseño específico de los experimentos. Su libro, *Statistical Methods for Research Workers* publicado en 1925, es el primer tratado moderno de inferencia científica, y *The Design of Experiments* de 1935, también obra suya, es el trabajo

responsable de los cambios más importantes en la metodología científica contemporánea, pero cabe resaltar, el experimento comparativo aleatorizado quizá sea su mayor contribución (Kuehl, 2007).

A principios del siglo XX, existía un gran interés por incrementar la productividad agrícola y por tanto la producción de alimentos. Con este fin se fundó en Inglaterra "La Estación Experimental de Rothamsted" que realizaría experimentos de campo para investigar el impacto de los fertilizantes sobre el rendimiento de los cultivos. En este tiempo la experimentación agrícola pasaba por serias dificultades. Para mencionar la más importante, estaba en cuestión la fiabilidad de los resultados experimentales. Según Moore (2010), la pregunta central al cual se enfrentaba Fisher y sus colaboradores era la siguiente: ¿Cómo se debe organizar la siembra de distintas variedades de un cultivo o la aplicación de diferentes fertilizantes para obtener comparaciones fiables? Es decir, ¿cuál es el impacto factores controlados (clase de semilla y fertilizante) sobre el rendimiento en la producción? El dilema estaba pues sobre la identificación del impacto.

Como la fertilidad de la tierra y otras variables cambian a medida que se desplaza a lo largo de un campo (factores incontrolados), tales como el clima, el drenaje, la luz, el viento entre otros, era difícil determinar la asignación de los diferentes tratamientos (nuevos fertilizantes) y la cuantificación del impacto sobre la producción. Por lo tanto, Fisher (1935) tuvo una idea para resolver este problema: escoger deliberadamente al azar las parcelas en las que se iba a plantar.

Para Fischer (1935) era indispensable formar grupos homogéneos, entendiendo por ello que tanto los factores que puedan controlarse como los que no, se distribuyeran de forma análoga en ambos grupos. Para él, el método que mejor garantiza esta homogeneidad es la aleatorización, esto es, asignar los tratamientos al azar (Saavedra, 2000). De esta manera, Fisher (1935) percibió que la asignación aleatoria de los tratamientos simularía el efecto de independencia en la distribución de la razón de varianzas, y en consecuencia el análisis de varianza y las pruebas de significancia usuales en la teoría normal serían aproximadamente válidas, siempre que la asignación de los tratamientos a las parcelas hubiese sido deliberadamente aleatoria (Box, 1980). Por lo tanto, a partir de dicha hipótesis se podría determinar si las diferencias entre los tratamientos correspondían a los efectos controlados.

Es por ello que en 1947, Fisher destacó el papel crucial de la asignación al azar:

> La teoría de la estimación presupone un proceso de muestreo aleatorio. Todas nuestras conclusiones están sobre la base de esa teoría; sin ella nuestras pruebas de significancia serían inútiles. [...] En la experimentación controlada se ha encontrado que no es difícil introducir la asignación al azar de una manera explícita y objetiva tal que las pruebas de significación sean correctas. En otros casos hay que actuar con la fe de que la Naturaleza ha realizado la asignación al azar por nosotros [...] Ahora reconocemos la asignación al azar como un postulado necesario para la validez de nuestras conclusiones. (Fisher, 1947, p.566)

1.4 El primer ensayo clínico aleatorizado

Como se presentó en la sección anterior, Fisher (1935) presentó la aleatorización como un ingrediente esencial en el diseño y análisis de experimentos, no obstante unos años más tarde Austin Bradford Hill, promulgaría la asignación aleatoria de los tratamientos en los ensayos clínicos como el único medio de evitar el sesgo sistemático entre las características de los pacientes asignados a diferentes tratamientos. Los dos enfoques fueron complementarios, Fisher apelando a la teoría estadística, Hill a las necesidades prácticas (Armitage, 2003).

A principios del siglo XX, con el desarrollo científico y técnico de la medicina, surgieron nuevos antibióticos en la lucha contra las enfermedades, y si bien estos descubrimientos eran prometedores, los médicos se enfrentaban al siguiente problema: ¿Cómo realizar una prueba objetiva y contundente que muestre la eficacia de estos nuevos medicamentos? De esta manera en 1943, Albert Schatz descubrió la estreptomicina, una nueva droga antimicrobiana que había demostrado desde los experimentos realizados en el laboratorio ser eficaz para contrarrestar las micobacterias causantes de la Tuberculosis. Sin embargo, se desconocía su efectividad en los seres humanos, por lo tanto, era necesario realizar un experimento en humanos, pero ¿qué tipo de experimento?

Austin Bradford Hill, un epidemiólogo y estadístico inglés, conocedor de los trabajos de Fisher, propuso realizar un ensayo clínico aleatorizado para demostrar la eficacia de la

estreptomicina en el tratamiento de la Tuberculosis. De acuerdo a Mukherjee (2010), Hill pensaba que los médicos eran los menos indicados para la realización de un experimento de este tipo, pues era inevitable que los médicos (aunque fuera de manera inconsciente) seleccionaran por adelantado ciertos tipos de pacientes y juzgaran los efectos de la estreptomicina sobre esta población sesgada por medio de criterios subjetivos, lo que significaba un verdadero riesgo en la fiabilidad de los resultados experimentales. Para Hill (1952) la asignación al azar:

> Asegura que ni nuestras idiosincrasias personales (nuestros gustos o disgustos aplican consciente o inconscientemente) ni nuestra falta de juicio equilibrado ha entrado en la construcción de los diferentes grupos de tratamiento, la asignación ha estado fuera de nuestro control y los grupos son, por tanto, no sesgados. (p.113)

De tal forma, Hill con el propósito de eliminar el sesgo de selección de los pacientes, asignó aleatoriamente unos pacientes al tratamiento con estreptomicina y otros al tratamiento con un placebo. De esta manera, al aleatorizar a los pacientes en los dos tratamientos desaparecería cualquier parcialidad de los médicos en su distribución y se impondría la neutralidad, por lo cual, la hipótesis en cuestión podría someterse a una prueba estricta.

Así pues, realizado el experimento se encontró que el grupo tratado con estreptomicina mostraba claramente una mejor respuesta que el grupo tratado con el placebo, posicionando de esta manera este antibiótico como un nuevo medicamento para el tratamiento de la tuberculosis, pero más importante, instauró la metodología de los Ensayos Controlados Aleatorios en la investigación médica, que dio lugar al posterior desarrollo de los métodos experimentales que ahora son usados a nivel internacional en la evaluación de nuevos medicamentos y terapias.

1.5 Historia de una conversión: de la Medicina Basada en la Evidencia a la Política Basada en la Evidencia.

El experimento de Bradford Hill sobre la estreptomicina traería consigo no solamente una invención metodológica, sino también un nuevo paradigma al campo de la medicina, sustituyendo así a la práctica médica tradicional. Por lo tanto, se da inicio a la introducción sistemática de estudios clínicos, como fuente habitual para la evaluación de fármacos, sueros, vacunas, agentes físicos, intervenciones quirúrgicas e incluso exámenes diagnósticos y otros procedimientos médicos (Chamorro et al, 2001). De hecho, si hubo algo característico de la evolución de la investigación clínica en la segunda mitad del siglo XX, fue precisamente el desarrollo de estudios clínicos destinados a evaluar los efectos de distintos agentes terapéuticos o preventivos.

Con esta revolución científico-técnica de la medicina, surge el movimiento que se conoce con el nombre de Medicina Basada en la Evidencia (MBE). Esta corriente nace a principios de la década de los noventa, y es sin duda el pensamiento dominante en la medicina de hoy (Tajer, 2010). Los pioneros de este enfoque fueron Gordon Guyatt y David Sackett, destacados dentro de un grupo de epidemiólogos clínicos y bioestadísticos de la Universidad de Mc Master en Ontario, Canadá. Aunque el MBE comenzó a gestarse en 1988, su declaración oficial data de 1992 con la publicación en el *Journal of the American Medical Association* (JAMA) del artículo “La MBE: un nuevo enfoque para la docencia de la práctica de la medicina".

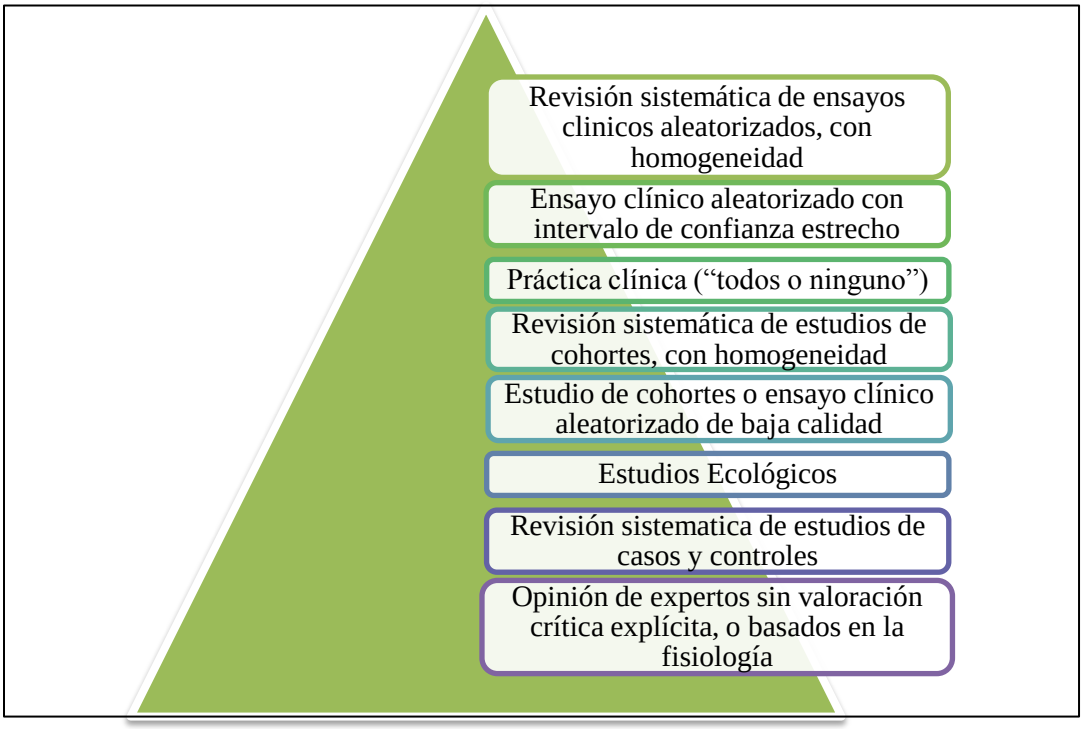
De acuerdo a Sackett (1996):

> La MBE es la utilización meticulosa, explícita y sensata de las mejores evidencias existentes para ayudar a tomar decisiones referentes al tratamiento de los pacientes. Su práctica integra la experiencia clínica personal con la mejor evidencia clínica proveniente de estudios clínicos de la mejor calidad metodológica. (p.71)

De esta definición surge naturalmente la siguiente pregunta: ¿Qué cuenta como “evidencia” para la MBE? La respuesta de los creadores de este concepto es clara, los mejores resultados son las que provienen de los ECA. Consideran que las decisiones de los

médicos deben ajustarse a las conclusiones de todos aquellos ensayos que están hechos con rigor científico, y propugnan una práctica médica basada en pruebas estadísticas, con datos confiables y evidencias experimentalmente confirmadas (Rodríguez, 2005). Quizás el mayor mérito de la MBE es defender una metodología precisa para la evaluación de la información científica y un ordenamiento jerárquico de las evidencias (Tajer, 2010).

Ilustración 1 Niveles de evidencia de la medicina



Elaboración propia con base en Primo (2003)

En la ilustración 1 se observa el nivel jerárquico de las evidencias provenientes de estudios médicos, lo cual se denomina como niveles de evidencia. En la cima de la pirámide se encuentran los estudios médicos considerados como los más rigurosos; de ellos se desprenden los resultados experimentales más fiables. Mientras que en la base de la pirámide se sitúan las prácticas médicas menos rigurosas.

A partir de la anterior ilustración, el método ECA es aquel que proporciona la evidencia más rigurosa y fiable dentro del conjunto de metodologías en el campo de la medicina. Otro elemento que se debe destacar del MBE es su procedimiento, que consta de cinco pasos fundamentales: (1) formular preguntas que puedan responderse, claras y precisas sobre un problema médico; (2) reunir evidencias, (3) valorarlas, (4) ponerlas en práctica y (5)

evaluar el proceso (Bernstein, 2004). Todo lo que se ha visto hasta ahora sobre el MBE condujo a un movimiento similar, pero esta vez, en el campo político.

En 1997 con la entrada del Gobierno de Tony Blair en el Reino Unido se impulsó una "modernización del Estado", que consistía en contribuir al cambio de la práctica de la "Política basada en la ideología", que confiaba en puntos de vista individuales y en supuestos ideológicos, para promover en su lugar, una práctica política que buscará activamente la mejor evidencia disponible en la toma de decisiones (Pinilla, 2006). Por lo tanto, la idea era sencilla: las políticas públicas deben estar informadas por evidencias rigurosas fruto de investigaciones. De ahí nace lo que se conoce como el movimiento de Política Basada en la Evidencia (PBE), naturalmente oriundo e inspirado por el MBE.

Según Davies (2004), el enfoque de la PBE es uno que "ayuda a las personas a tomar decisiones informadas sobre políticas, programas y proyectos, al colocar la mejor evidencia posible de las investigaciones en el centro del desarrollo e implementación de las políticas" (p.3). Para el PBE, el pragmatismo debe sustituir a la ideología, por ende apoya un enfoque más racional, riguroso y sistemático. Este movimiento puede tener un gran impacto sobre los países en desarrollo donde la mejor utilización de la evidencia en la política y la práctica económica puede ayudar a salvar vidas y reducir la pobreza (Sutcliffe & Court, 2005). El enfoque promovido por J-PAL en la economía del desarrollo hace parte del PBE, y los dos objetivos de J-PAL, los cuales son la producción y el uso de la evidencia en las decisiones políticas, claramente están en línea al objetivo general del PBE.

1.6 El uso de la aleatorización en la economía del desarrollo

Para Banerjee y Duflo (2011), el fracaso de las políticas para el desarrollo y las causas de que la ayuda no tenga el efecto que debería tener, radican a menudo en las llamadas "tres íes", es decir, ideología, ignorancia e inercia, por parte de expertos o dirigentes locales. Con el propósito de acabar con el escepticismo que se ha producido en la batalla contra la pobreza y por el desarrollo económico, el enfoque propuesto por este equipo de investigadores se basa en realizar Ensayos Controlados Aleatorios (ECA) en la economía, al mejor estilo de la investigación médica, pues según Duflo: "El principio general es acercarse a lo mejor del método de ensayo clínico" (citado en Mayneris, 2009, p. 2).

Esther Duflo, no fue la primera en tener la idea de realizar experimentos aleatorios en los países en desarrollo. A comienzos de la década de los noventa, Michael Kremer, profesor adscrito al MIT, ya había realizado una evaluación aleatoria en Kenia, la cual había mostrado que el suministro de nuevos libros de textos oficiales en las escuelas rurales no había mejorado las puntuaciones medias de las pruebas de los estudiantes. No obstante, a Duflo se le atribuye el poner en marcha lo que se ha llamado "una industria de la aleatorización" (Parker, 2010).

Este enfoque ha tenido en poco tiempo un sorprendente ascenso en la investigación de la pobreza y el desarrollo económico, que se atribuye a un "colapso de pensar en grande" como lo han descrito Cohen & Easterly (2009), el cual se resume en cuatro grandes aspectos:

> […] Los macroeconomistas fueron incapaces de ofrecer un mayor crecimiento; la credibilidad de la literatura de regresión del crecimiento disminuyó; tasas de crecimiento sumamente volátiles no podían ser explicadas […] La literatura sobre el crecimiento fue criticada por su incapacidad para hacer frente a la causalidad. (p.2-3).

Por otra parte, existía una evidencia muy pobre sobre el impacto de las políticas públicas en la lucha contra la pobreza, que generó un mayor interés en éste campo por responder a la pregunta: "¿qué funciona en la economía del desarrollo?". Lo anterior dio el espacio necesario para que la microeconomía, por medio de las evaluaciones de impacto, entrara a "llenar" el vacío que hasta ese momento existía en la economía del desarrollo. Esto obligó al separamiento de la tutela teórica, saliendo de los despachos y oficinas académicas, trasladándose hacia los países pobres con "batas de laboratorio" buscando comprender el fenómeno de la pobreza.

El método ECA es simple, se basa en hallar un efecto causal a través de la diferencia de los resultados en promedio, provenientes de los grupos de tratamiento y control. Aparentemente, éste método contiene menos supuestos para realizar una afirmación causal, permitiendo facilidad al transmitir los resultados obtenidos al público objetivo (hacedores de política, académicos y grupos de interés). De esta forma, el gran éxito y popularidad se

debe en gran parte a una reputación de objetividad que emite el método ECA, que es visto como un antídoto contra los prejuicios y problemas metodológicos que plagan a la econometría convencional (Salomone, 2012). Angrist & Pischke (2010) le dieron la bienvenida al nuevo movimiento de "Pensar en pequeño" considerándolo una "revolución de la credibilidad en la economía empírica" (p.3).

Ravallion (2009) describe el ascenso del enfoque experimental de la siguiente manera:

> Los randomistas, con el Laboratorio de Acción contra la pobreza del MIT se encuentran a la vanguardia, y están ganando influencia en la comunidad de desarrollo. Los investigadores están rechazando oportunidades para evaluar los programas públicos, cuando la asignación al azar no es factible. Los estudiantes de doctorado están buscando algo para aleatorizar. Agencias filantrópicas son a veces reacias a financiar las evaluaciones no experimentales. Incluso el Banco Mundial está respondiendo, después de haber sido criticado por los randomistas por no hacer suficientes experimentos sociales en sus operaciones; ha dedicado el mayor fondo (hasta la fecha) a las evaluaciones de impacto en preferencia explícita para diseños aleatorios. ¿Ha ido esto demasiado lejos? (p.1)

La pregunta que hace Ravallion, es una reacción explícita a la inminente transformación que ha tenido la economía del desarrollo en los últimos quince años. Por lo tanto, para describir a plenitud en qué consiste dicha transformación es necesario detallar un ECA para ilustrar su funcionamiento y sus elementos principales. Bajo éste objetivo, se describirá una evaluación aleatoria realizada en Colombia entre 2010 y 2011.

Attanasio, Meghir & Santiago (2011) llevaron a cabo un Ensayo Controlado Aleatorio para medir el impacto de dos intervenciones de desarrollo de la primera infancia en la cognición y la salud física, implementados a través de un programa de bienestar nacional existente, Familias en Acción. El contexto de la evaluación se enmarcó dentro de la evidencia de que los niños que viven en los hogares de más bajos ingresos en Colombia, a menudo acumulan retrasos sustanciales en el desarrollo, es decir, presentan diferencias significativas en el desarrollo (cognición y lenguaje) frente a niños de distinto nivel socioeconómico. Esta brecha se ensancha con la edad, y a la edad de 3 años las diferencias

entre los niños de clase media y los de las clases socioeconómicas más bajas son realmente considerables.

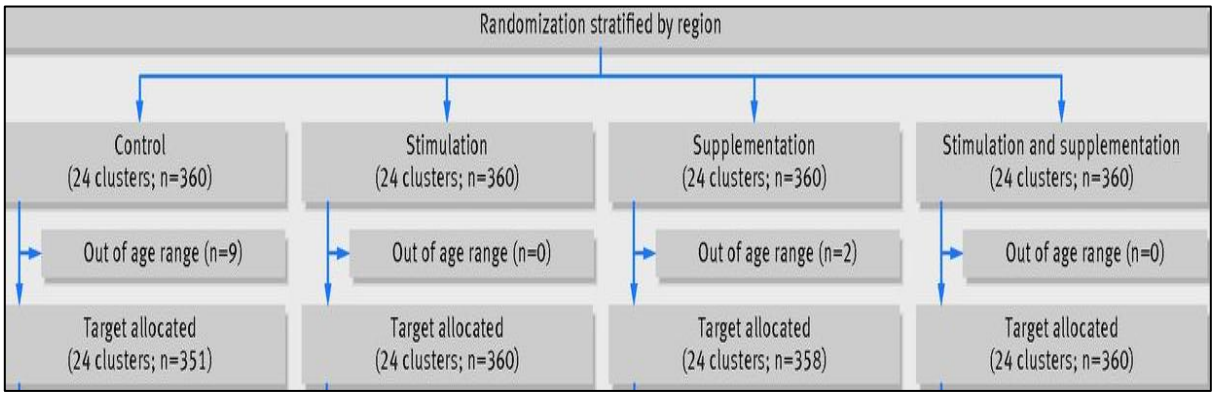
De esta manera, el objetivo de este estudio era responder a la siguiente pregunta: ¿Qué papel, si lo hay, puede desempeñar la política para remediar deficiencias tempranas entre los niños? En este contexto, se diseñó e implementó el experimento en 96 municipios de Colombia a través de ocho de sus 32 departamentos, con la participación de 1420 niños de 12 a 24 meses y sus cuidadores primarios.

Dos intervenciones serían evaluadas por el experimento. La primera intervención sería un programa de estimulación psicosocial, el cual consistía en visitas domiciliarias semanales por medio de "Madres Líderes" que realizaban actividades de juego utilizando juguetes de bajo costo, libros ilustrados y tableros. El objetivo era mejorar la calidad de las interacciones materno-infantiles y ayudar a las madres a participar en actividades de aprendizaje apropiadas para el desarrollo. La segunda intervención consistía en un programa de suplementación con micronutrientes en forma de polvo para tratar la anemia infantil. Los micronutrientes fueron suministrados por visitadores que proporcionaban a las madres instrucciones y formularios para la ingesta. El objetivo era contrarrestar la deficiencia de hierro y otros micronutrientes que puede afectar el desarrollo de los niños.

Por consiguiente, elaboraron cuatro grupos de comparación: (1) Estimulación, (2) La suplementación, (3) Estimulación + suplementación, (4) Grupo de control. Por su parte, los niños fueron asignados al azar en los cuatro grupos, utilizando la versión 9.0 del análisis de datos y el software estadístico STATA (StataCorp, College Station, TX). La intervención se puso en marcha en un período de cuatro meses, de febrero a mayo de 2010, y por etapas a lo largo de un periodo similar desde septiembre a diciembre de 2011.

La aleatorización se realizó de la siguiente manera:

Ilustración 2. Aleatorización estratificada por región



Fuente: (Attanasio et al., 2014)

Antes de seguir con el experimento de Attanasio et al. (2014), es importante destacar cómo la asignación aleatoria juega un papel clave en la evaluación de impacto. De manera que, como anteriormente se ha visto, cuando los grupos son seleccionados aleatoriamente se puede decir que son estadísticamente equivalentes, tanto a la población total como entre los mismos grupos de la muestra, es decir, los atributos de esos individuos elegidos son representativos del grupo entero del que fueron elegidos[8]. De ahí que Duflo (citado en Parker, 2010) defienda la asignación aleatoria:

> Este enfoque filtra el ruido de la estadística; se conecta causa y efecto. [..] La asignación al azar elimina las conjeturas, la magia, la destreza técnica, la intuición, de averiguar si algo marca la diferencia […] (p.2-3)

Retomando el análisis de la evaluación aleatoria de Attanasio et al. (2014), los resultados demostraron que el programa 1 de estimulación, mejoró significativamente el desarrollo cognitivo de los niños, mientras que el programa 2 de suplementos, no tuvo ningún efecto destacable. El programa de estimulación psicosocial aumentó las puntuaciones de cognición de los niños por 0.26 desviaciones estándar, y el lenguaje receptivo por 0.22 desviaciones estándar, en relación con los niños de la misma edad en las comunidades de comparación. Este estudio, llevado a cabo en Colombia, demostró que un programa de desarrollo infantil en estimulación psicosocial mejora el desarrollo cognitivo de los niños. Los resultados se muestran en la siguiente tabla:

[8] Análisis con base en la información de J-PAL, https://www.povertyactionlab.org

Tabla 1. Resultados de estimulación psicosocial

	Stimulation			Supplementation			Stimulation×supplementation		
	β (95% CI)	P value*	D†	β (95% CI)	P value*	D†	β (95% CI)	P value*	D†
Bayley-III scores‡:									
Cognition	1.139 (0.538 to 1.776)	0.002	0.260	0.196 (−0.294 to 0.676)	>0.50	0.045	−0.352 (−1.254 to 0.424)	>0.50	−0.080
Receptive language	0.776 (0.270 to 1.332)	0.032	0.218	0.128 (−0.458 to 0.706)	>0.50	0.036	−0.330 (−1.138 to 0.384)	>0.50	−0.093
Expressive language	0.455 (−0.286 to 1.250)	>0.50	0.084	0.403 (−0.580 to 1.269)	>0.50	0.074	−0.375 (−1.741 to 0.803)	>0.50	−0.060
Fine motor§	0.567 (−0.060 to 1.247)	0.34	0.122	0.455 (−0.184 to 1.102)	>0.50	0.098	−0.492 (−1.444 to 0.308)	>0.50	−0.106
Indexes:									
Aggregate index (Z score)¶	0.182 (0.050 to 0.314)	0.007	0.182	0.091 (−0.038 to 0.219)	0.165	0.091	−0.116 (−0.303 to 0.072)	0.223	−0.116
Factor index (Z score)**	0.217 (0.075 to 0.359)	0.003	0.217	0.075 (−0.061 to 0.210)	0.276	0.075	−0.091 (−0.286 to 0.104)	0.357	−0.091

Fuente: (Attanasio et al., 2014)

Ahora bien, la economía del desarrollo se ha volcado hacía a la búsqueda de lo que sí funciona para promover el desarrollo por medio de evaluaciones de impacto. En la base de datos de J-PAL, se almacenan ya 938 evaluaciones aleatorizadas que han sido lideradas por académicos afiliados en más de 64 países, todas ellas con el mismo tipo de metodología del experimento que se revisó en esta sección. Estas evaluaciones buscan evidencias a favor de políticas públicas eficaces para combatir la pobreza, por medio de ocho áreas del conocimiento: agricultura, crimen, salud, economía política y gobernabilidad, educación, finanzas, medio ambiente y energía, y mercado laboral.

Lo que quiere Duflo y sus colaboradores es que todos estos experimentos se conviertan en un "bien público internacional" para combatir la pobreza y promover el desarrollo. No obstante muchos críticos han salido a su paso, por ejemplo Easterly dijo: "adoptar los ECA ha llevado a los investigadores del desarrollo a rebajar sus aspiraciones" (citado en Banerjee y Duflo, 2011, p.292), y la respuesta de sus fundadores es:

> Puede que no tengamos mucho que decir sobre políticas macroeconómicas o sobre reformas institucionales, pero no se deje engañar por la aparente modestia de la empresa: los cambios pequeños puede tener efectos grandes. (Banerjee & Duflo, 2011, p.333)

1.7 Conclusiones Capítulo 1

Establecer el origen del método ECA en los trabajos de James Lind (1747) y Charles Sanders Peirce (1883), permitió mostrar en detalle los elementos básicos que cubren el método, tanto las características principales de lo que es un ensayo clínico como la adquisición de la aleatorización en la investigación y sus posteriores consecuencias sobre la economía del desarrollo. A partir de Peirce, se pudo ir elaborando la justificación y la defensa de la asignación al azar en el diseño experimental que terminó por ratificarse en el trabajo de Ronald Fisher (1935), quién a partir de demostraciones estadísticas, pudo estabilizar los atributos de la aleatorización. Paralelamente, se realizó una discusión frente al problema de identificación que va de la mano del problema de sesgo de selección, y se fueron mostrando cómo diferentes autores han tratado de enfrentarlos y resolverlos.

Con el experimento de Austin Bradford Hill (1943) y el realizado por Attanasio et al. (2014), se puede descubrir la gran analogía entre la actual economía experimental del desarrollo y la investigación médica, que ha tenido lugar en la última década. Esto se ha reflejado no sólo en la "importación" del método ECA, sino también en el aparente propósito de transferir todo el pragmatismo del MBE que lleva en sus hombros un determinado proceso de generación de conocimiento y una jerarquía de la evidencia.

El MBE y PBE, aunque separados por diferentes campos de investigación, guardan fuertes lazos, que se expresan en objetivos similares: "traduciendo la investigación en acción". Este acercamiento entre ambos movimientos ha llevado a la economía hacía un razonamiento inductivo y una clara apuesta a reformar el campo empírico de la economía. Las nuevas direcciones de la economía del desarrollo están abandonando los modelos teóricos por el enfoque empírico, quebrantando el predominio de la teoría. Un objetivo explícito de éste capítulo introductorio era estudiar cómo se introdujo la aleatorización en el diseño experimental y cómo logró obtener relevancia en el campo económico. De esta manera cobran vida algunas preguntas que serán estudiadas en el segundo capítulo: ¿Cuál es el tipo de causalidad que resalta de los ECA?, ¿Una para cada experimento?, ¿Cuál es el marco teórico que sustenta la realización de este tipo de experimentos en la economía?

CAPÍTULO 2: La inferencia causal desde el modelo Neyman-Rubin-Holland y su relación con la economía del desarrollo.

2.1 Evaluación de impacto; esencialmente un problema de inferencia causal.

Con la Declaración del Milenio en el año 2000, que fijó ocho objetivos de desarrollo humano para el año 2015, reafirmados y extendidos en los Objetivos de Desarrollo Sostenible, se ha difundido un profundo interés de los gobiernos, líderes mundiales, diferentes ONG´s y expertos del desarrollo, en aumentar la eficacia y la eficiencia de las políticas públicas con el propósito de dar cumplimiento a la nueva Agenda del Desarrollo. A causa de ello, en la última década, se han hecho importantes avances en cuanto a metodologías y herramientas para la evaluación de los resultados de las inversiones públicas en la reducción de la pobreza y mejora de los indicadores sociales de los países (Navarro et al., 2006).

Una de las metodologías más relevantes hoy en día en la investigación del desarrollo es la evaluación de impacto. La evaluación de impacto tiene como propósito determinar si un programa produjo los efectos deseados en las personas, hogares e instituciones a los cuales éste se aplica y la obtención de una estimación cuantitativa de estos beneficios (Aedo, 2005). Las evaluaciones de impacto forman parte de un programa más amplio de evaluación, en el que también se encuentran: evaluación de necesidades, evaluación teórica del programa, evaluación de procesos, análisis costo-beneficio, costo-efectividad, y costo-comparación, entre otras herramientas. Este trabajo se centra en las evaluaciones de impacto.

A diferencia de las otras evaluaciones, las de impacto, se preocupan por saber cuál es el efecto causal de un programa sobre un resultado de interés. Es decir, sólo interesa el efecto directo que tiene en los resultados (Gertler et al., 2011). De ahí que, Baker (2000) de una definición más técnica de evaluación de impacto: "es la obtención de evidencia contrafáctica acerca de qué habría ocurrido si el programa no se hubiera implantado, aislando otros efectos diferentes a la intervención para poder identificar con precisión" (p.7). Los diseños de evaluación de impacto se diferencian entre estudios experimentales y no experimentales; su principal diferencia está en la forma como se asignan los participantes a los grupos de tratamiento y control, elementos que se habrán de detallar más

adelante. En últimas, el objetivo fundamental de una evaluación de impacto es poder medir las relaciones causales.

2.2 El problema fundamental de la inferencia causal

Casi ninguna persona pasa por un día completo sin haber pronunciado frases como: "el pegamento no se adhirió a la madera a causa de la humedad", "el computador se dañó a causa de un virus", "la congestión vehicular me hará llegar tarde". Todas estas declaraciones tienen un tipo de relación entre dos entidades: causa y efecto. La causa conduce a un efecto, y el efecto sucede a consecuencia de la causa. No es sorprendente, por lo tanto, que los seres humanos dependan todo el tiempo de la causalidad para explicar lo que les ha ocurrido (Brady, 2002). Ahora bien, en el ámbito académico, el debate sobre la causalidad se ha protagonizado principalmente por filósofos y estadísticos.

La discusión sobre la causalidad se remonta a la Antigua Grecia con Aristóteles (384-322 a. C.), que realizó la primera y más sistemática definición de este concepto. Según Aristóteles, no bastaba una sola clase de causa para la producción de un efecto, sino que se necesitaban cuatro clases de causa, a saber: causa material, causa formal, causa eficiente y la causa final. Las dos primeras eran causas del ser y las dos últimas del devenir (Bunge, 1961). De las cuatro causas aristotélicas, sólo la causa eficiente se tuvo como merecedora de investigación científica, la cual se define como: "el primer comienzo del cambio y del reposo; por ejemplo, quien da un consejo es una causa, el padre es causa del hijo, y en general, el agente lo es de lo que hace, y el que produce cambio lo es de lo que cambia" (citado en Bunge, 1961, p.45). Esta doctrina sobre la causación persistió hasta el Renacimiento.

Con la llamada Revolución Científica y el comienzo de una ciencia predominantemente experimental, Galileo Galilei (1564-1642) propone una nueva definición de causa, refiriéndola como una "condición necesaria y suficiente para la aparición de algo […] aquella, y no otra, debe llamarse causa, a cuya presencia siempre sigue el efecto y a cuya eliminación el efecto desaparece" (Citado en Bunge, 1961, p.45). Este concepto de causalidad, manifiesta una transformación compleja pues pasa la causalidad metafísica de

Aristóteles a la causalidad lógico-fenoménica (Martínez, 1995), un cambio que sienta las bases para dos movimientos que nacen en la época de la Ilustración.

La evolución de la ciencia y el desarrollo del método científico que se gestó desde el siglo XVII condujo a grandes preocupaciones filosóficas que se expresaban en la siguiente pregunta: ¿Cuál es el origen del conocimiento y de la formación de las ideas? Consistía pues, en el interés por formular una "teoría del conocimiento". Bajo ese propósito, surgieron dos grandes líneas de pensamiento: el racionalismo y el empirismo.

El Racionalismo, de origen francés, fue liderado y defendido por autores como René Descartes, Malebranche, Spinoza, Leibniz y Wolff. Este movimiento defendió que la razón es la única o principal fuente de las ideas y de la verdad del conocimiento. Su investigación se fijaba por un razonamiento deductivo basado en principios evidentes o axiomas (Rojas, 2006). Mientras que, en oposición, el Empirismo defendió la experiencia como fuente principal de la verdad, promoviendo el uso del método empírico de la observación, la experiencia y la inducción. Sus principales exponentes fueron Francis Bacon, Thomas Hobbes, John Locke, George Berkeley y David Hume, una escuela de origen inglés. Ambos movimientos hicieron grandes aportes al problema de la causalidad, sin embargo, con los empiristas se fundó el escepticismo acerca de probar en la vida real las conexiones causales.

De la primera corriente, el pensamiento de René Descartes (1596-1650) fue el más influyente, su frase célebre "Pienso, luego, existo" reduce toda su concepción racionalista sobre la verdadera generación de conocimiento. Su formación académica estuvo fuertemente relacionada con las matemáticas y la geometría, por lo cual, sus preocupaciones fueron sobre todo de orden teórico. Según Xirau (2000) la vocación de Descartes era "la de un matemático que quería precisar las matemáticas y afinar un nuevo método para llegar a la verdad absoluta y necesaria" (p.214). Con este propósito Descartes desarrolló lo que se conoce como el "método cartesiano" que se fundaba principalmente en dos reglas, Xirau (2000) las describe de la siguiente manera:

> […] 1) Si queremos conocer algo debemos evitar la precipitación y la prevención; es decir, poner en duda, el conocimiento de oídas, el que proviene de lo que nos enseña

> la familia y el grupo social en que vivimos. 2) Una vez evitadas ambas, debemos proceder con claridad y distinción; debemos poner en duda la realidad para alcanzar la verdad. (p.217)

De este modo, Descartes renuncia a la experiencia como la forma correcta de generar conocimiento y defiende que es la deducción, con cada uno de sus pasos, una forma del descubrimiento inmediato y de creación, y en consecuencia, todos los casos particulares (teoremas, pruebas) son pasos sucesivos de una deducción que lleva de lo general a lo particular (Xirau, 2000). En cambio, el empirismo se fundamentó en la inducción, es decir aquel razonamiento que transfiere los datos de la observación de casos particulares de la experiencia hacia la instauración de leyes generales, suponiendo la existencia de una regularidad de los hechos naturales y sociales. Es por eso que Rojas (2006) dice lo siguiente: "el Yo del racionalismo es el Yo pensante, mientras que el Yo del empirismo es la experiencia" (p.206).

De la segunda corriente, fue el filósofo escocés David Hume el que tuvo un impacto revolucionario, no sólo en la filosofía sino en todas las ciencias. Según Schopenhauer (1967) fue Hume el primero que exigió de la inducción y la causalidad que mostrasen sus credenciales. Para Hume, los razonamientos inductivos implicaban siempre el principio de causalidad (Rojas, 2006). No obstante, el significado que Hume da a la causalidad es simplemente como una sucesión de impresiones, es decir, la causalidad no es una sucesión real que pueda encontrarse en los hechos naturales sino tan sólo una sucesión de hechos mentales (Xirau, 2000). De esta forma, según la doctrina empirista la causación es sólo una relación y no una conexión, que vincula experiencias y no hechos en general. Por esta razón Hume llegó a la siguiente conclusión:

> No es posible verificar empíricamente que la causa produce o engendra el efecto, sino tan sólo que el acontecimiento (experimentado) llamado causa está invariablemente asociado con el acontecimiento (experimentado) llamado efecto, o que el primero es invariablemente seguido por el segundo, argumento que, desde luego, se funda en la suposición de que sólo entidades empíricas pueden figurar en cualquier discurso relativo a la naturaleza o a la sociedad. (Citado en Bunge, 1961, p.18)

La tesis que sostenía Hume es que no hay una conexión entre el pasado y el futuro, razón por lo cual la doctrina empirista se apoyó de las siguientes conjeturas:

> a) Las impresiones de los sentidos son los únicos datos que deben tenerse en cuenta, tal como conviene a una teoría empirista; b) las impresiones de los sentidos son momentáneas, o sea, ajenas tanto al pasado como al futuro; c) como el pasado ya no es actual, no puede obrar sobre el presente, de modo que todo suceso es una nueva entidad por completo inconexa con las entidades que existieran en el pasado. (Bunge, 1961, p.57)

La crítica humeana de la causalidad y la inducción no debe entenderse, en primer lugar, como la negación a la existencia de la causalidad y, en segundo lugar, que la deducción es superior al pensamiento empírico. Hume lo que afirmó es que a partir de la experiencia, que se configura como la única fuente de conocimiento, no es posible determinar si hay o no la causación. Además, en *Una investigación sobre el entendimiento humano*, declaró la superioridad de la inducción sobre la deducción:

> […] Esta operación de la mente, por la que podemos inferir los efectos de las causas y viceversa, es esencial para la subsistencia de todas las criaturas humanas, es probable que pueda confiarse más en ella que en las falacias de la deducción de nuestra razón, que es lenta en sus operaciones; no aparece en los primeros años de la infancia; y como mucho es, en cualquier edad y periodo de la vida humana, extremadamente proclive al error. (Hume, 1777, p.40)

Otros filósofos, a partir de Hume, desarrollaron nuevas contribuciones al problema de la causalidad. Autores como Emmanuel Kant, John Herschel, John Stuart Mill, Bertrand Rusell y Patrick Suppes seguirían en la confrontación filosófica de este concepto. No obstante, en 1888 Francis Galton, concibió el concepto de correlación, que posteriormente fue desarrollado por Karl Pearson, y resultó en un giro radical hacia el tratamiento del problema de la causalidad. Esta vez, los nuevos interesados por el problema de la causalidad, centraban su atención en poder establecer un estudio acerca de la relación causal entre las variables que pudiera ser expresada en forma matemática, definida en

términos empíricos tanto desde los métodos observacionales como experimentales y, claramente, por fuera de los asuntos que hasta ese momento se había ocupado la filosofía[9].

Para Pearson, la correlación se presentaba como una poderosa arma frente a la causalidad. Defensor del positivismo, expuso que la misión de la ciencia no era explicar sino describir, es decir, descubriendo una formula descriptiva la cual fuera capaz de predecir la naturaleza (Porter, 2004). En 1911, sostuvo en su libro *La gramática de la ciencia*, que todo lo que podemos saber acerca de la causalidad está contenido en las tablas de contingencia: "Una vez que el lector se da cuenta de la naturaleza de dicha tabla, habrá captado la esencia del concepto de asociación entre la causa y el efecto" (citado en Norton et. al., 2014, p. 28). Por lo tanto, Pearson veía la causalidad como una consistencia probabilística de la asociación.

Esta corriente estadística ha prevalecido por mucho tiempo y ha influenciado disciplinas como la economía. Sin embargo, otros autores llegaron a la conclusión de que la "correlación no es causalidad", imponiendo un límite hacia la estadística y la economía en su contribución al análisis de la inferencia causal (García, 2010). Estos autores fueron Jersy Neyman, Donald Rubin y Paul Holland que de manera separada desarrollaron el "Modelo de Resultados Potenciales", sintetizado en 1986 por el trabajo de Holland titulado *Statistics and Causal Inference*. De este modelo se basan las experiencias de J-PAL y es ahí donde se concentrará el siguiente análisis.

2.3 Definición formal del problema de la causalidad[10]

Para ilustrar el problema de la causalidad, se usará un ejemplo cinematográfico: película *Mr. Nobody*[11]. Esta película narra la historia de Nemo Nobody, un niño de aproximadamente diez años el cual se enfrenta a una verdadera encrucijada en una estación de trenes. Sus padres han tomado la determinación de separarse y Nemo debe tomar la decisión de quedarse con alguno de los dos. Con el tren a punto de partir ¿Debe subir al

[9] Tal como dice Bunge (1961), muchos llegaron a declarar la causalidad como: "una ficción analógica" (Vaihinger), una "superstición" (Wittgenstein) o un "mito" (Toulmin) (p.347).

[10] La notación y definiciones que se seguirán a partir de esta sección están influenciadas por los trabajos de Holland (1986), Favereau (2014), Duflo et. al (2006), Lee (2005), Gertler et al. (2011), y García (2010).

[11] Godeau, P. (Productor), & Dormael , J. V. (Dirección). (2009). *Mr. Nobody* [Película]. Bélgica .

tren con su madre o quedarse con su padre? De esa elección dependerán muchas vidas posibles, todas son completamente distintas, incluso unas son más prometedoras que otras.

Formalizando lo expuesto, supongamos que se desea conocer el efecto que tiene para el nivel educativo de Nemo el subir al tren con su madre frente a no hacerlo. Para simplificar, la variable $\boldsymbol{X}$ (la variable "causa") puede tomar únicamente dos valores para cada unidad $\boldsymbol{i}$ ($\boldsymbol{i}$ en este caso es Nemo): $\boldsymbol{X_{0i}}$ **y** $\boldsymbol{X_{1i}}$, tomando el valor de 1 si Nemo se va con su madre y 0 si no lo hace. A su vez, la variable $\boldsymbol{X}$ tiene un efecto potencial sobre la variable $\boldsymbol{Y}$ (variable "efecto") que en este caso representa el nivel educativo de Nemo, los resultados potenciales serían por lo tanto, $\boldsymbol{Y_{0i}}$ y $\boldsymbol{Y_{1i}}$, respectivamente. De esta manera, si se quisiera conocer el efecto que tiene la madre sobre el nivel educativo de Nemo, se tendría que realizar simplemente la siguiente diferencia: $\boldsymbol{Y_{1i} - Y_{0i}}$ **,** suponiendo lo demás constante, donde esa diferencia simple se llamará efecto tratamiento.

Ahora bien, como señala Holland (1986), el papel del tiempo es muy importante debido al hecho de que cuando una unidad se expone a una causa esto debe ocurrir dentro de un período de tiempo específico, es decir, la unidad $\boldsymbol{i}$ (Nemo) debe ser expuesto tanto a $\boldsymbol{X_{0i}}$ como a $\boldsymbol{X_{1i}}$ al mismo tiempo y bajo exactamente las mismas condiciones, procurando entonces que cualquier otro factor que influencie a la variable $\boldsymbol{Y}$ impacte con la misma magnitud para ambos resultados potenciales (García, 2010).

Volviendo al ejemplo de Nemo, para conocer el efecto de la madre sobre el nivel educativo de su hijo se tendría que exponer a Nemo a las dos situaciones diferentes en el mismo momento, es decir, observando su futuro junto a su madre y a su padre simultáneamente. Por supuesto, que dentro de la ficción de la película esto se pudo contemplar, sin embargo, en la vida real es imposible dado que sólo uno de los dos resultados potenciales es observable, por lo cual la diferencia $\boldsymbol{Y_{1i} - Y_{0i}}$ es irrealizable.

En la siguiente tabla se ve el problema de la causalidad visto como datos perdidos:

Tabla 2. La causalidad vista como datos perdidos

i	Ti	Y_{0i}	Y_{1i}	$Y_{1i} - Y_{0i}$
1	0	3	5	2
2	1	2	5	3
3	1	5	4	-1
4	0	2	7	5

Elaboración propia con base en Lee (2005)

Si se toma el ejemplo de Nemo y se define *i*, como los individuos; *Ti*, como el tratamiento; $\boldsymbol{Y_{0i}}$ y $\boldsymbol{Y_{1i}}$, como los resultados potenciales de cada tratamiento; y $\boldsymbol{Y_{1i} - Y_{0i}}$, el efecto del tratamiento o efecto causal. Se puede observar que para el individuo 1, siendo *Ti=0*, el resultado observado de tomar la decisión de no irse con su madre es igual a 3, en una escala de calificación educativa de 1 a 10. No obstante, *Ti=1*, para este mismo individuo, sería un resultado imposible de observar y por ello se considera como un dato perdido (color rojo). Lo mismo ocurre para todos los individuos expuestos en la tabla. De aquí que se hace imposible calcular la diferencia $\boldsymbol{Y_{1i} - Y_{0i}}$; por lo cual nace la necesidad de encontrar un contrafactual válido para cada situación.

Para afinar la intuición del problema causal, ampliemos este ejemplo a dos personas: cada uno de estos individuos tendrá, por lo tanto, dos resultados posibles (resultados potenciales). Para la segunda persona, Pedro (denotado por $\boldsymbol{k}$) serán $\boldsymbol{Y_{1k}}$ y $\boldsymbol{Y_{0k}}$ tomando el valor de 1 si Pedro se va con su madre y 0 si no lo hace. Por consiguiente, si se quisiera observar el efecto que tiene la decisión de irse con su madre o no hacerlo para ambos individuos, el efecto se obtendría de la siguiente forma:

$$\frac{1}{2}[\boldsymbol{Y_{1i} - Y_{0i} + Y_{1k} - Y_{0k}}] \qquad \textbf{(1)}$$

Ecuación que representa la media de las diferencias de los resultados potenciales tanto de Pedro como de Nemo. De esta ecuación sólo dos resultados son observables:

1. $Y_{1i} - Y_{0k}$, en este caso, Nemo toma la decisión de irse con su madre y Pedro el no hacerlo.
2. $Y_{1k} - Y_{0i}$, Pedro toma la decisión de irse con su madre y Nemo toma la decisión de quedarse.

Como Nemo, al igual que Pedro, no pueden tomar dos decisiones simultáneamente, es necesario siempre considerar a otro individuo para realizar una comparación entre ellos. Si Pedro y Nemo tuvieran las mismas características, Pedro podría ser el clon perfecto de Nemo y así simular el *contrafactual* del mismo. Por contrafactual se entiende como la estimación de cuál habría sido el resultado *Y*, para un participante en ausencia de un programa. Para éste caso, el contrafactual es la estimación de los resultados no observables en cada uno de los observables.

Para la determinación del efecto causal, es necesario que Nemo y Pedro sean lo más idénticos posibles. En otras palabras, es necesario que $\boldsymbol{Y_{1i} - Y_{0k} = Y_{1i} - Y_{0i}}$, ó que $\boldsymbol{Y_{1k-}Y_{0i} = Y_{1k} - Y_{0k}}$. Sin esta condición, el efecto causal no sería plenamente identificado, pues los resultados podrían no ser producto del tratamiento en cuestión, sino de las diferentes características de los individuos.

En síntesis, el ***problema fundamental de la inferencia causal*** como lo ha descrito Holland (1986) es que nunca se puede encontrar el verdadero efecto causal a nivel individual.

2.3.1 Definición del Efecto Tratamiento Promedio (ATE)[12]

Tal como se mencionó en la anterior sección, el efecto tratamiento individual $\boldsymbol{\delta = Y_{1i}}$ - $\boldsymbol{Y_{0i}}$ no es posible identificarlo, pues uno de sus elementos no es observable. Ahora bien, si el estudio se centra en una población determinada y no en un individuo en particular, el camino hacia la identificación del efecto causal sí es factible, pues se trabajará a nivel de promedio y no a nivel individual. Así, el nuevo objetivo es analizar el efecto tratamiento promedio para la población (ATE, siglas en inglés), que en este caso constituye el impacto de un programa sobre una serie de resultados. De esta forma, se define el parámetro poblacional ATE:

$$\boldsymbol{\delta = ATE = E(y_1 - y_0) = E(y_1) - E(y_0)} \qquad \textbf{(2)}$$

Esta expresión es la manera como el Modelo Neyman-Rubin-Holland expresa la noción más básica de todas las declaraciones causales que se observa en la simple diferencia entre

[12] Siglas en inglés ATE: Average Treatment Effect.

el valor esperado del resultado potencial si se recibe el tratamiento y el resultado potencial si no se recibe; en el caso de Nemo, es la diferencia de si decide irse con su madre y el resultado potencial de no hacerlo. De tal forma, ante la imposibilidad de observar el efecto causal en una unidad específica, la solución estadística propone estimar el efecto causal promedio en una población de unidades ($\boldsymbol{i}$). Por consiguiente, si $\boldsymbol{\delta > 0}$ daría evidencia a favor del tratamiento en cuestión, siempre y cuando éste indique un beneficio. Por el contrario, si $\boldsymbol{\delta = 0}$ indicaría la falta de efecto medio del tratamiento a nivel poblacional. Por lo tanto, si se cumple la siguiente condición:

$$\boldsymbol{E(y_1) - E(y_0) \quad \neq \quad 0} \qquad \textbf{(2)}$$

Se concluye, que si el efecto medio es diferente de cero, existe un efecto causal del tratamiento en la variable respuesta de interés. Es importante señalar que el efecto causal promedio ($\boldsymbol{\delta}$) como su nombre lo indica, es un promedio y como tal goza de todas las ventajas y desventajas de los mismos. Dicho esto, se da paso a la definición del análogo muestral de $\boldsymbol{ATE}$:

$$\boldsymbol{\widehat{\delta} = \frac{1}{n_T}\sum_{i\epsilon T} y_i - \frac{1}{n_C}\sum_{i\epsilon C} y_i = \overline{y}|_{i\epsilon T} - \overline{y}|_{i\epsilon C}} \qquad \textbf{(3)}$$

Donde n_T es el número de unidades que se encuentran en el tratamiento y n_C es el número de unidades en el grupo de control.

De esta manera, si se quisiera estimar el impacto de un programa de desparasitación sobre la educación en niños, se tendría que encontrar un grupo de niños con características similares entre ellos, que al establecer un grupo de tratamiento y de control las características en promedio de cada grupo sean similares. Cumpliéndose esta condición, el efecto tratamiento promedio, se representa de la siguiente forma:

$$\boldsymbol{E\left[Y_i^T - Y_i^C\right]} \qquad \textbf{(4)}$$

Donde Y_i^T, es el resultado potencial si la unidad i recibe el programa de desparasitación (d) y Y_i^C cuando no lo recibe.

Si los dos grupos son iguales, a excepción de que uno de ellos participa en el programa y el otro no, cualquier diferencia de los resultados deberá provenir del programa. Es decir, las características que presenten los niños en el grupo donde se aplicó el programa de desparasitación, deben ser en promedio similares al del grupo de comparación, que en este caso sería el grupo que no recibe el programa. De tal manera:

$$\boldsymbol{\delta = E\left[Y_i^T | Niños\ Desparasitados\right] - E\left[Y_i^C | Niños\ no\ desparasitados\right] = E\left[Y_i^T | T\right] - E\left[Y_i^C | C\right]} \quad \textbf{(5)}$$

Si no existe un grupo válido de comparación, la evaluación de impacto del programa a realizarse será inválida, por lo tanto, no se estará estimando el impacto real del programa y será sesgado en términos estadísticos.

Sumando y restando la expresión $E\left[Y_i^C | T\right]$, que indica el resultado esperado para el grupo de tratamiento que no ha sido tratado (cantidad que no ha sido observada pero que lógicamente debe ser definida), obtenemos:

$$\boldsymbol{\delta = E\left[Y_i^T | T\right] - E\left[Y_i^C | T\right] - E\left[Y_i^C | C\right] + E\left[Y_i^C | T\right] = E\left[Y_i^T - Y_i^C | T\right] + E\left[Y_i^C | T\right] - E\left[Y_i^C | C\right]} \quad \textbf{(6)}$$

De tal manera, existe un grupo que se beneficia del programa de desparasitación $[Y_i|T]$, el cual tiene dos resultados potenciales:

1. $[Y_i^T|T]$, El resultado potencial del grupo *i* beneficiado por el programa de desparasitación cuando se recibe realmente.
2. $\left[Y_i^C|T\right]$, el resultado potencial del grupo *i* beneficiado por el programa de desparasitación si no se ha beneficiado realmente (situación hipotética, es decir, el contrafactual).

El grupo que no se beneficia del programa de desparasitación $[Y_i|C]$, a su vez cuenta con dos resultados potenciales:

1. $\left[Y_i^C|C\right]$, resultado potencial del grupo *i* no beneficiario del programa de desparasitación cuando en realidad no se beneficia.
2. $[Y_i^T|C]$, resultado potencial del grupo *i* que no se beneficia del programa de desparasitación si hubiesen recibido el programa (contrafactual).

Así, el primer término $E\left[Y_i^T - Y_i^C | T\right]$ es el efecto tratamiento sobre el grupo tratado, es decir, responde a la pregunta de ¿qué diferencia generó el programa de desparasitación? Término que es totalmente observable. Mientras que el segundo término $E\left[Y_i^C | T\right] - E\left[Y_i^C | C\right]$ muestra el sesgo de selección, el cual representa la diferencia entre el efecto promedio de los niños que hacen parte del grupo de tratamiento, que en realidad no recibieron el tratamiento y el efecto promedio en el grupo no tratado cuando no recibe el tratamiento. Entonces, se tiene que el primer término corresponde al sesgo de selección no observable y el segundo al observable.

Incluso, los niños que no fueron desparasitados, en promedio podrían haber tenido diferentes resultados sin haber sido tratados. Esto podría deberse a que los niños que se encuentran en el grupo de desparasitados tienen padres que se preocupan por el higiene de sus hijos y por ejemplo pueden incitar a que el niño se lave las manos antes de comer o a no andar descalzos, lo que haría que $E[Y_i^C | T]$ fuera más grande que $E[Y_i^C | C]$. Sin embargo, esto también se podría ver expresado en el caso contrario, si por ejemplo, una Organización No Gubernamental hubiera promovido una jornada nacional de cuidados de higiene por medios de comunicación oral y televisiva. Esto provocaría que $E[Y_i^C | T]$ fuera más pequeño que $E[Y_i^C | C]$ y el efecto de otras intervenciones se mostraran dentro de la medición. Para remover completamente el problema de sesgo de selección los niños en ambos grupos deben ser elegidos *aleatoriamente,* según el modelo Neyman-Rubin-Holland.

No obstante, si el sesgo de selección es igual a cero, el efecto promedio de tratamiento se calcula simplemente con la diferencia de los resultados esperados ($E\left[Y_i^T - Y_i^C | T\right]$). Siendo de esta manera, el análogo muestral del ATE ($\hat{\delta}$) un buen estimador poblacional. No obstante, para que lo anterior se cumpla, el modelo Neyman-Rubin- Holland requiere de los supuestos tratados en la siguiente sección.

2.3.1.1 Supuestos ATE

1. Condición de independencia

Supongamos que se realiza un sorteo en donde cada individuo tiene la misma probabilidad de hacer parte del programa de tratamiento. Así, el tratamiento será

independiente de los resultados potenciales de y_i, para *i=1,0*. A lo que formalmente diremos:

$$(\boldsymbol{y_1}, \boldsymbol{y_0}) \coprod \boldsymbol{d} \qquad \textbf{(7)}$$

Donde los resultados potenciales son estadísticamente independientes del tratamiento (*d*). De acuerdo a esta condición de independencia, entonces:

$$\boldsymbol{ATE = \delta = E(y_1 - y_0) = E(y_1) - E(y_0) = E(y_1|d = 1) - E(y_0|d = 0) = E(y|d = 1) - E(y|d = 0)} \qquad \textbf{(8)}$$

La última expresión ocurre debido a que y_1 sólo es observable cuando *d* = 1, lo mismo sucede para y_0 cuando *d* = 0. Este supuesto es el más relevante en el modelo de Neyman-Rubin-Holland. Intuitivamente, se refiere a que el tratamiento debe ser asignado entre los individuos del grupo de tratamiento y de control independientemente del resultado potencial que cada uno de ellos hubiera obtenido sin tratamiento. De manera que, si los individuos automáticamente se autoseleccionan para participar en un programa específico, teniendo la expectativa de obtener una ganancia esperada por participar en el grupo de tratamiento, Y_i no sería independiente de d y el efecto sería sesgado.

Por lo tanto, si tal supuesto se cumple, entonces ATE puede ser estimado consistentemente con la simple diferencia de los promedios de las observaciones de los grupos de control y de tratamiento, es decir, el estimador $\hat{\delta}$. Sin embargo, esta condición también se puede lograr con una independencia de medias de *d*, si $E(y_i|d) = E(y_i)$, lo que se transformaría en $E(y_1|d = 1) - E(y_0|d = 0)$ con *i* = 1, 0.

2. Estabilidad temporal y transitoriedad causal

La estabilidad temporal significa que hay una respuesta constante del tratamiento a través del tiempo, es decir, que la respuesta del grupo de control y del tratamiento no varía con respecto al tiempo, esto significa en términos matemáticos que:

$\boldsymbol{Y_{it}^T = Y_i^T}$, $\boldsymbol{Y_{it}^C = Y_i^C}$, para todo *i* y *t*, donde *t* = tiempo. **(9)**

Por otra parte, la transitoriedad causal afirma que el efecto de la causa en Y_i^C es transitoria y no cambia a los individuos suficientemente como para afectar más tarde a Y_i^T, es decir, el efecto que se pueda producir en el grupo de control no afecta significativamente al grupo de tratamiento.

3. Homogeneidad de la unidad

Se asume que $Y_i^T = Y_k^T$ y que $Y_i^C = Y_k^C$ para dos unidades *i* y *k*. Es decir, que el efecto que produce un tratamiento en un individuo *i*, debe ser exactamente igual al efecto que produce el tratamiento al individuo *k*. Supuesto que también ha sido aplicado en laboratorios científicos, cerciorándose a simple vista que sean idénticos en los aspectos más relevantes. Por supuesto que no se puede decir que el supuesto de homogeneidad es válido, pero sí hace que el supuesto sea verosímil.

4. Efecto constante

En el caso de que el efecto causal de $Y_i^T - Y_i^C$ sea grande, δ no sería un buen estimador del efecto causal de una unidad específica, debido a que habría una gran variabilidad del efecto causal entre los diferentes individuos. Si *i* es la unidad de interés, δ sería irrelevante, sin importar qué tan cuidadosamente se haya calculado (la cual es una de las desventajas de los promedios).

Así, el supuesto de efecto constante hace que el efecto del tratamiento sobre cada uno de los miembros del grupo de tratamiento sea el mismo, haciendo que δ se convierta en un estimador válido, tal como se ilustra en el siguiente ejemplo:

Retomando a Nemo (i) y Pedro (k), con sus respectivos efectos causales promedios $\delta_i = Y_i^T - Y_i^C$ y $\delta_k = Y_k^T - Y_k^C$, el supuesto de efecto constante implica que los resultados obtenidos a través de la aplicación del tratamiento son iguales tanto para Nemo como para Pedro, matemáticamente expresados como $Y_i^T = Y_k^T$ y $Y_i^C = Y_k^C$, los cuales determinan la siguiente igualdad:

$$\boldsymbol{\delta_i = \delta_k}, \text{ ó lo que es lo mismo, } \boldsymbol{Y_i^T - Y_i^C = Y_k^T - Y_k^C} \qquad \textbf{(10)}$$

Con este supuesto se concluye la especificación del modelo Neyman-Rubin-Holland, el cual da el marco teórico a los experimentos realizados en la Economía del Desarrollo por el equipo de investigadores de J-PAL. Finalizada la exposición del modelo contrafactual, es

pertinente destacar la importancia de la aleatorización, pues sin esta técnica este modelo no tendría ninguna validez.

2.4 El papel de la aleatorización sobre el modelo de Neyman-Rubin-Holland

Para poder abordar el papel que cumple la aleatorización sobre el modelo NRH es menester, en primer lugar, definir este concepto. En el sentido más simple, la aleatorización es la incorporación del azar como recurso en un proceso de selección. Esto sucede cuando, por ejemplo, utilizamos un juego de naipes, una moneda y/o una lotería para determinar quién participará en alguna actividad. Cuando alguna de estas herramientas (juego de naipes, moneda, lotería) se utilizan para tomar decisiones, se puede decir que el resultado se dejó en manos del azar, o que el resultado es aleatorio[13].

En la investigación científica, la asignación al azar se utiliza a menudo en la ejecución de encuestas o en la realización de experimentos, dado su gran poder en la construcción de grupos de comparación válidos. La razón estriba, en que cuando una cantidad suficiente de personas son seleccionadas aleatoriamente para participar en un experimento o encuesta, los atributos de esos individuos seleccionados son representativos del grupo entero del que fueron elegidos. Por tal motivo, lo que se descubra en ellos se puede inferir del grupo más grande, esto se conoce como *muestreo aleatorio*.

Una vez escogido el grupo seleccionado, es decir, después de haberse realizado el muestreo aleatorio, el siguiente paso es distribuir a estos individuos aleatoriamente en dos grupos (en un experimento simple), esto se conoce como *asignación aleatoria*. En este punto, se dice que estos dos grupos no son sólo estadísticamente equivalentes al grupo grande, sino que también son estadísticamente equivalentes el uno del otro.

Supóngase que se tiene un conjunto de unidades elegibles que está compuesto por 2000 personas, entre las cuales se seleccionan aleatoriamente la mitad para formar el grupo de tratamiento y la otra mitad el grupo de comparación. La selección aleatoria se realiza mediante un sorteo, donde a cada persona de las unidades elegibles se le asigna un número del 1 al 2000. Las que obtengan un número par corresponderán al grupo de tratamiento, dejando como grupo de control a los restantes. Así, se obtendrá un grupo de tratamiento

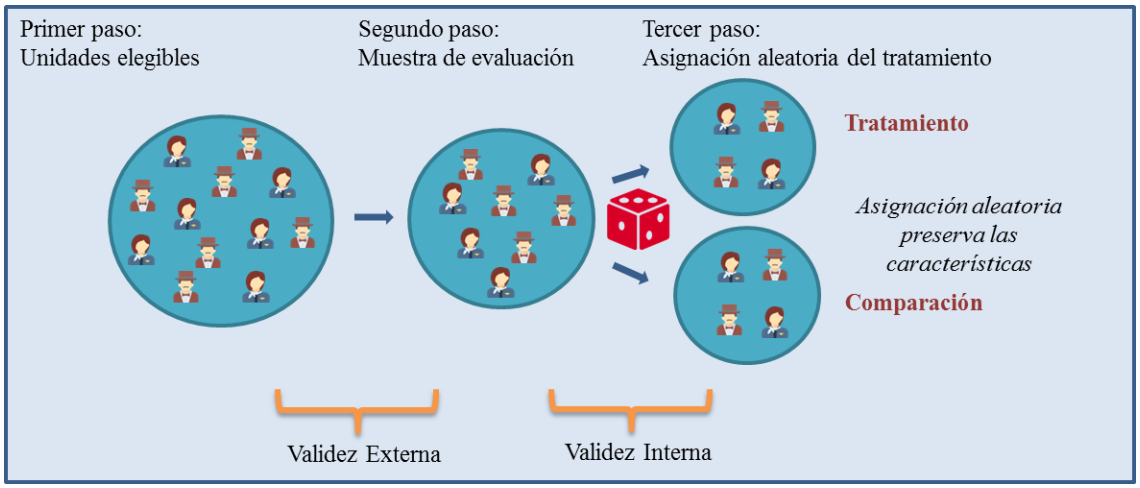
[13] Información con base en www.povertyactionlab.com

asignado aleatoriamente de 1000 personas y un grupo de comparación asignado del mismo modo.

De tal manera, que si el 25% de las 2000 personas elegibles son afrodescendientes, debido a la aleatoriedad en la selección, se esperaría que tanto en el grupo de tratamiento como en el de comparación se vea reflejado aproximadamente la misma proporción de afrodescendientes. En otro caso, si supiéramos que el 38% de las 2000 personas son antipáticas, se esperaría que esa misma proporción se distribuya aproximadamente igual entre ambos grupos. Por lo tanto, se puede llegar a la conclusión, de que la asignación aleatoria garantiza que, en general, los grupos de tratamiento y de comparación sean similares en todos los sentidos, es decir, tanto en sus características observadas como las no observadas, teniendo en cuenta que las no observadas son las características más difíciles de detectar.

Así, concretamente, lo anterior se recoge en 3 pasos como se muestra en la siguiente ilustración:

Ilustración 3. Asignación aleatoria



Fuente: elaboración propia con base en Gertler et. al (2011).

Los pasos descritos en la Ilustración 3 para la asignación aleatoria del tratamiento, aseguran tanto la validez interna como externa de la evaluación de impacto, siempre que la muestra de la evaluación sea suficientemente grande (Gertler et. al, 2011). Dos aspectos relevantes se derivan a partir de la anterior ilustración: (a) una evaluación es internamente válida si se emplea un grupo de comparación válido; (b) una evaluación es externamente

válida, si la muestra de evaluación representa al conjunto de unidades elegibles, de ahí que, los resultados puedan generalizarse a la población.

Si estas dos condiciones se cumplen, cualquier diferencia entre el grupo de tratamiento y de comparación sólo puede deberse a la existencia del programa en el grupo de tratamiento. Es ésta una de las razones por las que las evaluaciones aleatorias gozan de un estándar de oro dentro de las metodologías de evaluaciones de impacto. No obstante, una evaluación de impacto puede producir estimaciones internamente válidas, pero si la muestra no fue escogida aleatoriamente dentro de las unidades elegibles, estos resultados no son representativos para la población y por tanto no son generalizables. A su vez, puede ocurrir que se tenga una muestra aleatoria de la población y la asignación de los grupos de tratamiento y de comparación no hayan sido aleatorios, por lo cual, el efecto causal no sería identificado.

Cuando no hay aleatorización en ninguno de los pasos mencionados, provocará el llamado problema de sesgo de selección, cuyas fuentes radican en el desbalance existente de las características observables y no observables entre el grupo de tratamiento y de comparación. Si el sesgo se encuentra en las características observables, se denomina *sesgo en observables*, mientras que si difieren en las variables no observables se dirá que hay *sesgo en no observables*. Esto se conoce en la literatura inglesa como "overt bias" y "covert bias", respectivamente (García, 2010).

En presencia de la aleatorización, según el modelo NRH, el sesgo de selección desaparece, dado que la asignación aleatoria producirá grupos con un promedio de todas sus características estadísticamente equivalentes, como si se generara a nivel muestral un "clon perfecto" para cada grupo, haciendo que $\hat{\delta}$ se convierta en un buen estimador de ATE.

Por otra parte, sin la asignación al azar no se daría cumplimiento a los supuestos de ATE, especialmente al supuesto de independencia, pues al haber una correlación entre el resultado potencial y el programa a realizar, se generaría un sesgo de selección que afectaría el estimador muestral.

2.5 El modelo de Neyman-Rubin-Holland dentro de la economía del desarrollo

Para proporcionar una evaluación rigurosa de los programas del desarrollo, Banerjee & Duflo con todo el equipo de investigadores de J-PAL, han adoptado el modelo causal NRH en el desarrollo económico. Esta vez el interés está en determinar los efectos de una política para la lucha contra la pobreza, por ejemplo, la evaluación de una política que reduce la incidencia de la malaria en la población vulnerable de la India ó la evaluación de una política educativa que mejora la asistencia de los niños a las aulas escolares en México (PROGRESA).

En primer lugar, J-PAL "es una red de 136 profesores afiliados provenientes de más de 40 universidades alrededor del mundo. Su misión es reducir la pobreza garantizando que las políticas públicas estén informadas por evidencia científica." Su objetivo es: "apoyar el uso de evaluaciones aleatorias, desarrollar capacidades sobre métodos rigurosos de evaluación científica, y promover cambios en las políticas con base a los resultados de las evaluaciones aleatorias". [14]

Esta institución se estableció en el 2003 con la ayuda de tres profesores del Instituto de Tecnología de Massachusetts (MIT): Abhijit Barnerjee, Esther Duflo y Sendhil Mullainathan, quienes posteriormente se aliarían con la filósofa, politóloga y economista Rachel Glennester para pertenecer a la dirección de J-PAL. Después de tres importantes donaciones de Mohammed Abdul Latif Jameel, J-PAL empezó a tener un crecimiento significativo en los siguientes años, instruyendo a investigadores dentro de "programas de formación" destinados a evaluaciones aleatorias, los cuales fueron los encargados de realizar evaluaciones en diferentes países. Así, fue en 2007 donde J-PAL experimentó un punto de inflexión particular, con nuevas antenas en el sur de Asia, África, Nigeria, Europa, Chile y Sudáfrica; haciendo presencia en los cinco continentes.

El laboratorio mantiene una estrecha relación con muchos centros de investigación, lo cual permite que sus investigadores hagan uso de su método en muchas partes del mundo, esto en colaboración con los llamados "socios de campo" e instituciones que a través de los

[14] Información tomada de la página oficial: www.povertyactionlab.org

años, han venido contribuyendo con aportes monetarios significativos, permitiendo que la asignación al azar tome fuerza en la Economía del Desarrollo. Es así como el éxito de J-PAL como organización en la Economía del Desarrollo, se fundamenta en sus grandes aliados y en la construcción de sus redes sólidas de centros de investigación con antenas en los cinco continentes.

Cada antena se encuentra estructurada de la misma manera, diferenciados en tres categorías: investigadores afiliados, equipo administrativo y equipo de investigación. Esta estructura permite trabajar con un número importante de investigadores a través del mundo, permitiendo la posibilidad de tener investigadores que conocen especificidades de su propio país o inclusive de su continente, lo que a su vez permite que las redes de J-PAL sean cada vez más extensas.

¿Cómo opera? En la aplicabilidad de las evaluaciones, J-PAL cuenta con la colaboración de tres socios: un socio de campo que esté interesado en realizar el proyecto y que a su vez quiera evaluarlo. Por lo regular, estos socios son gobiernos, organizaciones no gubernamentales o empresas privadas. También, se necesitan un investigador o un equipo de investigadores que estén suficientemente interesados en llevar a cabo el proyecto; y por último, se necesita un ente financiador que en ocasiones suelen ser los socios de campo los que cumplen esta función (Duflo en Mayneris de 2009, p.3).

2.5.1 La investigación de J-PAL en ocho áreas del conocimiento

El número de evaluaciones y de profesores afiliados ha aumentado considerablemente, pasando de tres investigadores a más de cien. J-PAL ha puesto en marcha 935 experimentos en 67 países distribuidos en 8 áreas del conocimiento: Agricultura, Crimen, Educación, Medio ambiente y Energía, Finanzas, Economía política y Gobernabilidad, Salud y Trabajo. Tal y como se muestra en la siguiente tabla:

Tabla 3. Número de evaluaciones aleatorias realizadas por J-PAL

Áreas	Completas	En proceso	Total
Agricultura	32	37	**69**
Crimen	13	2	**15**
Economía política y Gobernabilidad	122	46	**168**
Educación	140	42	**182**
Finanzas	144	70	**214**
Medio Ambiente y Energía	18	16	**34**
Trabajo	67	34	**101**
Salud	98	54	**152**
Total	**634**	**301**	**935**

Fuente: elaboración propia con base en www.povertyactionlab.org

La generación de conocimiento para J-PAL se basa en evaluar los programas del desarrollo para cada una de las 8 áreas del conocimiento. En la Agricultura centran sus esfuerzos en la mejora de los sistemas agrícolas en el mundo en desarrollo, estudiando maneras de ayudar a los agricultores a adoptar prácticas y tecnologías rentables o ambientalmente sostenibles, con programas que les permita vincularse mejor al mercado. Por otro lado, sabiendo que el crimen y la violencia son los principales factores que impiden el desarrollo económico, los profesores afiliados determinan los mejores programas que reduzcan la delincuencia y la violencia a los menores costos, buscando maneras de reformar la justicia penal.

El programa de Educación de J-PAL promueve investigaciones que busquen darle a los encargados de políticas públicas los conocimientos que necesitan para diseñar e implementar políticas y programas de educación más efectivas, los cuales centran su atención en siete sectores: pedagogía, tecnología, capacitación para los profesores, incentivos para los profesores y administración de las escuelas, incentivos para los estudiantes, desarrollo infantil temprano y educación post-primaria. En cuanto al medio ambiente, los afiliados de J-PAL investigan cómo mejorar la regulación de contaminantes, los efectos de estufas limpias en la salud y el medio ambiente, y los pagos condicionales por la conservación de bosques; además, por medio de las evaluaciones aleatorias

identifican soluciones con alta ración de costo efectividad para los desafíos ambientales y energéticos que enfrentan los países en desarrollo.

El programa de finanzas de J-PAL utiliza las evaluaciones aleatorias para medir el impacto de productos y servicios financieros como un mecanismo para reducir la pobreza y catalizar el desarrollo económico, ayudando a mejorar la cantidad y la calidad del acceso financiero para todos los niveles de ingreso. Ahora bien, dentro del programa de Economía política y Gobernabilidad, se busca promover evaluaciones de impacto de programas que busquen mejorar la participación política y el proceso político que reduzca fugas de recursos públicos, desarrollando programas que minimicen la corrupción y mejoren la eficacia de los servicios públicos.

Por otra parte, el programa de salud de J-PAL busca identificar las maneras por las cuales las evaluaciones aleatorias pueden contribuir con mejoras en los servicios de salud y así incidir en un aspecto de la pobreza. Este programa se interesa, por ejemplo, en responder este tipo de preguntas: ¿cuáles son las soluciones efectivas a los problemas asociados a la entrega de servicio de salud en áreas urbanas? Así como esta y otras preguntas importantes hacen parte de esta línea de investigación.

Por último, el programa de mercado laboral o trabajo de J-PAL usa evaluaciones aleatorias para identificar los efectos causales de políticas laborales y de programas que se enfocan en las personas en busca de empleo, centros de empleo y empleadores, bajo estas líneas de investigación: capacitaciones, orientación e inserción laboral, emprendimiento, discriminación laboral, entre otros, respondiendo a preguntas como ¿qué intervenciones son las más efectivas para ayudar a los desempleados a encontrar empleo? ¿Cuáles son los retornos a las capacitaciones profesionales? ¿Son las pasantías eficientes en ayudar a los jóvenes a realizar la transición de la universidad al empleo? Es así como J-PAL dedica a partir de estas 8 áreas del conocimiento sus esfuerzos en conocer y comprender la pobreza para contribuir en la reducción de la desigualdad global.

2.5.2 PROGRESA, "la curiosa historia de las transferencias monetarias condicionadas"

En México, en el año de 1997, se realizó uno de los experimentos aleatorios más conocidos que se ha llevado a cabo en el mundo, por parte de los investigadores de J-PAL. Progresa no sólo fue notable por su diseño original, sino también porque cuando el programa comenzó, el gobierno inició una rigurosa evaluación de sus efectos. Es tal vez el más visible de una nueva generación de intervenciones, cuyo objetivo principal era mejorar el proceso de acumulación de capital humano en las comunidades más pobres (Attanasio, Meghir & Santiago, 2011).

El gobierno mexicano implementó un programa nacional de Transferencias Monetarias Condicionadas (TMC), llamado Progresa, con el objetivo de mejorar la nutrición, la salud, y la educación de los niños dirigido a las familias de bajos ingresos en localidades rurales.[15] Este programa ofrecía dinero a las familias pobres, pero solamente si sus hijos iban a la escuela con regularidad y si la familia buscaba servicios de salud preventivos. Si los niños estudiaban en enseñanza secundaria recibían más dinero que si lo hacían en primaria, al igual que si quien iba a la escuela era una niña en lugar de un niño. Para que el programa fuese aceptado políticamente, los pagos se presentaron como compensaciones a la familia por los salarios que dejaban de ingresar cuando sus hijos iban al colegio en el lugar de ir a trabajar (Banerjee & Duflo, 2011).

Para realizar la evaluación de impacto en el programa Progresa, se utilizaron los datos del censo del 1997 para identificar las comunidades elegibles sobre la base de la situación socioeconómica. De 50.000 comunidades elegibles, 506 fueron elegidas al azar para convertirse en la muestra de evaluación, para un total de más o menos 25.000 hogares. De las 506 comunidades, 320 fueron asignadas al grupo de tratamiento, recibiendo los beneficios del programa y 186 restantes fueron elegidos como grupo de comparación. Los hogares en el grupo de tratamiento recibieron la transferencia de dinero todos los meses, condicionada a completar los componentes de salud y educación que se requerían. Además de la encuesta de referencia (censo 1997), se llevaron a cabo encuestas de seguimiento para todas las familias elegibles cada 6 meses entre noviembre de 1998 y abril de 2000.

[15] Información con base a www.povertyactionlab.org

Los resultados sugieren que Progresa tuvo un efecto positivo en la matrícula de los niños, sobre todo, después de la escuela primaria. El aumento de la matrícula de los niños de 14 años de edad aumentó en 14 puntos porcentuales. Además, el número de matriculados, especialmente en el nivel de secundaria, las niñas pasaron del 67% al 75% y los niños del 73% al 77%. Progresa fue eficaz en el aumento de la participación escolar, dando lugar a una reducción de la oferta de trabajo infantil disponible en los pueblos de México.

A partir de los impactos positivos que reflejó esta evaluación aleatoria, Progresa se extendió a nivel nacional en México, beneficiando a 5,8 millones de familias (un cuarto de la población total del país) y hoy en día sigue en vigencia bajo el nombre de Oportunidades (2007-2016). El éxito de este programa conllevó a que más de 50 países en el mundo replicaran este modelo (Banco Mundial, 2014). Es esta una de las primeras pruebas de la capacidad de persuasión de un experimento aleatorio exitoso (Banerjee & Duflo, 2011).

2.5.3 Publicaciones académicas y lecciones de políticas

Progresa hace parte de las 634 evaluaciones aleatorias que se han realizado en la economía del desarrollo por parte del equipo de investigadores de J-PAL. Una vez completada una evaluación aleatoria, sus resultados son divididos en *publicaciones académicas* y *lecciones de política*.

Las publicaciones académicas van dirigidas a los debates teóricos y empíricos que hacen hincapié en la economía del desarrollo, los cuales son publicados en revistas económicas especializadas, principalmente en el American Economic Journal: Applied Economics. Mientras que las lecciones de política van dirigidas a los hacedores de política, gobiernos, organismos multilaterales, fundaciones, ONG's, y en general, a personas que deseen conocer las políticas o programas del desarrollo económico que funcionan en la batalla contra la pobreza, y así, incidir en las decisiones de política pública y programas sociales.

Uno de los propósitos fundamentales de J-PAL es construir lo que se ha denominado "un banco de conocimientos" que esté al alcance de la comunidad entera y que se alimente de todas aquellas evaluaciones aleatorias realizadas a nivel mundial, proporcionando

evidencia rigurosa de la efectividad de los programas de desarrollo, con el objetivo máximo de comprender y minimizar la pobreza en los cinco continentes.

2.6 Conclusiones Capítulo 2

Este capítulo se centró principalmente en el problema de la causalidad. En primer lugar, mostró la relación que existe entre la metodología de evaluaciones de impacto dentro de la economía del desarrollo y el problema de la inferencia causal. Este análisis conllevó a estudiar el problema fundamental de la causalidad a través de la visión filosófica y estadística de autores como Aristóteles, David Hume, Pearson y Galton.

Lo anterior, permitió adentrarnos a la visión propuesta por Neyman-Rubin-Holland para hallar efectos causales en los diseños experimentales a partir del modelo contrafactual. Se mostraron las propiedades básicas y sus supuestos con el propósito de entender la lógica del conocimiento que subyace de una evaluación aleatoria. En este aspecto, cobró vida la importancia de la asignación al azar y, cómo teóricamente logra contrarrestar el problema de sesgo de selección, además del impacto que genera en la validez interna y externa en este tipo de estudios. Es por ello que, bajo una perfecta asignación al azar, se concluye que las evaluaciones aleatorias gozan de una alta validez interna y externa, permitiendo obtener evidencia rigurosa sobre los efectos causales de los programas de desarrollo, considerándose así como el "estándar de oro" en la metodología de las evaluaciones de impacto.

Posteriormente se muestra, cómo J-PAL utiliza este método en 8 áreas de conocimiento; todas ellas con el propósito de encontrar programas que reduzcan la pobreza y mejoren la calidad de vida de las personas. Con el programa Progresa, se explica cómo el modelo NRH es aplicado en la economía del desarrollo a través de un estudio experimental elaborado por el equipo de investigadores de J-PAL. En este mismo sentido, se muestra la pretensión de esta institución de incidir en las decisiones políticas y no sólo en la academia.

Ahora bien, ya expuesto el método ECA a cabalidad, es menester evaluar esta propuesta tanto metodológica como teóricamente. Pues, a partir de lo visto en este capítulo, pueden surgir preguntas como: ¿Realmente los ECA generan información sobre la causalidad a partir del ATE? Si es así, ¿qué utilidad tiene sobre la economía del desarrollo? ¿Es el

modelo de resultados potenciales la mejor forma de hallar causalidades? ¿La asignación al azar realmente resuelve el sesgo de selección como se ha visto teóricamente? ¿Es este método originado de la medicina apropiado para la política social? Todas estas preguntas abren paso a una discusión que se desarrollará en el siguiente capítulo.

CAPÍTULO 3: La evaluación aleatoria en la práctica

3.1 Indagando el método ECA

Los Ensayos Controlados Aleatorios (ECA) han adquirido una cierta hegemonía en las evaluaciones de impacto sobre la base del gran poder de la asignación aleatoria para hallar efectos causales. Esta metodología, hoy en día es considerada como el "estándar de oro" de las evaluaciones de impacto y es, según el enfoque J-PAL, la única que produce evidencia rigurosa de la identificación causal. La principal virtud de la asignación al azar, como se discutió en el capítulo anterior, es la reducción del sesgo de selección, generando grupos de comparación válidos que otorgan una fuerte validez interna y externa de la investigación experimental.

De esta manera, los ECA se han convertido en el método que proporciona la mejor evidencia de efectividad y eficacia de los programas del desarrollo en la lucha contra la pobreza desde el enfoque J-PAL. Tal ha sido la fuerza del método ECA que ha desplazado a los estudios no experimentales a un segundo plano, conformando lo que Easterly (2009) ha denominado una *jerarquía de la evidencia*, clasificando estos métodos como "evidencia dura" y los demás como "evidencia blanda" (Cohen & Easterly, 2009).

Teniendo en cuenta el cambio que ha provocado los ECA en la investigación económica, basados en una serie de premisas y afirmaciones señaladas anteriormente, es pertinente entrar a evaluar esta propuesta metodológica, realizando un análisis y una discusión crítica de esta metodología experimental con el fin de investigar su poder, alcance y limitaciones. Esto conduce a evaluar propiamente el método, cuestionando tanto el papel de la aleatorización, como la utilidad de los resultados subyacentes y examinar la contribución del enfoque J-PAL a la comprensión de los fenómenos de la pobreza.

Con este objetivo, las críticas presentadas por reconocidos economistas del desarrollo como Deaton (2009), Ravallion (2009) y Rodrick (2008) y por filósofos reconocidos como Cartwright (2007), serán discutidas a lo largo del capítulo.

3.1.1 ¿Los métodos ECA como estándar de oro?

Retomando los dos objetivos principales del enfoque J-PAL - (1) Generar evidencia rigurosa de la efectividad de los programas y (2) Guiar las decisiones de política a partir de

los resultados experimentales- se encuentra que existe una estrecha relación entre los conceptos de validez interna y externa, respectivamente. Para cumplir con el primer objetivo, un experimento aleatorio debe comprobar que el impacto medido causado por el programa evaluado es, en últimas, una cuestión de validez interna. De este mismo modo, el cumplimiento del segundo objetivo depende de qué tanto los resultados son generalizables y replicables en otros lugares, refiriéndose al concepto de validez externa.

De acuerdo, a la filósofa Nancy Cartwright: *"no existe ningún estándar de oro: ningún método es universalmente mejor"* (Cartwright, 2007, p. 4). Para ella, un método que proporcione la información que se necesita, sea fiable y esté acorde con lo que se pueda hacer y conocer dependiendo de la ocasión, se puede denominar como un estándar de oro. Por lo tanto, el método que proporciona la mejor información será diferente de un caso a otro, es decir que, habrá lugar en el que ECA sea un buen método y en otros no.

En el caso del método ECA, una de las principales inquietudes es la utilidad de los resultados subyacentes, pues un programa puede ser efectivo para disminuir la deserción estudiantil en México, pero eso no significa que el tratamiento sea eficaz en otras regiones del mundo. No obstante, los investigadores de J-PAL bajo su lema "*Traduciendo la investigación en acción*", pretenden generar evidencia rigurosa por medio de experimentos aleatorios y promover aquellas políticas que hayan sido efectivas bajo este método de evaluación. Lo anterior requiere, entonces, que la efectividad del programa adquirida en México, se vea reflejada al aplicarse, por ejemplo en Colombia. De esta manera, cabe preguntarse: ¿cómo es posible asegurar que el efecto que se produjo en México se presentará de igual manera en Colombia?

Esta discusión ha sido ampliamente detallada por autores como Francesco Guala (2005), Easterly (2009), Deaton (2009), entre otros. El punto crítico del método ECA es que goza de una fuerte validez interna frente a otro métodos, sin embargo, su validez externa se ve seriamente comprometida, lo que ha denominado Favereau (2014) como una *contradicción epistemológica* que consiste en que los dos objetivos de J-PAL son opuestos, pues al esforzarse en encontrar la identificación causal en una población determinada con características especiales, adquiere un grado de complejidad mayor al momento de extrapolar los resultados encontrados de una evaluación a otras situaciones o lugares.

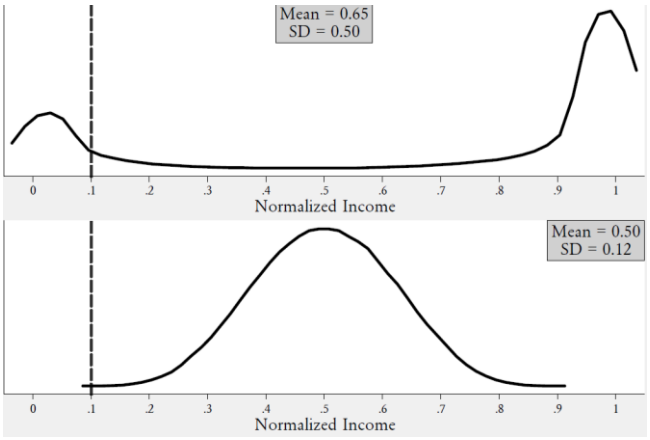
Sin embargo, para poder hablar de lo anterior, se necesita primero, verificar si el efecto causal promedio (ATE) producido por un experimento aleatorio, en realidad cuenta como un efecto causal. De esta forma, antes de preguntarse por la validez externa, se debe cuestionar los hallazgos producidos por este método. Por lo tanto, es menester estudiar cómo se ve comprometida la validez interna y externa de los experimentos aleatorios.

3.2 Validez Interna: ¿la joya de oro golfi?

3.2.1 Efecto heterogéneo del tratamiento

A partir de un experimento aleatorio, en el cual un programa de desarrollo es asignado aleatoriamente entre el grupo de tratamiento y de control, se obtiene el resultado de interés, ATE. Como se ha mencionado, este resultado corresponde a un promedio, el cual representa el efecto causal de una intervención; no obstante, este promedio no permite observar la distribución de los efectos del programa entre los individuos que conforman dichos grupos. En este sentido, en la ilustración 4 se observa el efecto heterogéneo del tratamiento en dos programas diseñados para un mismo propósito (evaluar un programa del desarrollo). En el panel superior (programa 1), se muestra un efecto causal promedio mayor que el panel inferior (programa 2), sin embargo, el programa 1 presenta una desviación estándar más grande, es decir, los efectos observados en las personas en el panel superior fueron más dispersos que en el panel inferior. De esta manera, si se quisiera elegir cuál de los programas es el mejor con base en su promedio, se optaría por elegir el programa 1.

Ilustración 4. Distribución de los efectos promedio



Fuente: Harrison (2011)

Sin embargo, elegir el panel superior desconociendo su desviación estándar, puede llegar a ser un grave error, ya que se estaría implementando un programa que tenga efectos positivos sólo para unos pocos. Esto ocurre cuando las personas reaccionan al programa de manera diferente, habiendo tanto "ganadores" como "perdedores" dentro de un mismo programa. No obstante, estos *efectos intradistributivos*, como los denomina Harrison (2011) no son observables a partir de los resultados de un ECA, ya que éste sólo proporciona efectos promedio. Esta información que no es capturada por el método es de rigurosa importancia a la hora de realizar análisis de bienestar adecuados, pues de ellos dependen las recomendaciones o lecciones de política. En otras palabras, no se puede conocer qué tan homogéneo sea el efecto causal dentro de la población estudiada.

Esta situación plantea la idea de que dos personas o más puedan reaccionar de manera diferente al mismo tratamiento. Lo anterior corresponde, a una de las críticas más relevantes del método ECA, debido a que en la realidad los efectos de un programa pueden ser significativamente diferentes entre los individuos. Kravitz, Duan & Baslow (2004) ofrecen dos razones que pueden determinar las diferencias de los efectos de los tratamientos: la primera, hace referencia a que todos los seres humanos no responden de la misma forma al tratamiento, es decir, la *receptividad* es diferente para cada individuo; y en segundo lugar, existen personas que responden demasiado fuerte al tratamiento, por lo tanto son muy *susceptibles* al incentivo que ofrecen los programas.

Así, el hecho de que una persona reciba un programa y se vea beneficiada por este, no quiere decir que la otra persona que también recibe el programa va a ser beneficiada de la misma manera, ya que estos individuos pueden diferenciarse por razones fisiológicas, ambientales, institucionales, etc. (Favereau, 2014). Por lo tanto, los efectos producidos por el tratamiento en cada una de las personas serán diferentes. La aleatorización no permite observar la heterogeneidad potencial de los efectos del tratamiento y es por esta razón, que Deaton (2009) argumenta que los ECA sólo están interesados en brindar información acerca de la media de los efectos del tratamiento, dejando de lado información referente a las características de la distribución de la población objeto de estudio, la cual es de vital importancia en el tipo de políticas que combaten la pobreza.

Cuando el enfoque J-PAL se centra en el efecto promedio del tratamiento, asume que este impacto será igual para todos sin importar quién lo recibe, no obstante, en la lucha contra la pobreza quien recibe el impacto parece crucial. Conformarse con ATE es asumir que el promedio es representativo, pero también es suponer que el impacto es el mismo para todas las personas (Favereau, 2014). Heckman (1991) también se une a esta crítica al afirmar que los efectos de los programas no son transferibles entre los individuos, es decir, los efectos positivos de un individuo son sólo para éste y no para ningún otro.

Otra de las críticas principales a los ECA es su falta de respuesta a por qué un programa funciona, pues a partir de los experimentos se puede determinar si el programa funciona, pero no por qué. Deaton (2009) ha establecido que los ECA funcionan como un "recuadro negro", donde a partir del método se hace muy difícil explicar los resultados. No saber por qué funciona un programa es no aprender de esta experiencia, lo cual no proporciona la información necesaria para los hacedores de política. De acuerdo a Cartwright (2007) la utilidad de los resultados del método ECA se ve seriamente comprometida debido a la falta de un modelo causal que explique los efectos causales, sin ello las experiencias de J-PAL difícilmente contribuirá a comprender los fenómenos de la pobreza.

3.2.2 Los problemas de ECA en la práctica: Perturbadores de ATE

Los experimentos se ven afectados por una serie de situaciones que sólo son visibles en la práctica, y afectan tanto a los experimentos económicos como a los experimentos clínicos. Con la presencia de los efectos que serán estudiados a continuación, el poder de la aleatorización se desvanece al afectar la creación de grupos con características similares, distorsionando el efecto promedio de tratamiento, ATE. Lo anterior, entra en el análisis de la validez interna, la cual cuestiona la identificación del efecto causal.

Un elemento diferenciador entre los experimentos clínicos y económicos se refiere al efecto placebo. En la medicina, el uso del placebo[16] se encuentra en los Ensayos Clínicos Controlados; comúnmente, es utilizado para que el paciente no sepa que está haciendo parte de un experimento. También, en los ensayos clínicos se utilizan los estudios de doble ciego, es decir, que los participantes del experimento no deben saber si están recibiendo el

[16] Un placebo es una sustancia farmacológicamente inerte que se utiliza como control en un ensayo clínico.

fármaco real o el placebo, y el administrador del fármaco no debe saber cuál de los dos está suministrando.

Por su parte, en la economía los investigadores de J-PAL tratan de incorporar tanto el efecto placebo como estudios de doble ciego a las evaluaciones aleatorias. No obstante, el efecto placebo no es muy claro al momento de evaluar los programas del desarrollo; usualmente es difícil encontrar un equivalente al placebo utilizado en medicina por lo que en algunos casos es casi inevitable que los sujetos no solo noten que están siendo sometidos al experimento, sino que noten a qué grupo pertenecen (García, 2010). Sería muy extraño en un programa de capacitación laboral que otorgaran charlas de capacitación completamente inútiles a los participantes del grupo de control con el propósito de que no noten que no son beneficiarios del programa.

En este sentido, si el efecto placebo no se cumple en las evaluaciones aleatorias en la economía, provocaría algunos problemas como la *autoselección* en los programas, ya que es difícil que se pueda obligar a las personas a aceptar un tratamiento que es voluntario; lo mismo puede ocurrir, con los sujetos que hacen parte del grupo de control y decidan abandonar el programa. La autoselección y el abandono en los grupos del programa debilitan el poder de la aleatorización, generando un sesgo de selección que se traduce en grupos de comparación con características diferentes afectando, en últimas, la identificación del ATE.

Además de presentar problemas de sesgo de selección, las evaluaciones aleatorias presentan críticas de orden ético. Con la asignación al azar, los investigadores no asignan arbitrariamente quiénes participan, o no, en los programas que van a hacer evaluados. Ahora bien, todo programa generará consecuencias tanto en el corto como en el largo plazo. Imagínese un programa que otorgue crédito educativo: para aquellas personas que no lo reciban podrían tener consecuencias negativas muy grandes por el resto de sus vidas (García, 2010). Algo parecido podría ocurrir con un programa alimenticio donde la vida de las personas puede estar en juego. Todo ello genera una serie de cuestionamientos de orden ético a la aplicación de experimentos en la economía del desarrollo.

3.2.3 Efecto Hawthorne y John Henry

En 1924, la compañía Western Electric Company Works en Illinois realizó una serie de experimentos para examinar el efecto de diversos niveles de iluminación sobre la productividad del trabajador. Se conformaron así, un grupo de tratamiento y un grupo de control; el primer grupo, fue expuesto a diversas intensidades de iluminación, y el otro grupo, trabajó bajo una intensidad luminosa constante. Los investigadores esperaban que la producción individual estuviera directamente relacionada con la intensidad de la luz. Sin embargo, se halló que cuando el nivel de la luz aumentaba en el grupo experimental, la producción se acrecentaba en ambos grupos. A su vez, cuando el nivel luminoso se redujo en el grupo experimental, la productividad no se veía afectada, por lo cual concluyeron que la iluminación no estaba directamente relacionada con la productividad de los trabajadores (McCarney et al., 2007).

Años más tarde, con la colaboración del profesor Elton Mayo se siguieron realizando los experimentos. Al terminar los estudios, se concluyó que existía un efecto motivador en los obreros al saber que estaban siendo objeto de estudio, actuando como el experimentador quisiera que actuara, específicamente en el grupo de tratamiento. Esta modificación en la conducta se define como el *efecto Hawthorne*. Los experimentos económicos liderados por J-PAL, como toda investigación experimental sufre de estos mismos problemas, debilitando de igual manera el accionar de la aleatorización y generando un sesgo significativo en ATE.

Por otra parte, el *efecto John Henry* hace referencia al grupo de control, cuyos participantes en reacción hostil al no ser parte del grupo de tratamiento, modifican su comportamiento. Esto genera, al igual que el efecto Hawthorne un sesgo de selección, y una mala identificación del efecto causal.

3.2.4 Efecto desbordamiento (Spillover Effect)

El efecto desbordamiento se puede presenciar en casos particulares, donde existe el riesgo que se filtre el efecto que ha obtenido el tratamiento hacia el grupo de los que no han sido tratados. Este efecto se ilustra en el experimento realizado por Cohen & Dupas (2010) en el cual se evalúa un programa para reducir la infección de la malaria por medio de la

distribución de toldillos. En aquella evaluación se suministró toldillos de manera gratuita a la población que hace parte del grupo de tratamiento, mientras que al grupo de control no se le suministró. En el desarrollo del experimento se evidenció que sin haber obtenido mosquiteros los del grupo de control, igualmente se vieron beneficiados por quienes sí los tenían, reduciendo la incidencia de la malaria para los dos grupos. La razón estribó en que mientras la mitad de la población se protegía de la malaria, se redujo la probabilidad de contagio para toda la población en general.

Este es propiamente un ejemplo del efecto desbordamiento en los experimentos, que aunque brindó una externalidad positiva para la población que hacía parte del grupo de control, afectó la identificación del efecto tratamiento promedio al sesgar los resultados. Otra forma de generar un sesgo de selección y afectar la validez interna del estudio, sucede cuando las personas deciden no participar en el experimento. Según Heckman (1991), esto introduce un sesgo significativo, dado que si algunas personas de la muestra representativa no quieren participar, se excluirían desde el principio introduciendo un sesgo de selección que la aleatorización supone cancelar.

3.2.5 Sesgo debido a la violación de los supuestos del modelo NRH

Como se estudió en el segundo capítulo, la asignación al azar es el elemento fundamental para que el modelo NRH tenga validez y robustez. En la práctica, hay algunos efectos que debilitan la aleatorización (Efecto Hawthorne, John Henry y Spillover), generando un sesgo de selección en el modelo NRH.

El supuesto de independencia del modelo se refiere a que los resultados potenciales son estadísticamente independientes del tratamiento. Esta condición de independencia se logra sólo en presencia de la aleatorización, por lo cual, si los efectos señalados se introducen en un experimento aleatorio, la condición de independencia se rompería.

Por otro lado, el supuesto de efecto constante, asume que para dos unidades, el efecto que se produce para ambos individuos debe ser exactamente igual. Este supuesto se rompe con la presencia del efecto heterogéneo del tratamiento, en el cual, los individuos reaccionan de manera diferente ante la presencia de un mismo programa. Por lo cual, en la práctica es muy difícil pensar en que se encuentre un efecto constante para una determinada

población. En esta misma línea, se encuentra el supuesto de la homogeneidad de la unidad, el cual se refiere a que los individuos que son seleccionados de la muestra comparten características similares, hecho que en la vida real es poco factible, ya que las personas en una población determinada cuenta con características socioculturales diferentes.

Ahora bien, una vez presentadas las limitaciones que afectan la validez interna del estudio experimental, es consecuente realizar un análisis sobre los alcances del método ECA en la economía del desarrollo. La validez externa del método ECA, ha sido el más criticado tanto por los más reconocidos economistas del desarrollo, como por filósofos interesados en la inferencia causal.

3.3 La escasa validez externa de los experimentos aleatorios

Para Nancy Cartwright (2009c) el análisis de la validez externa de los experimentos aleatorios se concentra en tres conceptos fundamentales, los cuales son: *eficacia real*, *efectividad potencial* y *capacidad causal*. Como primera medida, existe una distinción entre eficacia real y efectividad potencial; la primera, hace referencia al resultado obtenido a partir de un experimento en particular, mientras que la segunda, se refiere a resultados que se pueden obtener de un experimento en otra situación o lugar a partir de un experimento en particular.

Retomando el ejemplo anterior, los conceptos de eficacia real y efectividad potencial se ilustran de la siguiente manera: el hecho de considerar que en México el programa Progresa disminuya la deserción estudiantil se refiere a una eficacia real, pues es éste el que resulta de un experimento en particular. Ahora bien, si éste programa puede ser eficaz de igual manera en Colombia, se refiere, entonces, a una efectividad potencial del resultado obtenido en México. Analizando la validez externa, lo que se está buscando es encontrar el alcance de las experiencias realizadas por el enfoque J-PAL, en últimas, es preguntarse por ¿cuál es la utilidad de los resultados experimentales?

La distinción entre ambos conceptos sugiere el paso entre los resultados obtenidos de una situación en particular a otra situación, esto es de suma importancia para el análisis de las evaluaciones aleatorias utilizadas por J-PAL, debido a que si los ECA sólo encuentran eficacia real y no logran encontrar efectividad potencial, entonces no se cumpliría el

segundo objetivo de ellos. Luego, la pregunta a resolver es ¿cómo pasar de una eficacia real a una efectividad potencial? Para Cartwright (2009c) es posible realizar el paso de tener una buena razón a pensar que el resultado puede tener el mismo efecto en otro lugar (Favereau, 2014). Estas buenas razones se basan en el tercer concepto desarrollado por Cartwritght, llamado *capacidades causales*—que no es más que el poder intrínseco de causar algún tipo de efecto.

De acuerdo a Cartwright (1989) las afirmaciones causales que se presentan en la ciencia no se basan en conjunciones constantes o regularidades de eventos, sino en capacidades que subyacen de los fenómenos. Cuando se dice que la aspirina tiene la capacidad de aliviar el dolor de cabeza, lo que se está diciendo es que existe una entidad con la propiedad de producir un resultado (Ivarola, 2015). Lo anterior, no se debe entender como que siempre que se tome una aspirina se aliviará la migraña, ni tampoco que la alivie la mayoría de veces; en lugar de ello, lo que simplemente se dice es que existe una capacidad estable y relativamente duradera que una entidad lleva consigo misma de caso en caso (Cartwright, 1989).

En el movimiento de la Política Basada en la Evidencia de la cual hace parte J-PAL, se tiene el principio fundamental de usar políticas que funcionaron en algún lugar y momento determinado. El esfuerzo de este enfoque se concentra en encontrar factores causales invariantes a partir del método ECA, los cuales se utilizan en diferentes escenarios. No obstante, para Cartwright & Hardie (2013) estos presentan problemas de validez externa debido a que sólo proporcionan información respecto de una política que funcionó en *alguna parte.* Para llegar al objetivo propuesto por J-PAL, es necesario responder a tres interrogantes causales: ¿funciona en alguna parte? ¿Funciona en general? ¿Sólo funciona aquí y ahora?

Según Ivarola (2015) estas preguntas son ineludibles si se pretende tener éxito en la implementación de una política. Para ello, es necesario conocer cómo un factor causal produce un determinado efecto (*funciona en alguna parte*), sin embargo, saber lo anterior no implica que funcionará para un caso objetivo (*funcionará aquí y ahora*), ni mucho menos que funcionará siempre (*funciona en general*). Luego, la extrapolación de políticas no está garantizada simplemente porque está basada en una inferencia inductiva

(Cartwright, 2011). Al momento de considerar la aplicación de una política exitosa en un escenario diferente al cual fue evaluado, se debe llegar a la pregunta: ¿por qué esperar a que esta política tenga éxito en condiciones diferentes?

Para Cartwright (2009c):

> No tiene sentido hablar de la eficacia real en otro entorno. El concepto sólo se aplica a aquellas situaciones en las cuales se obtiene el diseño experimental, y no en otros. (p.6)

Por lo tanto, es un exabrupto tratar a la eficacia real como si fuera una efectividad potencial cuando en realidad no lo es, y no sirve de ninguna manera para dar recomendaciones de política, tal como lo pretende el enfoque J-PAL. Con la aleatorización se produce eficacia real, pero creen producir efectividad potencial. Es un error atribuir una dependencia de una población específica a la efectividad potencial, ya que así sería imposible pensar en otro contexto; y es por ello que el concepto de capacidades causales es de suma importancia, pues para llegar a desarrollar recomendaciones de política se debe hacer con base a las capacidades causales y no por medio de la eficacia real. Por consiguiente, es la capacidad causal la que permite hacer el paso entre la eficacia real a la efectividad potencial.

Con el programa liderado por el Banco Mundial para combatir la desnutrición infantil en Bangladesh (BINP, siglas en inglés) realizado en 1995, se observa cómo pueden llegar a fracasar las políticas basadas en la evidencia. Este programa fue implementado en Bangladesh sobre la base de un programa exitoso llevado a cabo en la India llamado *"Indian Tamil Nadu Integrated Project"* (TINP). Este programa fracasó en Bangladesh en su objetivo de disminuir la desnutrición de los niños, pero ¿por qué no funcionó el BINP, si el TINP logró tener éxito? ¿Qué fue lo que falló?

El programa BINP tenía como objetivo mejorar el estatus nutricional de mujeres embrazadas y en periodos de lactancia, así como también de infantes en las comunidades pobres de Bangladesh. Este proporcionaba asesoramiento nutricional a mujeres embarazadas y brindaba alimentación suplementaria para niños por debajo de los 24 meses. Por medio del asesoramiento se esperaba cambiar ciertas creencias de los habitantes de

Bangladesh que perjudicaban la nutrición de los niños, por ejemplo, la creencia de que al comer menos, los bebés serían más pequeños, lo cual facilitaba cargarlos (White, 2009). Mientras que la alimentación suplementaria mejoraba significativamente el nivel nutricional de los infantes.

Ahora bien, las razones por las cuales el programa BINP no tuvo el efecto esperado se debió fundamentalmente a quién ejercía el control de los alimentos en el hogar, el cual permitió la existencia de dos factores que no se tuvieron en cuenta en la implementación del programa en Bangladesh: La *filtración* y la *sustitución* de alimentos. Por un lado, la comida que era suministrada para ser complementaria se convirtió en sustituta y por el otro lado, los alimentos destinados a las madres y niños terminaban siendo desviados hacia otros miembros de la familia. Estos dos factores estaban vinculados con un patrón sociocultural muy significativo para el resultado de la política, y es el hecho de que en Bangladesh son las suegras las que llevan la administración del hogar, por el contrario, en la India, las madres eran las encargadas (Cartwright, 2011b).

Así, para Bangladesh, resultó ser significativo el factor "las madres no tienen el control del hogar", distinto a lo que ocurría en la India, lo cual contribuyó a la generación de resultados muy diferentes a los esperados. Por lo tanto, Cartwright & Hardie (2013) señalan de que se necesita tener mucha más información que simplemente el efecto causal; es necesario también conocer los factores contextuales que la complementan, denominados *factores coadyuvantes.*

De esta manera, el conocimiento de factores causales es necesario mas no suficiente, y es aquí donde adquiere relevancia el desarrollo de una investigación tanto vertical como horizontal. El primero, permite la identificación correcta del efecto causal, que una vez identificado, produzca un principio general que sea aplicable tanto al sistema bajo estudio como a otras poblaciones. Mientras que el segundo, consiste en examinar si los factores coadyuvantes obtenidos en la población estudiada también se van a obtener en otras poblaciones. Es decir, se trata de especificar los elementos puntuales que serán cruciales para el éxito de una política en un contexto determinado.

De acuerdo al ejemplo, bajo la investigación vertical se obtiene el principio 1:

> Mejor conocimiento nutricional de las madres más suministro de alimentación suplementaria pueden mejorar el estatus nutricional de los infantes. (Cartwright, 2011b, p.22).

Este principio está condicionado a las características de la población que fue objeto de estudio. Por lo tanto, sería un error tratar de extrapolarlo a otras poblaciones, de aquí que, sea necesario llegar a un principio más específico que se obtiene a partir de una investigación horizontal, tal como se muestra de la siguiente manera:

> Principio 2: Mayor conocimiento nutricional puede mejorar la nutrición de los niños si las personas que tienen el conocimiento son aquellas que: proveen al niño con alimentación suplementaria, controlan qué comida es adquirida, controlan cómo la comida es administrada y mantienen los intereses del niño como elemento central.(Cartwright, 2011b, p.23).

Lo anterior da cuenta de que el enfoque de J-PAL por medio de los experimentos aleatorios por sí solos, no puede dar recomendaciones de política sin pasar primero por una investigación horizontal. No es suficiente el descubrimiento de un efecto causal para implementar una política, ya que cada escenario contiene factores muy específicos. Por lo tanto, la generalización de los resultados experimentales contiene un grado de dificultad muy amplio si su interés se encuentra en reducir la pobreza global por medio de lecciones de política, provenientes de los métodos ECA.

Sin embargo, Banerjee & Duflo (2011) han planteado que la escasa validez externa de los experimentos aleatorios se puede corregir mediante la replicación de los experimentos en contextos diferentes, dando así evidencia de la eficacia real sujeta a dichas condiciones. No obstante, autores como Rodrick (2008) y Cohen & Easterly (2009) han mostrado que el incentivo de replicar un experimento se pierde en la medida que éste se repite. Por otra parte, no es tan claro cuántas repeticiones son necesarias o qué tipo de repeticiones son suficientes para establecer una generalización. Además se suma el hecho de que los experimentos aleatorios son costos y el incentivo de invertir el dinero en un proyecto de evaluación también se pierde con la replicación, lo mismo sucede con el incentivo académico para mostrar algo innovador. Aun así, si se tuviera la oportunidad de repetir un

experimento n-veces buscando establecer un efecto causal estable, siempre existiría el interrogante de saber si en la próxima oportunidad funcionará.

2.4 Conclusiones capítulo 3

Desde la perspectiva de Cartwright (2007) los ECA no son un estándar de oro y no deben considerarse como tal. No hay un solo método en la investigación que sea mejor que otro, ni que proporcione una mejor evidencia, habrá lugar en el que ECA sea un buen método y en otros no. El estudio de la validez interna del método arrojó debilidades en la identificación de los efectos causales, principalmente por la presencia de factores incontrolados que ocurren en la práctica de los diseños experimentales (efecto Hawthorne, John Henry y desbordamiento). Esta debilidad afecta el poder de la aleatorización, ya que genera sesgos de selección los cuales, se supone, serían cancelados con la asignación al azar.

Ante esto, se rompe la idea de que las evaluaciones aleatorias gozan de una alta validez interna, pues como se vio en este capítulo, la identificación del efecto causal puede verse comprometida y no ser plenamente identificada. Si lo anterior ocurre, la extrapolación de los resultados no sería una tarea de gran valor, debido a que el resultado a generalizar sería consecuentemente erróneo.

Estudiando los conceptos de eficacia real, efectividad potencial y capacidad causal, se encontró que los ECA proporcionan eficacia real, la cual corresponde a ATE, sin embargo es tratada como si fuera una efectividad potencial por parte del equipo de investigadores de J-PAL. Estos pretenden generar recomendaciones de política a partir de ATE y no de las capacidades causales, las cuales crean el puente para llegar a la extrapolación de los resultados y es el camino idóneo hacia la generalización. Los problemas estudiados en este capítulo que afectan la validez interna y externa ponen en tela de juicio la supuesta rigurosidad del método. No obstante, dichos problemas no invalidan como tal el método, pero sí pone limite a los conocimientos que se puedan extraer de este.

Conclusiones

El debate metodológico y epistemológico siempre es necesario en la medida en que la ciencia económica es una disciplina en constante cambio. Los nuevos desarrollos dentro de la disciplina exigen siempre un examen filosófico y una mirada crítica de los mismos. La ciencia económica ha adquirido un apreciable grado de autoconciencia metodológica que subyace a las constantes preocupaciones, problemas y retos que nacen día a día de cada coyuntura en la práctica económica.

En la economía del desarrollo con la incursión de los Ensayos Controlados Aleatorios se ha transformado completamente la manera en que se genera conocimiento. El desarrollo de dicha metodología se debió principalmente a la falta de evidencia de los macroeconomistas en no poder ofrecer evidencia contundente sobre las políticas que funcionan en la promoción del desarrollo económico. Este vacío dio lugar a un nuevo movimiento empírico el cual sustituyó las grandes preguntas sobre los factores determinantes de la pobreza por otras más pequeñas, pero no por ello menos importantes, que pudieran ser respondidas por medio de la metodología ECA.

Lo anterior dio paso a que los ECA fueran considerados como un estándar de oro en las evaluaciones de impacto, creando evidencia rigurosa sobre la eficacia y efectividad de los programas del desarrollo. No obstante, a partir de las críticas realizadas por Rodrick (2008), Deaton (2009), Cohen & Easterly (2009) y Cartwright (2011) se determinó que la metodología ECA en el campo experimental no es, en primer lugar, un estándar de oro, ni es superior a otras metodologías que buscan efectos causales. Esto debido a que en la práctica económica la validez interna se ve comprometida con la presencia de efectos que distorsionan la identificación causal por medio de ATE, los cuales no invalidan el método pero sí generan sesgo en los resultados. Por otra parte, aun así, teniendo una identificación causal correcta, la generalización de los resultados es compleja, pues la metodología ECA encuentra eficacia real y no efectividad potencial.

Por lo tanto, los dos objetivos de J-PAL se debilitan en el sentido de que es poco posible pasar de una situación particular a generar políticas que combatan la pobreza mundial. Sin embargo, a pesar de sus limitaciones, los ECA se han convertido en otra herramienta válida para estudiar el desarrollo y son hoy en día, la apuesta que está revolucionando el campo empírico.

Bibliografía

Aedo, C. (2005). *Evaluación del impacto.* Santiago de Chile: División de Desarrollo Económico, CEPAL .

Angrist, J., & Pischke, J. S. (2010). The Credibility Revolution in Empirical Economics: how better research design is taking the con out of econometrics. *NATIONAL BUREAU OF ECONOMIC RESEARCH* (Working Paper 15794), 1-37.

Armitage, P. (2003). Fisher, Bradford Hill, and randomization. *International Journal of Epidemiology, 32*(6), 925-928.

Attanasio, O., Meghir, C., & Santiago, A. (2011). Education choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresa. *Review of economic Studies*.

Baker, L. (2000). *Evaluación de impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales.* Washington DC: Banco Mundial.

Banco Mundial. (19 de Noviembre de 2014). *The World Bank.* Recuperado el 20 de Abril de 2016, de www.worldbank.org

Banerjee, A. V. (2007). *Making Aid Work.* Cambridge, MA: MIT Press.

Banerjee, A. V., & Duflo, E. (2011). *Repensar la pobreza : un giro radical en la lucha contra la desigualdad global.* Madrid: Taurus.

Banerjee, A., & Duflo, E. (2009). L'approche expérimentale en économie du développement. *Revue d'économie politique, 119*, 691-726.

Baslow, J., Duran, N., & Kravitz, R. (2004). Evidence-Based Medicine, Heterogeneity of Treatment Effects, and The Trouble with average. *The Milbank Quaterly*, 661-687.

Bernstein, J. (2004). Evidence-based medicine. *J Am Acad Orthop Surg.*, 8-80.

Box , J. F. (1980). R.A. Fisher and the Design of Experiments. *The American Statistician, 34*, 1-7.

Brady, H. E. (2002). Models of Causal Inference: Going Beyond the Neyman-Rubin-Holland theory. *paper presented at the Annual Meetings of the Political Methodology Group University of Washington, Seattle.* Washington.

Bunge , M. (1961). *Causalidad; el principio de la causalidad en la ciencia moderna.* Buenos Aires: EUDEBA .

Cacheiro, L. (2011). *Introducción a la Inferencia Causal*. Obtenido de (Tesis de Licenciatura, UNIVERSIDAD DE BUENOS AIRES): http://cms.dm.uba.ar/academico/carreras/licenciatura/tesis/2011/Cacheiro_Laura.pdf

Cartwright, N. (1989). Nature's Capacities and Their Measurement. *Clarendon Press, Oxford*.

Cartwright, N. (2007). Are RCTs the Gold Standard? *Biosocieties, 2*, 11-20.

Cartwright, N. (2009c). What Is This Thing Called Efficacy? *Philosophy of the Social Science*, 1-20.

Cartwright, N. (2011a). A Philosopher's View of the Long Road From RCT's to Effectiveness. *The Lancet*, 1400-1401.

Cartwright, N. (2011b). Evidence, External Validity and Explanatory Relevance. *In Philosophy of Science Matters: The Philosophy of Peter Achinstein*.

Cartwright, N., & Hardie, J. (2013). *Evidence-Based Policy: A Practical Guide To Doing It Better.* Oxford: 1st Edition.

Chamorro, Á., Alonso, P., Arrizabalaga, J., Carné, X., & Camps, V. (2001). Luces y sombras de la medicina basada en la evidencia: el ejemplo del accidente vascular cerebral. *Medicina Clínica, 116*(9), 343-349.

Chenery, H. B. (1952). Overcapacity and the acceleration principle. *Econometrica, 20*(1), 1-28.

Clavijo, S. (2012). Economía de la pobreza: repensando la lucha contra la pobreza global. *Torre de Marfil*.

Cohen, J., & Dupas, P. (2010). Free Distribution Or Cost-Sharing? Evidence From a Randomized Malaria Prevention Experiment. *The Quartely Journal Of Economics*, 1-30.

Cohen, J., & Easterly, W. (Edits.). (2009). *What Works in Development?: Thinking Big and Thinking Small.* Washington, D.C.: Brookings Institution Press.

Davies , P. T. (2004). Is Evidence-Based Government Possible? *Presented at the 4 Campbell Collaboration Colloquium* . Washinton D.C.

Deaton, A. S. (2009). INSTRUMENTS OF DEVELOPMENT:Randomization in the tropics, and the search for the elusive keys to economic development. *NBER WORKING PAPER SERIES* (14690).

Demirdjian, G. (2006). Historia de los ensayos clinicos aleatorizados. *Arch.argent.pediatr, 104*(1), 52-61.

Dodgson, S. J. (2006). The evolution of clinical trials. *The Journal of the European Medical Writers Association, 15*(1), 20-21.

Godeau, P. (Productor), & Dormael , J. V. (Dirección). (2009). *Mr. Nobody* [Película]. Bélgica .

Duflo, E. (2009). *Expérience, science et lutte contre la pauvreté.* Paris : Fayard.

Duflo, E. (2010). *Experimentos sociales para luchar contra la pobreza*. . [Video Web]. Congreso llevado a cabo en Long Beach, CA. Recuperado de: http://www.ted.com/talks

Duflo, E., Glennerster, R., & Kremer, M. (2006). Using randomization in development economics research: a toolkit. *NBER Technical Working Paper Series*(333).

Favereau, J. (2014). *The J-PAL's experimental approach in development economics: an epistemological turn?* Tesis doctoral, Université Paris 1 Pantheon-Sorbonne.

Fechner, G. T. (1860). *Elemente der Psychophysik.* Lausanne: Breitkopf und Härtel.

Fienberg, S. E. (1971). Randomization and Social Affairs: The 1970 Draft Lottery. *Science, New Series, 171*(3968), 255-261.

Fisher, R. A. (1935). *The Design of Experiments.* Edinburgh: Oliver & Boyd.

Fisher, R. A. (1947). Development of the theory of experimental design. *Proceedings of the International Statistical Conferences, 3*, págs. 434-439.

García Núñez, L. (2010). Econometría de evaluación de impacto. *Pontifica Universidad Católica del Perú* (Documento de Economía núm. 283.).

Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. (2011). *La evaluación de impacto en la práctica.* Washington DC: Banco Mundial .

Guala, F. (2005). *The methodology of experimental economics.* New York: Cambridge University Press.

Guillen, H. R. (2015). La Economía del Desarrollo rebaja sus ambiciones: las experimentaciones por asignación aleatoria de Duflo. *Economíaunam, 12*(36), 34-48.

Hall, N. S. (2007). R. A. Fisher and his advocacy of randomization. *Journal of the History of Biology, 40*, 295-325.

Harberger, A. (1993). Secrets of Success. A Handful of Heroes. *American Economic Review, 83*(2), 343-350.

Harrison, G. (2011). Randomization and Its Discontents. *Journal of African Economies*, 1-33.

Heckman, J. (1991). Randomization and Social Evaluation. *NBER Working Paper*, 107.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association, 81*(396), 945-960.

Hume, D. (1777). *An Enquiry concerning Human Understanding.* Oxford: Clarendon Press.

Hurtado, J. (2014). Albert Hirschman y la economía del desarrollo: Lecciones para el presente. *Cuadernos de Economía, 33*(62), 7-31.

Ivarola, L. (2015). Sobre el Descubrimiento y Uso de Factores Causales: El Aporte de Nancy Cartwright. *Fragmentos de Filosofía*, 59-78.

Kuehl, R. O. (2007). *Diseño de experimentos.* México : Thomson Editores, S.A.

Labrousse, A. (2010). Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement. *Revue de la régulation*(7 (Primer Semestre)).

Lee, M. J. (2005). *Micro-Econometrics for Policy, Program, and Treatment Effects.* New York: Oxford University Press.

Martínez, R. (1995). LA FILOSOFIA DE GALILEO Y LA CONCEPTUALIZACION DE LA CAUSALIDAD FISICA. *Thémata*(14), 37-59.

Mayneris, F. (5 de mayo de 2009). « L'économie du développement à l'épreuve du terrain – Entretien avec Esther Duflo ». *La vie des idées*.

McCarney, R., Warner, J., Ilife, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: a randomised, controlled trial. *BMC Medical Research Methodology*, 7-30.

Meier, G. M. (2002). La vieja generación de economistas del desarrollo y la nueva. En G. M. MEIER, & J. E. STIGLITZ (Edits.), *FRONTERAS DE LA ECONOMÍA DEL DESARROLLO: El futuro en perspectiva* (págs. 1-37). Bogotá: ALFAOMEGA GRUPO EDITOR.

Milne, I. (2012). Who was James Lind, and what exactly did he achieve. *Journal of the Royal Society of Medicine, 105*(12), 503-508.

Moore, D. S. (2010). *Estadística aplicada básica.* Barcelona: Antoni Bosch.

Mukherjee, S. (2010). *El emperador de todos los males. Una biografía del cáncer* . Madrid: Taurus.

Navarro, H., King , K., Ortegón , E., & Pacheco , J. F. (2006). *Pauta metodológica de evaluación de impacto ex-ante y ex-post de programas sociales de lucha contra la pobreza.* Santiago de Chile: ILPES, CEPAL .

Norton, S. B., Cormier, S. M., & Suter II, G. W. (2014). *Ecological Causal Assessment.* Boca Ratón: CRC Press.

Nurkse, R. (1961). Equilibrium and Growth in the World Economy. En G. Harbeler, & R. Stern (Edits.), *economic essays by Ragnar Nurkse.* Cambridge: Harvard University Press.

Parker, I. (17 de Mayo de 2010). The Poverty Lab: Transforming development economics, one experiment at a time. *The New Yorker*.

Pinilla Palleja, R. (2006). Política Basada en la Evidencia para la Renovación del Estado de Bienestar. *XII Encuentro de Economía Pública* . Mallorca.

Porter, T. M. (2004). *Karl Pearson. The scientific life in a statistical age.* Princeton: Princeton University Press.

Prebisch, R. (1950). The Economic Development of Latin America and its Principal. *Economic Bulletin for Latin America, 7*(1), 1-22.

Ravallion, M. (2009). Should the Randomistas Rule? *The Economists' Voice, 6*(2), 1-5.

Rodríguez, M. (2005). La medicina basada en la evidencia y la práctica médica. *Revista Cubana de Medicina, 44*, 3-4.

Rodrik, D. (2008). The New Development Economics: We Shall Experiment, But How Shall we learn? *Harvard Kennedy School Working Paper* (RWP08-055).

Rojas Osorio, C. (2006). *La filosofía: sus transformaciones en el tiempo.* San Juan: Isla Negra Editores.

Rosenstein-Rodan, P. (1943). "Problems of Industrialization of Eastern and Southeastern Europe.". *Economic Journal*(53), 11-210.

Rostow, W. (1960). *The Stages of Economic Growth: A Non-Communist Manifesto.* Cambridge: Cambridge University Press.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology,* 688-701.

Saavedra Santana, P. (2010). Diseños experimentales y ensayos clínicos. En A. M. Cejas (Ed.), *Las matemáticas del siglo XX una mirada en 101 artículos* (págs. 147-150). Sociedad Canaria Isaac Newton de Profesores de Matemáticas.

Sackett, D. L., Rosenberg, W. M., Gray , J. A., Haynes, R. B., & Richardson , W. S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ, 312,* 71-72.

Salomone, J. (2012). Randomized controlled trials at war (Tesis de maestría). Paris: Paris School of Economics.

Schopenhauer, A. (1967). *De la cuadruple raíz del principio de razón suficiente.* Madrid: Editorial Aguilar.

Stern, N. (2002). Prólogo. En J. Stiglitz, & G. Meier (Edits.), *Fronteras de la economía del desarrrollo. El futuro en perspectiva* (Primera, en español. ed., págs. 8-9). Bogotá: Banco Mundial en coedición con Alfaomega Colombiana S. A.

Sutcliffe, S., & Court, J. (2005). *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?* London: Overseas Development.

Tajer, C. D. (2010). Ensayos terapéuticos, significación estadística y relevancia clínica. *Revista Argentina de Cardiología, 78,* 385-390.

White, H. (2009). Theeory-Based Impact Evaluation: Principles and Practice. *The International Iniciative For Impact Evaluation*.

Xirau, R. (2000). Introducción a la historia de la filosofía. *Textos Universitarios Universidad Nacional Autónoma de México*.