

## Métodos para evaluar políticas e intervenciones públicas

Javier Martín Román<sup>1\*</sup>

<sup>1</sup> Profesor del Departamento de Economía Aplicada y Gestión Pública, Facultad de Derecho, Universidad Nacional de Educación a Distancia (UNED)

### Resumen

Las políticas públicas están diseñadas para lograr ciertas metas y cubrir a un número concreto de potenciales personas beneficiarias. En este artículo se describen los principales métodos empleados para determinar si las intervenciones públicas realmente funcionan, así como el nivel y la naturaleza de los impactos generados sobre los destinatarios. La necesidad de evaluar el impacto generado por aquellas se justifica, fundamentalmente, por tres razones. En primer lugar, emplaza a los responsables de la gestión y formulación de políticas ("policy makers") a contrastar si los programas están alcanzando los resultados inicialmente previstos (identificación de efectos positivos o negativos). En segundo lugar, la puesta en marcha de evaluaciones de impacto encuentra respaldo en el fomento, promoción y búsqueda de transparencia en la asignación de recursos (rendición de cuentas). Finalmente, también contribuye a mostrar evidencia objetiva de lo que funciona y lo que no (eficacia de la política). El dominio de metodologías como las que aquí se presentan nos permite precisar en qué medida se han cumplido los tres objetivos anteriores. Y, más concretamente, nos ayuda a esclarecer si la causa del cambio observado en la variable de interés puede atribuirse íntegramente a la intervención, finalidad última de toda evaluación de impacto.

**Palabras clave:** *evaluación de impacto, políticas públicas, ensayos controlados aleatorizados, diseños cuasiexperimentales.*

\*Correspondencia: [jmartin@cee.uned.es](mailto:jmartin@cee.uned.es)



Open Access: los artículos publicados en esta revista están en acceso abierto bajo licencia Creative-Commons 4.0 Atribución-No Comercial-Compartir Igual (CC BY-NC-SA)



Publicación del Instituto Mixto de Investigación Escuela Nacional de Sanidad (IMI-ENS), de la UNED y el Instituto de Salud Carlos III



Se recomienda imprimir 2 páginas por hoja

Revista de Gobierno, Administración y Políticas de Salud \_\_\_\_\_ página 1

## Abstract

Public policies are designed to achieve certain goals and cover a defined number of potential beneficiaries. This paper describes the main methods used to determine whether public interventions really work, as well as the level and nature of the impacts generated on the recipients. The need to evaluate the impact of public interventions is justified for three main reasons. First, it requires policy makers to verify whether the programs are reaching the results initially planned (identification of positive or negative effects). Secondly, the implementation of impact evaluations is supported by the encouragement, promotion, and search for transparency in the allocation of resources (accountability). Finally, it also contributes to show objective evidence of what works and what does not (policy effectiveness). Mastery of methodologies such as those presented here enables us to determine to what extent the above three objectives have been met. More specifically, it helps us to clarify whether the cause of the observed change in the variable of interest can be fully attributed to the intervention, which is the main goal of any impact evaluation.

**Keywords:** impact evaluation, public policies, randomized controlled trials, quasi-experimental designs

## ÍNDICE

1. Introducción
2. Inferencia causal: el problema del contrafactual
3. El sesgo de selección
4. Diseños de evaluación
  - 4.1. Diseños experimentales
  - 4.2. Diseños cuasiexperimentales
5. Métodos de estimación
  - 5.1. Diferencias-en-diferencias (diff-in-diff)
    - 5.1.1. Descripción de la técnica
    - 5.1.2. Ejemplos de aplicación práctica
    - 5.1.3. Ventajas y limitaciones del método
  - 5.2. Controles sintéticos (Synthetic Control Methods)
    - 5.2.1. Descripción de la técnica
    - 5.2.2. Ejemplos de aplicación práctica
    - 5.2.3. Ventajas y limitaciones del método
  - 5.3. Matching o emparejamiento
    - 5.3.1. Descripción de las técnicas
    - 5.3.2. Ejemplo de aplicación práctica
    - 5.3.3. Ventajas y limitaciones del método
  - 5.4. Regresión en discontinuidad
    - 5.4.1. Descripción de la técnica
    - 5.4.2. Ejemplos de aplicación práctica
    - 5.4.3. Ventajas y limitaciones del método
6. Resumen y conclusiones

## Referencias bibliográficas

## 1. Introducción

La evaluación de impacto de un programa, proyecto o política, última etapa del denominado marco lógico, tiene por objetivo cuantificar los efectos causales de una intervención pública. Para que las relaciones causales establecidas sean fiables es necesario manejar conceptos como "contrafactual" o "sesgo de selección", así como dominar los principales métodos y técnicas cuantitativas a disposición de los investigadores.

El marco lógico de una política pública, también denominado "teoría del cambio", es una herramienta que permite llevar a cabo la planificación, ejecución y evaluación de aquella. A lo largo de esta "cadena de valor", se pueden identificar diferentes tipos de evaluación: evaluación de necesidades, evaluación del diseño, evaluación de la implementación, evaluación de resultados y, finalmente, evaluación de impacto. Este artículo se detiene en la última.

Cualquier evaluación de impacto debe incluir el cómputo y análisis de todos los efectos, ya sean específicos o globales, buscados o no, positivos o negativos, directos o indirectos. En cuanto a su definición, la Organización para la Cooperación y el Desarrollo Económicos (OCDE) la conceptualiza como sigue: "Término que indica si el proyecto tuvo un efecto en su entorno en términos económicos, técnicos, socioculturales, institucionales y medioambientales". Por su parte, el Banco Mundial

se expresa en los siguiente términos: “Tipo particular de evaluación que pretende responder a una pregunta específica de causa y efecto: cuál es el impacto (o efecto causal) de un programa en un resultado de interés”.

En este artículo se presentan e ilustran los principales elementos característicos de toda evaluación de impacto. De esta forma, en la Sección 2 ya se introduce el concepto de causalidad. En la articulación de cualquier evaluación de impacto se ha de presentar, desde un inicio y explícitamente, la relación existente entre dos variables clave: la política o intervención objeto de estudio ( $P$ ) y el outcome o variable de resultado ( $Y$ ).

A continuación, se define el concepto de “impacto”, se enumeran los principales objetivos asociados a su evaluación, y se aborda una de las cuestiones centrales del artículo: el problema del contrafactual. En última instancia, una evaluación de impacto siempre pretende aislar el efecto de la intervención pública de la influencia ejercida por otros factores externos durante el periodo de ejecución de aquella. El contrafactual se puede definir como la situación hipotética en que se habría encontrado una persona beneficiaria de una política pública en el caso de que no hubiese disfrutado de tal condición. Encontrar un buen contrafactual es, por tanto, imprescindible.

La Sección 3 está destinada al análisis del sesgo de selección. A fin de hacer frente a este obstáculo, tradicionalmente se recurre a dos propuestas metodológicas alternativas: los diseños experimentales y los diseños cuasiexperimentales (Sección 4). Los primeros constituyen la opción más robusta de implementar una evaluación de impacto. No obstante, los diseños experimentales, experimentos sociales o ensayos controlados aleatorizados (Randomized Controlled Trials, RCT), como también se les conoce, plantean importantes restricciones de implementación. Cuando no es posible su utilización, se recurre a técnicas cuasiexperimentales. En este artículo se exponen cinco de los métodos cuantitativos más relevantes en la actualidad (Sección 5).

El artículo concluye con un apartado destinado a las preceptivas conclusiones (Sección 6), remarcando la conveniencia de realizar evaluaciones de las intervenciones públicas, y subrayando la importancia y el potencial tanto de los diseños experimentales como de los cuasiexperimentales, herramientas esenciales para el establecimiento y medición de relaciones causales.

## 2. Inferencia causal: el problema del contrafactual

El análisis de evaluación de impacto se basa en relaciones causales: la política pública ( $P$ ) es la causa y el resultado ( $Y$ ) es el efecto. Para aislar el verdadero efecto de la intervención, es crucial considerar el problema del “contrafactual” y aplicar la ecuación básica de evaluación de impacto. Este proceso implica comparar los resultados observados con otros simulados. La diferencia entre ambos determina el efecto causal de la intervención pública.

El análisis de evaluación de impacto parte de la existencia de relaciones causales. Si consideremos que tenemos a nuestra disposición dos variables,  $P$  e  $Y$ , donde  $P$  (política pública) denota la causa e  $Y$  (outcome) representa el efecto, entonces la causalidad habría de formularse en los siguientes términos (y no al revés):

$$P \rightarrow Y \quad [1]$$

expresión que refleja que los cambios de  $P$  propician u originan modificaciones en  $Y$ .

En relación con el análisis de causalidad, es preciso tener presente el cumplimiento de ciertos criterios metodológicos. Básicamente, los que hacen referencia a la validez y robustez de las estimaciones. Por lo que atañe al primero, cabe destacar la existencia de dos tipos de validez:

- *Validez interna*: aquella que intenta generar un grupo de comparación que sea estadísticamente equivalente al grupo de tratamiento en la línea de base.
- *Validez externa*: aquella vinculada a la posi-

bilidad de extrapolar o generalizar los resultados alcanzados.

La robustez, en cambio, alude a la "estabilidad" de las estimaciones llevadas a cabo. Una medición se considera robusta siempre que un instrumento aplicado de manera sistemática sobre un mismo objeto de análisis proporcione unos resultados semejantes.

Introducidos los conceptos más relevantes del análisis de inferencia causal, el siguiente punto a abordar es la definición de "impacto", término que hace referencia a la variable objetivo a largo plazo que se desea modificar. Por ejemplo: las tasas de mortalidad infantil (reduciéndolas) o las de vacunación frente a la gripe (incrementándolas).

En lo que atañe a los principales objetivos perseguidos por cualquier evaluación de impacto, se pueden concretar en tres:

- 1) Determinar si puede atribuirse, o no, la causa del cambio observado a la intervención pública examinada.
- 2) Identificar si el programa o política produjo los efectos deseados en las personas, hogares o instituciones sobre los que se proyectó.
- 3) Obtener una estimación cuantitativa de sus beneficios potenciales.

La puesta en marcha de una intervención pública suele venir acompañada de otros acontecimientos (por ejemplo: la implementación en paralelo de otros programas alternativos) que pueden ejercer una influencia notable sobre el impacto que se pretende cuantificar. En consecuencia, para evaluar correctamente el mencionado impacto es necesario aislar el efecto generado por la intervención objeto de estudio de la acción de otros elementos o factores externos, puesto que pueden estar interfiriendo con los resultados derivados de la política pública examinada.

¿Cuál es entonces el impacto o efecto causal de un programa o política ( $P$ ) sobre un resultado de

interés ( $Y$ )? La respuesta a esta pregunta se puede hallar aplicando la ecuación básica de la evaluación de impacto, que es la siguiente:

$$\tau = E(Y|P=1) - E(Y|P=0) \quad [2]$$

donde  $\tau$  refleja el impacto o efecto causal del programa,  $Y$  representa la variable de resultado, y  $P$  es una variable dicotómica que indica si el individuo ha recibido ( $P=1$ ), o no ( $P=0$ ), el programa. Asimismo,  $E(\cdot)$  denota el operador de valor esperado o esperanza matemática (valor promedio) de la variable objetivo condicionada a la participación, o no, en el programa.

En otras palabras, para medir el impacto de un programa o intervención pública ( $\tau$ ) basta con comparar, transcurrido cierto intervalo de tiempo, la media que toma el *outcome* en las unidades beneficiarias de la política, frente a ese mismo valor en el supuesto de que esas mismas unidades beneficiarias no hubiesen recibido los efectos del tratamiento. Si bien el primero de los resultados se puede observar y medir fácilmente, resulta imposible determinar, simultáneamente, cuál habría sido el resultado para ese mismo beneficiario, o conjunto de beneficiarios, en ausencia del programa.

Surge entonces un problema de identificación y de falta de observación de uno de los elementos del análisis (problema del *contrafactual*). No podemos conocer el resultado producido en un individuo o unidad tras recibir el tratamiento y, al mismo tiempo, observar el resultado producido en ese mismo individuo o unidad sin pasar por el tratamiento.

El objetivo, por tanto, consiste en simular cómo habrían estado las unidades beneficiarias de la intervención en caso de que no hubiesen recibido los efectos del programa. La comparación de la realidad con el escenario *contrafactual* generado es el mecanismo empleado para aislar el efecto del programa de la influencia de potenciales factores externos agregados.

### 3. El sesgo de selección

El análisis del sesgo de selección es un aspecto esencial en el diseño y ejecución de evaluaciones de impacto. Se pueden identificar hasta cuatro tipos diferentes: de focalización y autoselección, según su origen, y en variables observadas y no observadas, según el tipo de variable analizada.

Uno de los aspectos más importantes a tener en cuenta en el diseño e implementación de evaluaciones de impacto está relacionado con el denominado *sesgo de selección*, cuya existencia produce una desviación en el cálculo del verdadero valor del impacto o efecto causal.

Aplicando un doble criterio de clasificación, se pueden identificar hasta cuatro tipos diferentes de sesgo de selección. Atendiendo al origen o fuente del sesgo, identificamos los dos siguientes:

- a) *Sesgo de focalización*: aquel imputable a la propia política pública llevada a cabo. Una inmensa mayoría de programas no se implementan de manera aleatoria, sino que van dirigidos a grupos concretos de población. Siempre que esto suceda, aparecerá un sesgo de focalización.
- b) *Sesgo de "autoselección"*: aquel atribuible a las potenciales personas destinatarias de una política pública. Surge entonces cuando determinadas personas son capaces de "convertirse" en beneficiarias de una intervención en particular, esto es, cuando tienen poder de decisión para escoger si desean participar en un programa ("autoselección") o, por lo contrario, descartan tal oportunidad.

Una clasificación alternativa es la que se establece en función del tipo de variable analizada:

- a) *Sesgo en variable observada*: es el que aparece cuando existe una característica accesible para el investigador a través una base de datos, como una encuesta o un registro administrativo, que condiciona

el hecho de que un individuo pertenezca al grupo de tratamiento o bien forme parte del de control.

- b) *Sesgo en variable no observada*: es el que aparece cuando existe una característica que no puede ser observada por el investigador a través de una base de datos (por ejemplo: el nivel de confianza o el grado de felicidad), que condiciona el hecho de que una persona pertenezca al grupo de tratamiento o bien forme parte del de control.

### 4. Diseños de evaluación

Existen dos mecanismos para evaluar el impacto de las políticas públicas: los diseños experimentales o experimentos sociales (RCT), opción óptima, pero éticamente cuestionable y costosa; y los diseños "cuasiexperimentales", aunque más económicos y sencillos de articular, siempre incorporan sesgos de selección.

Una estimación directa del impacto de una política se puede obtener de modo sencillo. Basta con comparar los valores promedio de la variable de resultado en el grupo de tratamiento y en el de control. La determinación y construcción de ambos grupos, por su parte, se puede realizar empleando diversos métodos y técnicas cuantitativas. Estas metodologías, a su vez, se subdividen en dos grandes categorías: los diseños experimentales y los diseños *cuasiexperimentales*:

- *Diseños experimentales*: son aquellos en los que, partiendo de una población de potenciales personas beneficiarias, estas terminan participando, o no, como consecuencia de la aplicación de un mecanismo de asignación puramente aleatorio.
- *Diseños cuasiexperimentales*: son aquellos en los que la decisión de participación en el programa no viene definida por un procedimiento aleatorio. En su lugar, surgen tres alternativas: (1) los propios individuos deciden si participan o no; (2) un agente o

regla externa predetermina quién puede beneficiarse de la intervención y quién no; (3) las dos opciones anteriores aparecen combinadas.

#### 4.1. Diseños experimentales

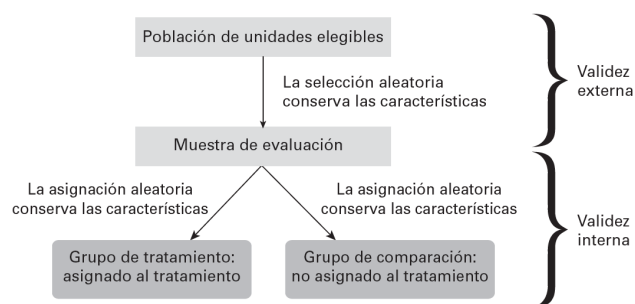
Los diseños experimentales o *Randomized Controlled Trials* (RCT), por su denominación en inglés, representan los métodos más robustos de todas las técnicas cuantitativas de evaluación de impacto. Metodológicamente, la implementación de una evaluación de impacto a través de un experimento social resulta muy parecida al procedimiento seguido en los ensayos clínicos.

Así, con posterioridad a la selección de aquellos individuos susceptibles de recibir los beneficios o efectos positivos de un programa, se lleva a cabo, bajo su consentimiento, un procedimiento de asignación puramente aleatorio. Una parte de la muestra termina formando parte del grupo de tratamiento (por ejemplo: quienes reciben la dosis de una vacuna), mientras que el resto va a parar al grupo de control.

La **Figura 1** esquematiza el funcionamiento de un RCT.

- En una primera etapa, se obtiene una muestra aleatoria a partir de un conjunto de unidades potencialmente beneficiarias del programa. Se garantiza así la *validez externa* del experimento, que mide la posibilidad de generalizar las conclusiones obtenidas a otros programas.
- Seguidamente, la muestra se asigna aleatoriamente a dos grupos: el de tratamiento y el de control. Operando de esta forma se asegura la *validez interna*.

**Figura 1.** Muestra aleatoria y asignación aleatoria de tratamiento



Fuente: [Gertler et al. \(2017\)](#).

#### Ventajas de los diseños experimentales:

- 1) Constituyen el método más sólido de evaluar el impacto de un programa. La asignación aleatoria se conoce como la "regla de oro" o *gold standard* de la evaluación de impacto.
- 2) La interpretación de los resultados es extremadamente simple. El efecto causal de la intervención se puede medir como la diferencia entre los valores promedio del *outcome* en ambos grupos.
- 3) La propia construcción del experimento, que parte de un mecanismo de asignación aleatoria, implica que no se contemple la posible existencia de un *sesgo de selección*.

#### Limitaciones de los diseños experimentales:

- 1) Problemas éticos y morales. Se puede considerar injusto privar de manera aleatoria a ciertas personas (las que forman parte del grupo de control) de los potenciales efectos beneficiosos asociados a una nueva política.
- 2) Los diseños experimentales suelen ser excesivamente caros y consumen mucho tiempo. Además, requieren trabajar con muestras de grandes dimensiones.
- 3) Se antoja complicado sostener que la asignación de los individuos a los grupos

de tratamiento y control sea íntegramente fruto del azar.

El alcance de las dos primeras limitaciones permite comprender y justificar por qué el método de diseño experimental no se ha utilizado en la práctica con la frecuencia deseada. En su lugar, los investigadores optan por métodos y técnicas de corte *cuasiexperimental*.

## 4.2. Diseños *cuasiexperimentales*

Tal y como sucede con los RCT, estos diseños permiten definir relaciones causa-efecto y contrastar hipótesis causales. Sin embargo, carecen, por definición, de distribución aleatoria. En su lugar, se recurre a potentes herramientas estadísticas y econométricas.

### Ventajas de los diseños *cuasiexperimentales*:

- 1) Se pueden llevar a cabo a partir de bases de datos ya existentes.
- 2) A diferencia de los RCT, proporcionan una aproximación a la estimación del impacto mucho más rápida y barata.
- 3) Cuando la asignación aleatoria no es posible, siempre facilitan la obtención de un grupo de comparación válido.

### Limitaciones de los diseños *cuasiexperimentales*:

- 1) En contraste con los diseños experimentales, la fiabilidad de los resultados no es tan elevada. Ninguna técnica o metodología estadística de corte *cuasiexperimental* es tan robusta como un RCT.
- 2) La implementación de algunos métodos *cuasiexperimentales* puede ser bastante compleja. Se recomienda acreditar un perfil claramente orientado hacia el análisis cuantitativo de datos.
- 3) El problema del *sesgo de selección* siempre está presente cuando manejamos métodos *cuasiexperimentales*. La ecuación básica

de la evaluación de impacto, mostrada en la expresión [2], habría de completarse ahora con un término adicional:

$$\tau = E(Y|P=1) - E(Y|P=0) + \text{sesgo de selección} \quad [3]$$

El *sesgo de selección* no puede eliminarse si optamos por un diseño *cuasiexperimental*. El objetivo perseguido en estos casos es el de su minimización.

No existe ninguna técnica *cuasiexperimental* superior o mejor que otra. Tampoco existe la técnica perfecta. Cada una de ellas se adecúa a un tipo diferente de programa, así como a las circunstancias y al contexto en que se vaya a desarrollar la evaluación ([Rossi et al., 2019](#)).

## 5. Métodos de estimación

La imposibilidad de llevar a cabo RCT en numerosos escenarios, obliga a los investigadores a aplicar diseños “*cuasiexperimentales*”. Las principales técnicas categorizadas como tales son las siguientes: diferencias en diferencias, controles sintéticos, métodos de emparejamiento y regresión en discontinuidad.

### 5.1. Diferencias-en-diferencias (diff-in-diff)

#### 5.1.1. Descripción de la técnica

El método de diferencias en diferencias, o dobles diferencias, compara los cambios observados en la variable de resultado considerando dos periodos de tiempo y dos grupos. Tan solo uno de ellos se beneficia de los efectos del programa a la conclusión del segundo periodo; ninguno lo hace durante el primero.

La estimación del impacto aquí se divide en dos etapas.

- Habitualmente<sup>1</sup>, en la primera fase se cal-

<sup>1</sup> En realidad, existen dos alternativas para calcular el impacto causal cuando se aplica un enfoque de diferencias en diferencias. Aquí se describe la “opción 1” (véase la **Figura 3**).

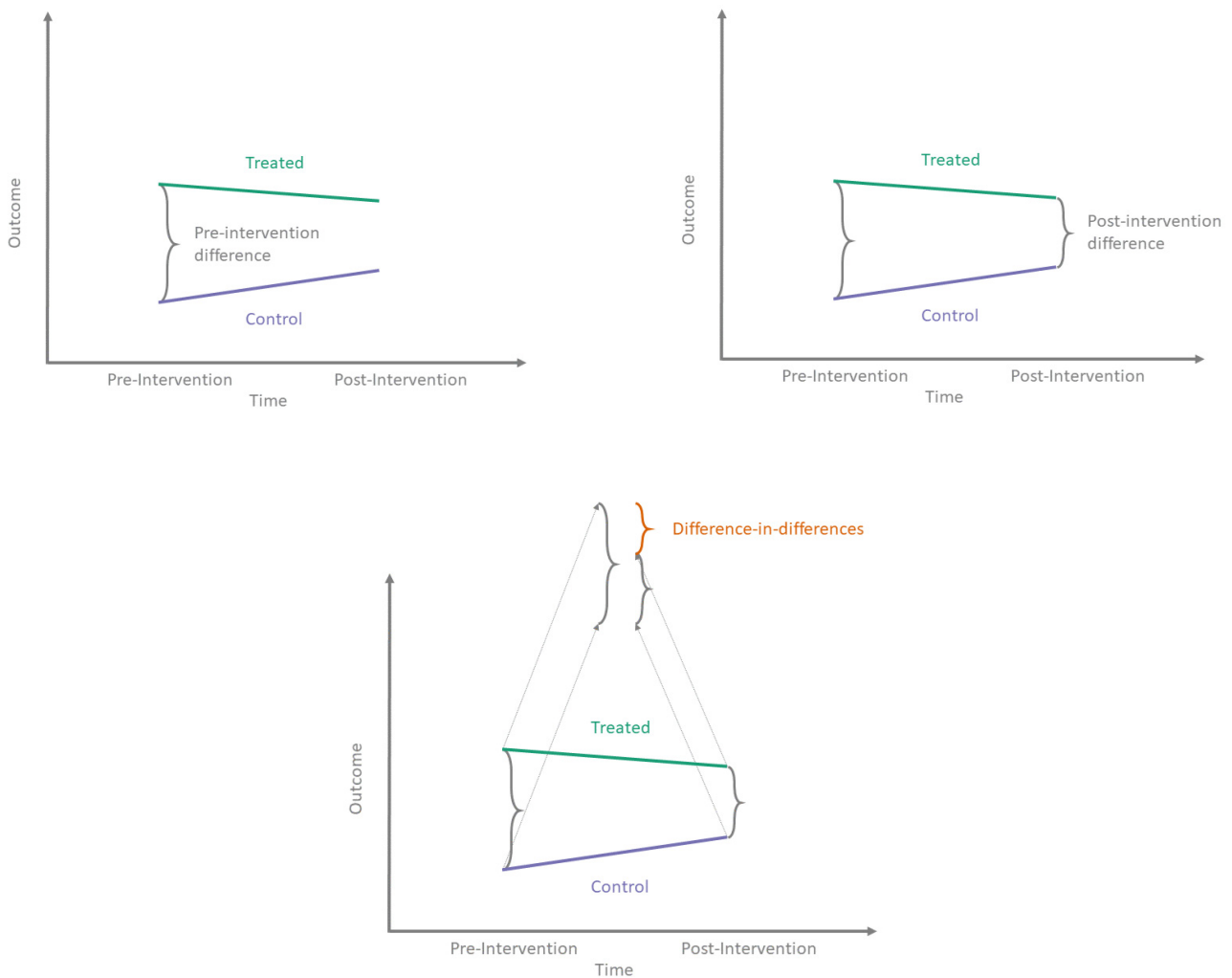
cula la diferencia de resultados (antes-después) en la variable objetivo y para cada uno de los grupos.

- Posteriormente, se obtiene la diferencia de las dos restas calculadas previamente (justificando así el nombre de la técnica). Operando de este modo logramos anular los

posibles sesgos atribuibles a diferencias permanentes entre los dos grupos, y también los procedentes de tendencias temporales.

Gráficamente, el funcionamiento de un diseño de dobles diferencias se puede esquematizar tal y como describe la **Figura 2**.

**Figura 2.** Modelo de diferencias-en-diferencias. Descripción gráfica



Notas: (a) La distancia resaltada en color naranja representa la estimación del impacto. (b) En este caso, la estimación del impacto se lleva a cabo recurriendo a la "opción 2" (véase la **Figura 3**).  
Fuente: <https://diff.healthpolicydatascience.org/>



La **Figura 3**, por su parte, resume el diseño de un modelo diff-in-diff desde una perspectiva analítica:

**Figura 3.** Modelo de diferencias-en-diferencias. Descripción analítica

|  | Control: $P = 0$ | Tratamiento: $P = 1$ | Dobles diferencias<br>(opción 2)             |
|--|------------------|----------------------|--|
| Periodo inicial: : $t=0$<br><i>línea de base</i> | $E(Y_0   P = 0)$ | $E(Y_0   P = 1)$     | $D_0$  |
| Periodo final: : $t=1$<br><i>seguimiento</i>     | $E(Y_1   P = 0)$ | $E(Y_1   P = 1)$     | $D_1$  |
| Dobles diferencias<br>(opción 1)                 | $D_c$            | $D_t$                | <i>Diff-in-diff</i><br>$D_T - D_C / D1 - D0$ |

Fuente: Elaboración propia.

Donde  $E(Y_0 | P = 0)$  representa el valor promedio del *outcome* en el momento inicial ( $Y_0$ ) condicionado (|) a que las unidades de análisis no participen en el programa ( $P = 0$ ); por su parte,  $E(Y_0 | P = 1)$  representa el valor promedio del *outcome* en el momento inicial ( $Y_0$ ), condicionado (|) a que las unidades de análisis sí participan en el programa ( $P = 1$ ). Y así, sucesivamente.

Es importante destacar asimismo que, a fin de garantizar la validez de las estimaciones, se debe verificar el cumplimiento del denominado "supuesto de tendencias paralelas". En otras palabras, es preciso comprobar que la tendencia de la variable de resultado en los grupos de tratamiento y control ha sido semejante en los periodos previos al inicio de la intervención. La violación de este supuesto conduce a una estimación sesgada del efecto causal.

### 5.1.2. Ejemplos de aplicación práctica

Las dos referencias clásicas de investigaciones que utilizan un modelo de diff-in-diff son los trabajos de [Card \(1990\)](#) y [Card y Krueger \(1994\)](#). Más recientemente, [Roth et al. \(2023\)](#) resumen los últimos avances econométricos relacionados con la metodología, y ofrecen diversas recomendaciones de interés para expertos en la materia.

Dentro del ámbito sanitario, cabe destacar la

contribución de [Wing et al. \(2018\)](#). En este artículo, los autores revisan los supuestos básicos de un diseño de dobles diferencias, poniendo en valor sus potenciales ventajas en la investigación de políticas de salud pública. Además, proponen diferentes tácticas y enfoques complementarios para mejorar y perfeccionar los ejercicios de inferencia estadística.

### 5.1.3. Ventajas y limitaciones del método

Ventajas del método de diferencias-en-diferencias:

- 1) Probablemente, se trata de la técnica cuasi-experimental más sencilla de implementar.
- 2) Al contrastar las variaciones previas y posteriores a la intervención entre un grupo que recibe tratamiento y otro que se utiliza como control, un diseño diff-in-diff garantiza la consecución de una sólida validez interna.
- 3) Ambos grupos pueden ser diferentes en características no observables (motivación, esfuerzo, etc.) siempre y cuando las diferencias permanezcan constantes a lo largo del tiempo.

Limitaciones del método de diferencias-en-diferencias:

- 1) No se puede garantizar el cumplimiento del "supuesto de igualdad de tendencias" en todos los escenarios. En especial, cuando se

trabaja en entornos manifiestamente dinámicos.

- 2) Para verificar el cumplimiento del “supuesto de caminos paralelos” es necesario recopilar datos de numerosos periodos previos al inicio de la intervención.
- 3) En este tipo de modelos, los problemas de endogeneidad asociados al tratamiento son bastante habituales.

## 5.2. Controles sintéticos (*Synthetic Control Methods*)

---

### 5.2.1. Descripción de la técnica

---

La metodología de controles sintéticos (SCM), se ha convertido en una herramienta ampliamente utilizada en los últimos años para estimar el impacto de políticas públicas de muy amplia y diversa naturaleza. La técnica resulta especialmente útil cuando se trata de estimar el efecto causal de intervenciones de política económica que se implementan a nivel agregado. Por otra parte, se trata de una técnica estadística muy transparente que proporciona un grupo de control “artificial” para la unidad tratada.

SCM atribuye un peso o ponderación a cada unidad de control (provincia, comunidad autónoma, país) a partir de herramientas estadísticas de optimización que minimizan la distancia entre vectores. Unos vectores que únicamente deben contener información sobre la variable de resultado (outcome) y un conjunto de predictores de aquella, tanto para la unidad beneficiaria del tratamiento como para los potenciales controles (donor pool)<sup>2</sup>. Este método sugiere que, cuando se trabaja con datos macro, una combinación ponderada de unidades sin tratar (control sintético) a menudo proporciona una aproximación más precisa y fiel a las características de la unidad de tratamiento<sup>3</sup> que una única entidad individual.

---

<sup>2</sup> Denominación que recibe el grupo de potenciales unidades de comparación en la metodología SCM.

<sup>3</sup> Durante el periodo previo al inicio de la intervención.

Desde un punto de vista gráfico, su interpretación es sencilla. La idea que subyace a la lectura de los principales resultados (“eyeball test”) es la de comprobar en qué medida las series representadas (evolución real del outcome y evolución simulada) se aproximan entre sí durante el pre-tratamiento. Cuanto más próximas se encuentren<sup>4</sup>, más fiables serán las estimaciones obtenidas. La determinación del impacto para cada uno de los años post-tratamiento se calcula mediante la simple distancia vertical entre ambas series.

### 5.2.2. Ejemplos de aplicación práctica

---

La técnica SCM surge por primera vez con la sugerente propuesta de [Abadie y Gardeazabal \(2003\)](#), pero es tras las contribuciones de [Abadie et al. \(2010\)](#) y [Abadie et al. \(2015\)](#) cuando la aplicación de la metodología experimenta una notable expansión.

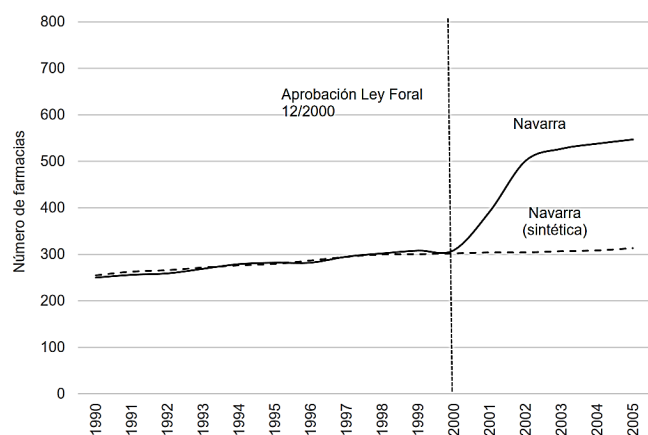
Desde entonces, su difusión no ha dejado de crecer, extendiéndose ya a intervenciones en numerosos ámbitos (por ejemplo: políticas y programas de salud pública). En relación con estas últimas, un ejemplo muy ilustrativo se halla en [Silos Ribas y Vega Vicente \(2015\)](#). La investigación examina los efectos de la [Ley Foral 12/2000, de 16 de noviembre](#) sobre el número de farmacias en Navarra. Esta ley introdujo una importante liberalización en la apertura de nuevas farmacias en Navarra en contraste con las regulaciones y legislaciones vigentes en otras comunidades autónomas.

---

<sup>4</sup> El escenario idóneo es de la superposición de las series.

La **Figura 4** presenta el “eyeball test” mostrado por los autores del trabajo arriba mencionado.

**Figura 4.** Evolución del número de farmacias en Navarra real y Navarra Sintética: 1990-2005 (“eyeball test”)



Fuente: [Silos Ribas y Vega Vicente \(2015\)](#).

El gráfico pone de manifiesto que el ajuste entre la unidad tratada (Navarra) y la unidad de control (Navarra sintética) es muy bueno. Ambas series prácticamente se superponen durante todo el periodo previo al inicio de la intervención. En consecuencia, se podría confiar plenamente en las estimaciones realizadas desde el año 2000 en adelante.

En palabras de [Athey e Imbens \(2017\)](#) “*el enfoque de controles sintéticos (...) es posiblemente la innovación más importante en la literatura de evaluación de políticas durante los últimos 15 años.*”

### 5.2.3. Ventajas y limitaciones del método

#### Ventajas del método de controles sintéticos:

- 1) Es mucho más sencillo y operativo trabajar con datos macro que con microdatos (muchas intervenciones públicas se pueden evaluar a nivel agregado).
- 2) La metodología de controles sintéticos no requiere asumir supuestos tan estrictos como sí ocurre con otras técnicas.
- 3) Es una técnica transparente. Los resulta-

dos obtenidos informan de la importancia relativa de cada “donante” en la unidad sintética.

#### Limitaciones del método de controles sintéticos:

- 1) Los ejercicios de inferencia estadística son menos formales que los implementados en el caso de otras técnicas cuantitativas.
- 2) Es recomendable disponer de un gran número de observaciones en el periodo pre-tratamiento. Es precisamente ahí dónde se lleva a cabo el ejercicio de optimización.
- 3) Resulta complicado encontrar unidades de control con características similares a la de tratamiento que no se hayan visto afectadas por el programa analizado u otro similar. Si esto sucede, no se pueden utilizar.

## 5.3. Matching o emparejamiento

### 5.3.1. Descripción de las técnicas

En el caso del matching, se busca un emparejamiento exacto para cada unidad sometida a tratamiento sobre la base de una serie de características observables fácilmente identificables (por ejemplo: el nivel de estudios, la edad o la ocupación). El problema surge cuando es necesario manejar un amplio número de regresores o características, ya que resulta muy complicado definir y aplicar el concepto de proximidad. [Rosenbaum y Rubin \(1983\)](#) solucionaron este problema proponiendo el cálculo de una medida única: la denominada puntuación (puntaje) de propensión o propensity score (PS). Este indicador sintetiza toda la información relevante del vector de características observadas en un único valor.

#### a) *Emparejamiento exacto*

Las técnicas de matching tratan de emparejar unidades tratadas y no tratadas en función de una serie de características observables. Y parten de la aceptación del supuesto de “selección en

observables”, también conocido como “supuesto de independencia condicional”, el cual implica que las características no observables no desempeñan

un papel relevante. La **Figura 5** muestra cómo proceder ante un ejercicio de emparejamiento exacto.

**Figura 5.** *Emparejamiento exacto en cuatro características*

| Unidades tratadas |        |                   |                       | Unidades no tratadas |        |                   |                       |
|-------------------|--------|-------------------|-----------------------|----------------------|--------|-------------------|-----------------------|
| Edad              | Género | Meses desempleado | Diploma de secundaria | Edad                 | Género | Meses desempleado | Diploma de secundaria |
| 19                | 1      | 3                 | 0                     | 24                   | 1      | 8                 | 1                     |
| 35                | 1      | 12                | 1                     | 38                   | 0      | 1                 | 0                     |
| 41                | 0      | 17                | 1                     | 58                   | 1      | 7                 | 1                     |
| 23                | 1      | 6                 | 0                     | 21                   | 0      | 2                 | 1                     |
| 55                | 0      | 21                | 1                     | 34                   | 1      | 20                | 0                     |
| 27                | 0      | 4                 | 1                     | 41                   | 0      | 17                | 1                     |
| 24                | 1      | 8                 | 1                     | 46                   | 0      | 9                 | 0                     |
| 46                | 0      | 3                 | 0                     | 41                   | 0      | 11                | 1                     |
| 33                | 0      | 12                | 1                     | 19                   | 1      | 3                 | 0                     |
| 40                | 1      | 2                 | 0                     | 27                   | 0      | 4                 | 0                     |

Fuente: Gertler et al. (2017).

Este tipo de emparejamiento solamente se puede llevar a cabo en situaciones muy concretas. De hecho, resulta muy complicado lograrlo cuando: (1) Alguna de las variables a incluir en el análisis es continua (por ejemplo: el nivel de colesterol en sangre, la edad o el peso); (2) La muestra de datos a disposición del investigador tiene un tamaño reducido; (3) El vector de características es excesivamente grande: problema de la “maldición de la dimensionalidad”.

Como solución a los problemas del emparejamiento exacto, se desarrolló la técnica del PSM, una metodología más sofisticada que, sin embargo, permite solventar las principales limitaciones de aquel.

*b) Propensity Score Matching (PSM)*

Uno de los aspectos críticos en la aplicación de técnicas de matching consiste en definir con exactitud lo que significa “similar”. Cuando únicamente manejamos unas pocas características, el concepto “similar” está claro<sup>5</sup>. No obstante, conforme vamos acumulando

variables, la idea de semejanza se difumina. Surge entonces el denominado problema de la “maldición de la dimensionalidad”. La solución que propone el PSM en relación con este último consiste en utilizar como instrumento de emparejamiento, no las características observadas, sino la probabilidad de participar en el programa en función de esas características. La idea es transformar un problema multidimensional en uno unidimensional.

El PS se define como la probabilidad estimada de recibir tratamiento, de participar en el programa (*P*), condicionada a un conjunto de variables observables (*X*). Esta puntuación, que es un escalar que adopta valores entre 0 y 1, resume la influencia de todas las características observables en la probabilidad de participar en el programa. Calculados los PS<sup>6</sup>, las unidades pertenecientes al grupo de tratamiento se emparejan con las unidades de control que presentan puntuaciones próximas<sup>7</sup>.

<sup>6</sup> El procedimiento empleado para calcular estas puntuaciones consiste en estimar un modelo de regresión de elección discreta o respuesta dicotómica: un modelo logit o un modelo probit.

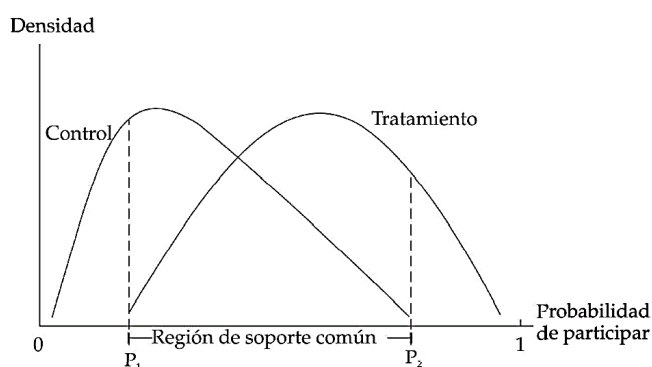
<sup>7</sup> En la práctica existen diversos algoritmos de emparejamiento. Entre otros, cabe destacar los siguientes: el estimador PSM por vecino más cercano, el de radio o el basado en funciones kernel.

<sup>5</sup> Véase la **Figura 5**.

Por otro lado, es preciso señalar que el PSM únicamente puede implementarse en la denominada región de “soporte común”. Esta condición implica que en la estimación tan solo se utilizan unidades del grupo de control que presentan probabilidades de participación parecidas a las del grupo de tratamiento. El procedimiento más intuitivo de demostrar el cumplimiento de la hipótesis de soporte común consiste en llevar a cabo una exploración visual.

Se trata, en definitiva, de comprobar en qué grado se superponen, para los grupos de tratamiento y control, las respectivas distribuciones de densidad de la probabilidad estimada de participación (véase la **Figura 6**).

**Figura 6.** Estudio del soporte común en un ejercicio de PSM



Fuente: Bernal y Peña (2011).

Una vez restringida la muestra al soporte común, y tras realizar la prueba de equilibrado de muestras (balancing test o balancing property) para verificar que las características observables son similares entre los grupos de tratamiento y control, se procedería a calcular el impacto o efecto causal del programa.

### 5.3.2. Ejemplo de aplicación práctica

Imagine que se desea estudiar el efecto de un programa que tiene por objetivo mejorar la nutrición de los menores de 1 año pertenecientes a familias con bajos ingresos. Para ello, se recurre a una submuestra de 7 familias. La **Figura 7**

incorpora las probabilidades estimadas de participación (PS) y la estatura de los destinatarios del programa (expresada en centímetros).

**Figura 7.** Ejemplo de Propensity Score Matching

| Grupo \ Familia      |  | 1    | 2    | 3    | 4    | 5    | 6    | 7    |
|----------------------|--|------|------|------|------|------|------|------|
| Grupo de tratamiento | PS: Probabilidad de participación estimada | 0,40 | 0,46 | 0,54 | 0,67 | 0,73 | 0,75 | 0,81 |
|                      | Y: Estatura de los menores de 1 año        | 66   | 75   | 72   | 75   | 79   | 68   | 81   |
| Grupo de control     | PS: Probabilidad de participación estimada | 0,42 | 0,49 | 0,57 | 0,58 | 0,61 | 0,67 | 0,81 |
|                      | Y: Estatura de los menores de 1 año        | 72   | 68   | 69   | 71   | 79   | 68   | 72   |

Fuente: Elaboración propia a partir de Bernal y Peña (2011).

En este caso, se ha escogido como método de emparejamiento el algoritmo del vecino más cercano. La familia tratada #1, cuya probabilidad de participación es 0,40, se emparejaría con la de control que tiene una probabilidad igual a 0,42 (resaltado en tono azul). Su contribución al impacto sería: 66 – 72 = -6 cm; La familia tratada #2, cuya probabilidad es 0,46, se emparejaría con el control de probabilidad igual a 0,49 (resaltado en color rojo); la contribución al efecto causal en este caso vendría dada por la siguiente diferencia: 75 – 68 = 7 cm. Y así, sucesivamente. Adicionalmente, se ha resaltado en color verde cuál sería el emparejamiento de la familia tratada #5 (la familia de control #6).

Luego, para el ejemplo propuesto, el impacto total estimado resultaría ser el promedio simple de las contribuciones individuales al efecto causal:

$$\tau_{PSM} = \frac{1}{7} \times (-6 + 7 + 3 + 7 + 11 - 4 + 9) = 3,85 \text{ cm}$$

Los menores de un año que recibieron el tratamiento resultaron ser, en promedio, casi 4 cm más altos frente a aquellos que no fueron beneficiarios de la política.

### 5.3.3. Ventajas y limitaciones del método

Ventajas de las técnicas de matching o emparejamiento:

- 1) En comparación con los RCT, las técnicas de *matching* sortean los reproches éticos y morales que surgen cuando, por meras razones de azar, se impide que ciertas personas accedan a un programa potencialmente beneficioso.
- 2) Las técnicas de emparejamiento no presentan un coste económico tan elevado como los RCT, que son muy caros.
- 3) Las técnicas de *matching* presentan una importante ventaja frente a los tradicionales métodos de regresión: los estimadores de emparejamiento toman en consideración el problema del "soporte común". Si no se produce una superposición clara entre tratados y controles, la robustez de los resultados queda en entredicho.

Limitaciones de las técnicas de matching o emparejamiento:

- 1) Resulta muy complicado lograr un emparejamiento exacto en casi todos los escenarios.
- 2) En comparación con un RCT, estas técnicas parten de un "supuesto de independencia condicional" que permite eliminar el *sesgo de selección* "únicamente" mediante el control en variables observables. Aquéllos garantizan que personas beneficiarias y controles sean similares tanto en características observables como en no observables.
- 3) Frente a un RCT, las técnicas de emparejamiento tan solo pueden estimar los efectos del tratamiento donde existe una superposición entre beneficiarios de la intervención y unidades de control ("soporte común").

### 5.4. Regresión en discontinuidad

#### 5.4.1. Descripción de la técnica

Este método resulta apropiado en aquellas situaciones en que las potenciales personas beneficiarias de una política pública pueden ordenarse a partir de un índice continuo de elegibilidad. Por ejemplo, el nivel de renta, la edad de los individuos o las calificaciones obtenidas durante un determinado curso escolar. En esta metodología, asimismo, la simple fijación de un umbral o punto de corte permite separar al grupo de tratamiento del de control. En los alrededores de ese punto de corte, los individuos que se sitúen justo por encima se considerarán similares a los ubicados justo por debajo<sup>8</sup>. De hecho, tan solo existirá un factor o característica que los diferenciará: la participación en el programa.

En cuanto al ámbito de aplicación de la metodología, cabe subrayar que resulta muy común encontrarse con programas sanitarios y políticas sociales que requieren de un índice de elegibilidad para definir las condiciones de acceso. Es el caso, por ejemplo, de campañas de vacunación o de programas de lucha contra la pobreza.

A fin de llevar a la práctica correctamente este método, es necesario cubrir las cuatro etapas que se enumeran justo debajo:

- 1) Disponer de una "variable de asignación" ( $Z$ ) que tome valores continuos.
- 2) Fijar o reconocer un umbral o punto de corte ( $Z^*$ ) que permita decretar qué unidades (personas o familias) recibirán el programa y cuáles no.

<sup>8</sup> La regla externa adoptada para fijar el punto de corte determinará a qué lado del umbral se sitúa el grupo de tratamiento y dónde el de control. Ambos pueden terminar emplazados tanto a la derecha como a la izquierda del punto de corte establecido. El resultado dependerá, por un lado, de la variable de resultado escogida ( $Y$ ), que representamos en el eje de ordenadas (por ejemplo: el nivel de gasto o de consumo); por el otro, de la variable de asignación tomada como referencia ( $Z$ ), que trazamos sobre el eje de abscisas (por ejemplo: la edad o el nivel de ingresos).

- 3) Implementar el programa sobre aquellas unidades que se consideran población objetivo del mismo ( $P = 1$ ).
- 4) Determinación del efecto causal. Para ello, se comparan los resultados del outcome en aquellas unidades que se sitúan justo por debajo y justo por encima del punto de corte:  $Z^*$ .

Finalmente, conviene aclarar que existen dos diseños de regresión en discontinuidad: el diseño sharp (brusco o nítido) y el diseño fuzzy (difuso o borroso):

- a) Diseño "sharp" (brusco o nítido):
  - Se aplica cuando la probabilidad de participar en el tratamiento es determinista, esto es, cuando cambia de 0 a 1 en un punto concreto ( $Z^*$ ) de la variable  $Z$ .
  - La estimación se basa en comparar qué sucede con la variable de resultado ( $Y$ ) en los puntos situados justo por debajo y justo por encima de  $Z^*$ .
- b) Diseño "fuzzy" (difuso o borroso):
  - Se aplica cuando la probabilidad de participar en el programa es nuevamente discontinua en  $Z^*$ , pero el cambio ahora no es de 0 a 1. En este segundo caso, la discontinuidad es estocástica.
  - Existen otras variables, además de  $Z$ , que también afectan a la probabilidad de ser tratado. Aparecen tratados y no tratados a ambos lados de  $Z^*$ . El proceso de estimación no es tan inmediato ni directo.

#### 5.4.2. Ejemplos de aplicación práctica

Dos de los principales trabajos que dieron a conocer la metodología fueron las contribuciones

de [Angrist y Lavy \(1999\)](#) y la de [Black \(1999\)](#). La repercusión de ambos artículos propició un crecimiento exponencial en el uso y aplicación de la técnica que aún hoy perdura.

Por otro lado, un ilustrativo y sencillo ejemplo del funcionamiento del método, basado en Carpenter y Dobkin ([2009](#), [2011](#)), es el ejercicio que se presenta a continuación. El ejercicio analizado tiene por objetivo determinar los efectos de alcanzar la edad mínima legal para el consumo de alcohol<sup>9</sup> ( $P$ ) sobre las tasas de mortalidad ( $Y$ ). A partir de los datos mostrados en la **Figura 8**, las **Figuras 9a** y **9b** representan los efectos mencionados haciendo uso de dos herramientas informáticas alternativas: Microsoft Excel y Stata.

A continuación, aplicamos las cuatro etapas descritas en la subsección 5.4.1 para comprender, de manera intuitiva, cómo se implementa esta técnica:

- 1) Disponer de una "variable de asignación" ( $Z$ ) que tome valores continuos: la edad.
- 2) Fijar o reconocer un umbral o punto de corte ( $Z^*$ ) que permita decretar qué personas cumplen la condición (tratados) y cuáles no: 21 años.
  - Personas con un valor  $Z < 21$  no se consideran tratados (controles).
  - Personas con un valor  $Z \geq 21$  sí se consideran tratados.
- 3) Verificar el cumplimiento del requisito legal en aquellas personas que se consideran población objetivo ( $P = 1$ ): personas con un valor  $Z \geq 21$ .
- 4) Analizar qué ha sucedido con la variable de resultado ( $Y$ ), comparando los resultados de aquellas unidades que se sitúan justo por debajo y justo por encima del punto de corte:  $Z^*$ .

$$\tau_{RDD}^{sharp} = E(Y|21^+) - E(Y|21^-) \quad [4]$$

De las 50 personas incluidas en la muestra (**Figura 8**),  
9 En Estados Unidos se sitúa en 21 años.

3 de ellas se encuentran ligeramente por debajo del umbral de 21 años (identificadores: 23, 24 y 25) y tres ligeramente por encima (identificadores: 26, 27 y 28).

El efecto causal de la intervención (cumplir con la condición:  $Z \geq 21$ ) se podría estimar entonces a través de la ecuación básica de la evaluación de impacto para diseños "sharp" (expresión [4]):

$$\tau_{RDD}^{sharp} = \frac{1}{3} \times (105,5 + 101 + 97,5) - \frac{1}{3} \times (94,3 + 94,5 + 94,7)$$

$$\tau = 101,3 - 94,5 = 6,8 \text{ muertes por } 100.000 \text{ habitantes}$$



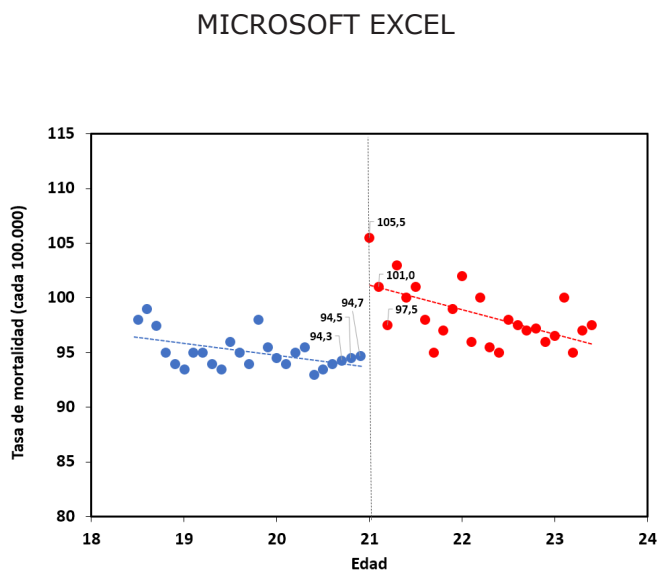
Figura 8. Datos ficticios de 50 personas utilizados en el ejercicio

| <b>Identificador</b> | <b>Control = 0 / Tratado = 1</b> | <b>Edad</b> | <b>Tasa de mortalidad</b> |
|----------------------|----------------------------------|-------------|---------------------------|
| 1                    | 0                                | 18,5        | 98,0                      |
| 2                    | 0                                | 18,6        | 99,0                      |
| 3                    | 0                                | 18,7        | 97,5                      |
| 4                    | 0                                | 18,8        | 95,0                      |
| 5                    | 0                                | 18,9        | 94,0                      |
| 6                    | 0                                | 19,0        | 93,5                      |
| 7                    | 0                                | 19,1        | 95,0                      |
| 8                    | 0                                | 19,2        | 95,0                      |
| 9                    | 0                                | 19,3        | 94,0                      |
| 10                   | 0                                | 19,4        | 93,5                      |
| 11                   | 0                                | 19,5        | 96,0                      |
| 12                   | 0                                | 19,6        | 95,0                      |
| 13                   | 0                                | 19,7        | 94,0                      |
| 14                   | 0                                | 19,8        | 98,0                      |
| 15                   | 0                                | 19,9        | 95,5                      |
| 16                   | 0                                | 20,0        | 94,5                      |
| 17                   | 0                                | 20,1        | 94,0                      |
| 18                   | 0                                | 20,2        | 95,0                      |
| 19                   | 0                                | 20,3        | 95,5                      |
| 20                   | 0                                | 20,4        | 93,0                      |
| 21                   | 0                                | 20,5        | 93,5                      |
| 22                   | 0                                | 20,6        | 94,0                      |
| 23                   | 0                                | 20,7        | 94,3                      |
| 24                   | 0                                | 20,8        | 94,5                      |
| 25                   | 0                                | 20,9        | 94,7                      |
| 26                   | 1                                | 21,0        | 105,5                     |
| 27                   | 1                                | 21,1        | 101,0                     |
| 28                   | 1                                | 21,2        | 97,5                      |
| 29                   | 1                                | 21,3        | 103,0                     |
| 30                   | 1                                | 21,4        | 100,0                     |
| 31                   | 1                                | 21,5        | 101,0                     |
| 32                   | 1                                | 21,6        | 98,0                      |
| 33                   | 1                                | 21,7        | 95,0                      |
| 34                   | 1                                | 21,8        | 97,0                      |
| 35                   | 1                                | 21,9        | 99,0                      |
| 36                   | 1                                | 22,0        | 102,0                     |
| 37                   | 1                                | 22,1        | 96,0                      |
| 38                   | 1                                | 22,2        | 100,0                     |
| 39                   | 1                                | 22,3        | 95,5                      |
| 40                   | 1                                | 22,4        | 95,0                      |
| 41                   | 1                                | 22,5        | 98,0                      |
| 42                   | 1                                | 22,6        | 97,5                      |
| 43                   | 1                                | 22,7        | 97,0                      |
| 44                   | 1                                | 22,8        | 97,2                      |
| 45                   | 1                                | 22,9        | 96,0                      |
| 46                   | 1                                | 23,0        | 96,5                      |
| 47                   | 1                                | 23,1        | 100,0                     |
| 48                   | 1                                | 23,2        | 95,0                      |
| 49                   | 1                                | 23,3        | 97,0                      |
| 50                   | 1                                | 23,4        | 97,5                      |

Nota: (a) La variable edad está expresada en años. (b) La tasa de mortalidad recoge el número de muertes por cada 100.000 personas; se divide el número total de muertes a una edad determinada por el número total de años vividos a esa edad por todos los miembros de la cohorte.

Fuente: Elaboración propia.

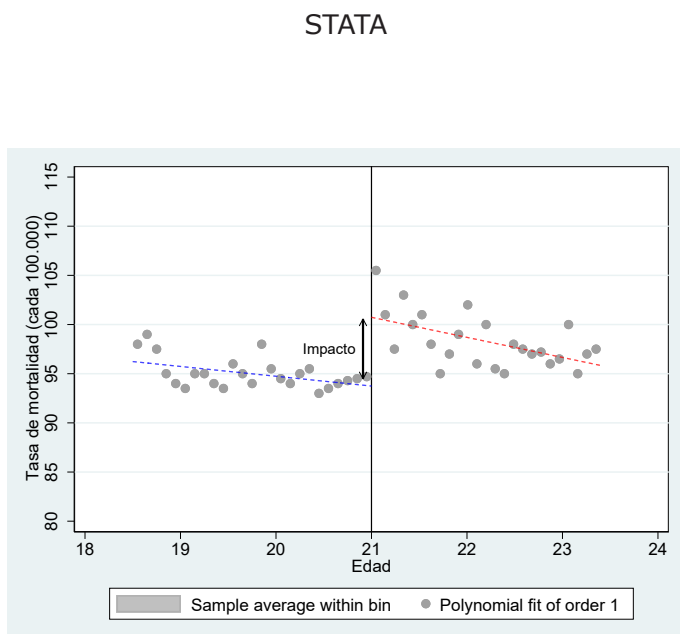
**Figura 9a.** Estimación de los efectos de alcanzar la edad mínima legal para el consumo de alcohol sobre la mortalidad



*Nota:* Edad mínima legal para el consumo de alcohol en Estados Unidos: 21 años.

*Fuente:* Elaboración propia.

**Figura 9b.** Estimación de los efectos de alcanzar la edad mínima legal para el consumo de alcohol sobre la mortalidad



*Nota:* Edad mínima legal para el consumo de alcohol en Estados Unidos: 21 años.

*Fuente:* Elaboración propia.

### 5.4.3. Ventajas y limitaciones del método

#### Ventajas de la regresión en discontinuidad:

- 1) Permite identificar los efectos causales del programa sin imponer restricciones arbitrarias de exclusión o relativas al proceso de selección.
- 2) El proceso de clasificación de unidades en tratados y controles parte de una regla externa conocida, criterio muy habitual en el diseño de políticas sociales.
- 3) Los análisis de regresión en discontinuidad permiten obtener estimadores insesgados del impacto de la política, lo cual es deseable.

#### Limitaciones de la regresión en discontinuidad:

- 1) Ofrece una estimación local. Los efectos del tratamiento alrededor del punto de corte no siempre son extrapolables.
- 2) Es preciso disponer de numerosas observaciones en torno al umbral fijado por la regla de acceso al programa.
- 3) Los individuos no puedan escoger en qué lado del punto de corte se encuentran. La estimación quedaría invalidada si existe manipulación.

## 6. Resumen y conclusiones

En la actualidad, la utilización de métodos y técnicas cuantitativas para medir el impacto de programas y políticas públicas ha cobrado una gran relevancia. Las crecientes demandas políticas en búsqueda de una mayor transparencia, y de unos mecanismos más adecuados para alcanzar el objetivo de rendición de cuentas, así lo ponen de manifiesto.

A fin de avalar el rigor metodológico, toda evaluación de impacto debe estimar un

contrafactual que recree de manera fiel y precisa lo que habría ocurrido si la política pública no se hubiese llevado a cabo. Se trata, en definitiva, de obtener un grupo de comparación que sea lo más parecido posible al de tratamiento en todas las características excepto en una: la participación en el programa.

Para lograr este objetivo, la literatura sobre evaluación de impacto pone a nuestro alcance dos instrumentos de potente alcance y repercusión: los RCT y los diseños cuasiexperimentales. Comenzando con los primeros, cabe destacar que, muy recientemente, en 2019, Abhijit Banerjee, Esther Duflo y Michael Kremer fueron galardonados con el Premio Nobel de Economía por sus aportaciones al estudio de la pobreza y su erradicación desde un enfoque de experimentos sociales. Sus contribuciones nos ayudan a enriquecer nuestro conocimiento sobre la economía conductual (behavioral economics), tratando de aprovechar dicha experiencia para el mejor diseño de políticas fundamentadas en esquemas de incentivos.

A pesar de que la asignación aleatoria al tratamiento, la que se lleva a cabo en los mencionados diseños experimentales, constituye una herramienta extraordinariamente útil por su robustez, no siempre resulta factible su implementación. Junto con los elevados costes económicos asociados a su proceso de puesta en marcha, en numerosas ocasiones van igualmente acompañados de ciertas críticas relacionadas con cuestiones morales o éticas (derivadas de los mecanismos de asignación aleatoria aplicados).

Cuando por alguna de las razones arriba expuestas no podemos recurrir a los RCT, la alternativa óptima, los investigadores acuden a los métodos cuasiexperimentales. El potencial de estas técnicas econométricas, en ocasiones complejas, pero también transparentes, no ha pasado desapercibido en los últimos años. Así, el Premio Nobel de Economía en 2021 fue a parar a manos de David Card, Joshua Angrist y Guido Imbens, quienes obtuvieron reconocimiento por

sus análisis y estudios sobre la trascendencia del análisis de causalidad.

Dentro de esta categoría de diseños cuasiexperimentales, se ha presentado el enfoque diff-in-diff, el método de controles sintéticos, las técnicas de emparejamiento y, finalmente, la metodología de regresión en discontinuidad. Todos estos instrumentos no solo resultan apropiados para identificar qué programas funcionan y cuáles no. Además, mejoran las capacidades de las instituciones para explotar herramientas cuantitativas al servicio de la gestión pública y los policy makers, y permiten cuantificar el impacto de los programas, políticas e intervenciones con un inestimable rigor metodológico.

## Bibliografía:

1. Abadie, Alberto, Alexis Diamond, y Jens Hainmueller. "Comparative politics and the synthetic control method." *American Journal of Political Science* 59.2 (2015): 495-510.
2. Abadie, Alberto, Alexis Diamond, y Jens Hainmueller. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105.490 (2010): 493-505.
3. Abadie, Alberto, y Javier Gardeazabal. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review* 93.1 (2003): 113-132.
4. Angrist, Joshua D., y Victor Lavy. "Using Maimonides' rule to estimate the effect of class size on scholastic achievement." *The Quarterly Journal of Economics* 114.2 (1999): 533-575.
5. Athey, Susan, y Guido W. Imbens. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31.2 (2017): 3-32.
6. Bernal, Raquel, y Ximena Peña. *Guía práctica para la evaluación de impacto*. Ediciones Uniandes-Universidad de los Andes, 2011.
7. Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education." *The Quarterly Journal of Economics* 114.2 (1999): 577-599.
8. Card, David, y Alan B. Krueger. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *American Economic Review* 84.4 (1994): 772-793.
9. Card, David. "The Impact of the Mariel Boatlift on the Miami Labor Market." *ILR Review* 43.2 (1990): 245-257.
10. Carpenter, Christopher, y Carlos Dobkin. "The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age." *American Economic Journal: Applied Economics* 1.1 (2009): 164-82.
11. Carpenter, Christopher, y Carlos Dobkin. "The Minimum Legal Drinking Age and Public Health." *Journal of Economic Perspectives* 25.2 (2011): 133-56.
12. Gertler, Paul J., Sebastián Martínez, Patrick Premand, Laura B. Rawlings, y Christel M. J. Vermeersch. *La evaluación de impacto en la práctica*. World Bank Publications, 2017.
13. Rosenbaum, Paul R., y Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70.1 (1983): 41-55.
14. Rossi, Peter H., Mark W. Lipsey, y Gary T. Henry. *Evaluation: A Systematic Approach*. Sage publications, 2019.
15. Roth, Jonathan, et al. "What's trending in difference-in-differences? A synthesis of the recent econometrics literature." *Journal of Econometrics* 235.2 (2023): 2218-2244.
16. Silos Ribas, Mateo, y Vega Vicente, Pilar. "Efectos de la eliminación de las restricciones de entrada geográficas en el sector de oficinas de farmacia." *Documento de Trabajo en Política de Competencia y Regulación N.º 001/2015*. CNMC, 2015.
17. Wing, Coady, Kosali Simon, and Ricardo A. Bello-Gomez. "Designing difference in difference studies: best practices for public health policy research." *Annual review of public health* 39 (2018): 453-469.