

# *Las potencialidades (y riesgos) de la evaluación de impacto en el aumento de la calidad de las finanzas públicas*

*The potential (and risks) of impact evaluation in increasing the quality of public finances*

Si se desean políticas públicas que logren solucionar los problemas de la sociedad es necesario, por un lado, generar evidencia para una mejor toma de decisiones, y, por otra parte, que estos análisis estén orientados a la obtención de resultados, y la evaluación de impacto (EI) cubre ambos objetivos. En este trabajo se presenta la necesidad e importancia de la EI en la mejora de las intervenciones públicas. Se muestran sus características, como el contrafactual y sesgo de selección, y se describen las principales metodologías de evaluación existentes agrupadas en diseños experimentales y cuasiexperimentales. En cada una de estos métodos se presta especial atención a los supuestos necesarios para su aplicación, necesidades de información y limitaciones y riesgos en su aplicación.

*Gizartearen arazoak konpontzeko moduko politika publikoak nahi baditugu, beharrezko da, alde batetik, ebidentzia sortzea, erabakiak hobeto hartzeko; eta, bestetik, analisiak emaitzak lortzera bideratzea. Eraginaren ebaluazioak (EE) bi helburuak betetzen ditu. Lan honen bidez nabarmendu nahi da EEa oso garrantzitsua dela esku-hartze publikoak hobetzeko, eta, hortaz, beharrezko dela. Ebaluazio mota horren ezaugarriak erakusten dira, hala nola kontrafaktuala dela eta hautaketa-alborapena egiten duela, eta gaur egungo ebaluazio-metodologia nagusiak ere deskribatzen dira, diseinu experimental eta ia experimental multzo-katuta. Metodo horietako bakoitzean, arreta berezia jartzen da aplikaziorako beharrezkoak diren suposizioetan, informazio-beharretan, mugetan eta aplikazio-arriskuetan.*

In order to have public policies that manage to solve society's problems, it is necessary, on the one hand, to generate evidence for better decision making, and on the other hand, that these analyses are results-oriented, and impact evaluation (IE) covers both objectives. This paper presents the need and importance of IE in the improvement of public interventions. Its main characteristics, such as counterfactual and selection bias, are shown, and the main existing evaluation methodologies grouped into experimental and quasi-experimental designs are described. Each of these methods is developed with special attention to the assumptions necessary for their application, information needs and limitations and risks in their application.

## Índice

---

1. Introducción
2. La evaluación de impacto
3. Diseño de la evaluación de impacto: experimental y cuasiexperimental
4. Diseño experimental: técnica de diferencia de medias
5. Diseño cuasiexperimental: método de emparejamiento
6. Método de diferencias en diferencias
7. Variables instrumentales
8. Regresión en discontinuidad
9. Conclusiones

### Referencias bibliográficas

**Palabras clave:** políticas públicas, evaluación de programas, evaluación de impacto.

**Keywords:** public policies, program evaluation, impact evaluation.

**Nº de clasificación JEL:** C21, H43

Fecha de entrada: 21/10/2022

Fecha de aceptación: 17/02/2023

---

## 1. INTRODUCCIÓN

En un entorno de recursos limitados, donde todos los gobiernos se esfuerzan por aumentar la efectividad, eficiencia y calidad de los servicios públicos que ofrecen, solucionando las diferentes necesidades y problemas que presenta la sociedad, resulta necesario tomar buenas decisiones, y para ello es imprescindible disponer de evidencia. La evaluación, junto al seguimiento, son las únicas herramientas que tienen los decisores y gestores para recopilar información, generar hallazgos y poder aprender y tomar mejores decisiones que ayuden a tener una intervención pública de mayor calidad (Feinstein, 2012, 2007), tratando de contestar preguntas como ¿Se ha logrado lo que nos propusimos hacer? ¿Podríamos haber hecho mejor las cosas? ¿Deberíamos continuar haciendo esto o intentar algo más?

Existen diferentes áreas en las que una intervención pública puede mejorar (Onrubia y Sánchez-Fuentes, 2019), y por ese motivo hay distintos tipos de evalua-

---

ciones (UNEG, 2020; Comisión Europea, 2013): algunas analizan el diseño del programa público, otras se centran en la implementación, o los procedimientos y reglas, en términos económicos, pero si nos interesa determinar si la intervención pública es efectiva, es decir, si está solucionando, o al menos mitigando, el problema que tenía la población, es necesario realizar una evaluación de impacto (Planas, 2005).

Hasta hace poco, cuando se evaluaba una intervención pública, era muy habitual que los gestores de los programas y los responsables de las políticas se centraran exclusivamente en medir e informar sobre los recursos utilizados y los productos obtenidos por el programa, sin prestar atención al logro de resultados, como indican Feinstein y Zapico (2010). Afortunadamente, en los últimos años se está produciendo un cambio hacia una formulación de políticas públicas en dos direcciones. Por un lado, intervenciones orientadas a resultados y, por otra parte, que estén basadas en evidencia (Jimeno, 2021; Zapico, 2022). Sin lugar a dudas, como observan Artés y Rodríguez-Sánchez (2022), la evaluación de impacto cumple ambos objetivos, dado que genera hallazgos que permiten aprender sobre el grado de efectividad.

Un gestor público interesado en solucionar necesidades y problemas de la sociedad debe realizar evaluaciones de impacto, ya que son herramientas que ayudan a seleccionar la mejor alternativa de política pública, así como un instrumento que permite saber si en realidad estamos (o no) logrando mejoras en la variable de resultado, reconociendo los programas efectivos, que solucionan problemas, de los que no. Otro motivo para realizar evaluaciones de impacto es la rendición de cuentas. Dado que vivimos en un contexto en que la sociedad demanda más información sobre la ejecución y efectividad de los programas públicos, la evaluación de impacto es un elemento de gran relevancia a la hora de proporcionar evidencia robusta y creíble sobre el desempeño y saber si un programa concreto está alcanzando sus resultados deseados. Además, este tipo de evaluaciones son fundamentales para construir conocimiento y aprendizaje acerca de la efectividad de los programas (Fedeia, 2021).

Este trabajo se estructura de la siguiente forma. En el siguiente apartado se presenta la evaluación de impacto y sus principales características, y a continuación se muestran los diferentes diseños en la EI. En el apartado cuarto se desarrollan los diseños experimentales, y los apartados del 5 al 8 se centran en los diseños cuasiexperimentales más habituales, finalizando con una sección dedicada a las conclusiones.

## 2. LA EVALUACIÓN DE IMPACTO

El elemento de inicio de cualquier evaluación viene dado por aquella característica del programa público que se quiere conocer, y en el caso de una Evaluación de Impacto (EI) es determinar la efectividad de la Intervención Pública (IP).

Es posible que solamente se quiera saber si un programa ha logrado los resultados que inicialmente tenía propuestos. Otra alternativa es no solamente estar interesado en si un programa es eficaz, sino también en comprender por qué y cómo los

programas tienen éxito o fracasan, a fin de mejorar el programa actual en el futuro. Por lo tanto, como indican Leeuw y Vaessen (2009) hay diferentes tipos de preguntas de «impacto» que se pueden contestar en una EI, o incluso más probablemente una combinación diferente de tales preguntas. Las opciones de preguntas típicas en IE son las siguientes:

**Cuadro nº 1. TIPOS DE PREGUNTAS EN EVALUACIÓN DE IMPACTO**

Tipo de pregunta	Pregunta
Sí (o no) /cuanto	¿Los beneficiarios de la intervención han mejorado (en la variable de resultado)?
Sí (o no) /cuanto	¿Cuánto de la mejora (en la variable de resultado) de los beneficiarios se debe exclusivamente a la IP?
Cómo	¿De qué forma la IP ha logrado (o no) generar diferencias (en el resultado)?
Dónde	¿Se pueden esperar estos resultados en otros lugares?

Fuente: Elaboración propia basada en Stern (2015).

Por lo que se refiere a la primera pregunta, se evalúa el efecto del programa unido al producido por otras intervenciones. Es lo que se conoce como análisis de contribución. Habitualmente, los programas que se desean analizar son una parte de toda la acción pública que trata de solucionar un problema, y que influyen simultáneamente en los resultados de los individuos. En esta situación, el efecto en la variable de resultado se debe a la multicausalidad y la combinación de «paquetes causales», donde el programa evaluado solamente es un factor más.

La segunda cuestión implica un análisis de atribución, donde el objetivo es determinar si la intervención pública realmente influyó en la mejora de los individuos, buscando aislar el efecto del programa del resto de factores que influyen en la variable de resultado analizada. La tercera cuestión analiza qué características internas y supuestos de la IP (y de la organización que la implementa) implican situaciones de éxito o fracaso, estudiando la teoría del programa y las características de la intervención pública. Finalmente, la cuarta pregunta considera la posibilidad de que una misma intervención, en un entorno tenga éxito pero en otro entorno no. En estos casos también será necesario analizar la teoría del programa, y los factores externos, para dar respuesta a esta pregunta.

Considerando estas opciones de preguntas en una evaluación de impacto, si el propósito de una evaluación es puramente la rendición de cuentas para mostrar que

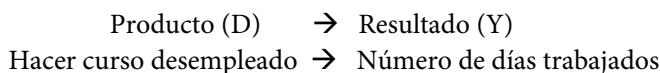
se han logrado los resultados, ya sea de manera conjunta o aislada, entonces el análisis causal será suficiente (preguntas 1 y 2), pero cuando el objetivo es aprender para mejorar el éxito o replicar programas en otros lugares, se necesitan explicaciones (preguntas 3 y 4).

## 2.1. Evaluación de impacto: el problema del **contrafactual**

62

Cuando se desea realizar un análisis de atribución de una política pública, la pregunta 2 del cuadro anterior se puede plantear en los siguientes términos: ¿Cuál sería la situación de los beneficiarios del programa si la intervención no hubiera tenido lugar?, o también ¿Las mejoras que se han producido en los individuos son el resultado directo del proyecto, o sencillamente es que se ha mejorado de todas formas? Estas preguntas buscan detectar la causalidad entre la política y la variable de resultado que trata de afectar, y no la (cor)relación entre ambas<sup>1</sup>.

La incorporación de la causalidad implica una mayor complejidad en el análisis, dado que a la variable de resultado que se desea analizar en una EI no solo le afecta la intervención pública estudiada, sino que existen muchos otros factores que también influyen sobre su valor final. Por ejemplo, supongamos que se desea estudiar si realizar un curso de capacitación del Ministerio de Empleo afecta a la empleabilidad de una persona, se tiene el siguiente diagrama:



Para que una persona trabaje (resultado, Y), no solo le afectará si realizó un curso del Ministerio de Empleo (D), sino que habrá otras variables o características también relevantes como la edad, si tiene hijos pequeños, si su pareja tiene trabajo, si el periodo es de crisis económica o no, el sector de actividad donde está, su nivel de estudios, etc., así que el objetivo de la evaluación de impacto es determinar cuánto supone esa intervención pública en la variable de resultados, sin considerar el resto de factores. El objetivo es aislar el efecto de la política (Rogers, 2012: OCDE, 2006).

Para poder contestar a la pregunta de la evaluación de impacto planteada inicialmente, es necesario una traducción matemática de esta cuestión: «¿Cuál es la situación (en la variable de resultado) que hubiera tenido un beneficiario de una intervención pública en el caso de no ser beneficiario?». En términos matemáticos, la pregunta anterior se expresa del siguiente modo:

$$\alpha_i = (Y_{1i}|D_i = 1) - (Y_{0i}|D_i = 1) \quad (1)$$

---

<sup>1</sup> Existe gran cantidad de referencias que exponen las diferentes metodologías en evaluación de impacto, algunas a destacar son los trabajos de Cunningham (2021), Frölich y Sperlich (2019), Gertler *et al.* (2016) o Khandker, Koolwal y Samad (2010), entre otros.

Donde  $Y_{1i}$  es el estado potencial de la variable de resultado de un individuo en el caso de recibir el programa, mientras que  $Y_{0i}$  es el estado potencial de ese mismo individuo en el caso de no recibir programa. Considerando que la variable D indica si un individuo es beneficiario de la política, D=1, o si no es beneficiario, D=0, entonces la ecuación anterior indica que el impacto de una política D en un individuo i, que denominamos  $\alpha_p$ , se puede expresar como la diferencia entre dos situaciones: el resultado en un individuo tratado si recibe la política, ( $Y_{1i} | D_i = 1$ ), menos el resultado que hubiera tenido ese mismo individuo tratado en el caso de no recibir la política, ( $Y_{0i} | D_i = 1$ ). Esta expresión permite calcular el impacto de un programa sobre un individuo que recibe tratamiento, porque en ambas situaciones, dado que es el mismo individuo, tienen las mismas características, salvo que en una de ellas recibe el programa mientras que en el otro no, por lo que la diferencia entre ambas situaciones solo puede ser debida al programa estudiado, aislando así el efecto de la intervención.

Lamentablemente, no es posible disponer de información de una persona en dos estados diferentes, por lo que la expresión (1) no se puede calcular en la realidad debido a que uno de los dos elementos de la parte derecha de la igualdad no es conocido por el evaluador. Es lo que en la literatura de evaluación se conoce como el problema del contrafactual.

Para solucionar esta situación se consideran dos individuos que presenten características similares, salvo que uno fue beneficiario (D=1) del programa y el otro no (D=0).

$$\alpha_i = (Y_{1i} | D_i = 1) - (Y_{0i} | D_i = 0) \quad (2)$$

En esta situación, se utiliza información de dos individuos distintos. Por un lado, el valor de la variable de resultado (cuando recibe tratamiento –  $Y_{1i}$ ) para un individuo que recibe tratamiento, es decir ( $Y_{1i} | D_i = 1$ ). Por otro lado, el valor de la variable de resultado (cuando no recibe tratamiento -  $Y_{0i}$ ) de otro individuo que en la realidad no recibe tratamiento, es decir ( $Y_{0i} | D_i = 0$ ). Dado que ambas cantidades son observadas por el evaluador, es una diferencia que se puede calcular en la realidad.

Sin embargo, el gran peligro que existe al utilizar este enfoque es que es difícil poder aislar el efecto del programa, porque estos dos individuos se deben parecer lo máximo posible, con la única diferencia entre ellos de que uno recibió el tratamiento y el otro no. Considerando la dificultad de disponer de dos individuos exactamente iguales en sus características, salvo por el hecho de que uno recibe el programa (D=1) mientras que el otro no (D=0), la solución de la estadística consiste en utilizar grupos de individuos en lugar de un solo individuo. Ahora, aunque existan factores diferenciales entre las personas, al utilizar grupos de individuos, siempre se puede seleccionar una serie de personas que, en promedio, presenten características

similares. Con la información de los individuos de cada grupo, el impacto de una política se obtiene como:

$$\alpha = E(Y|D = 1) - E(Y|D = 0) \quad (3)$$

Donde el término  $E(\cdot)$  se refiere a la esperanza (o la media). Ahora, el impacto del programa se obtiene calculando el promedio de la variable de resultado en el grupo de tratados,  $E(Y|D = 1)$ , menos el resultado promedio obtenido a partir de los individuos que no han recibido el programa,  $E(Y|D = 0)$ .

## 2.2. El sesgo de selección

Al evaluar el impacto de la intervención pública se consideran dos tipos de población: individuos que efectivamente han recibido la IP, llamados beneficiarios ( $D=1$ ) y que forman el grupo de tratamiento, e individuos que no son beneficiarios de la IP ( $D=0$ ), denominados grupo de control. Considerando estas dos tipologías de individuos, la pregunta que se desea contestar en la evaluación de impacto implica conocer si el programa ha creado una diferencia comparando la situación de un individuo que recibe tratamiento ( $D=1$ ) si recibe la ayuda,  $Y_1$ , –observado por el evaluador– respecto a su situación que hubiera tenido ese mismo individuo en el caso de no recibir ayuda,  $Y_0$ , –no observado por el evaluador–:

$$\alpha = E(Y_1|D = 1) - E(Y_0|D = 1) \quad (4)$$

Sin embargo, como no se dispone de información sobre qué le hubiera pasado a un tratado en ausencia de tratamiento,  $Y_0$ , se sustituye esta cantidad, por información de otro individuo, que en la realidad no recibe tratamiento ( $Y_0 / D = 0$ ) Por lo tanto, la estimación que hace el evaluador es:

$$dif_{tra-cont} = E(Y_1|D = 1) - E(Y_0|D = 0) \quad (5)$$

Comparando las dos ecuaciones previas, estas son diferentes en el segundo componente de la derecha de la igualdad. Por lo tanto, la estimación del evaluador no tiene por qué coincidir con el efecto verdadero de la intervención pública (IP). Esta diferencia entre el impacto verdadero y la estimación es lo que se conoce en la literatura como el sesgo de selección.

$$dif_{tra-cont} = \alpha + sesgo.selec \quad (6)$$

Y solamente la estimación que hace el evaluador es igual al verdadero impacto del programa cuando el sesgo de selección es cero.

Este sesgo de selección indica, de manera simplificada, que los individuos del grupo de tratamiento y control no son directamente comparables, porque hay diferencias entre ellos, y no solo porque unos reciben programa y los otros no. Estas características diferenciales, como edad, sexo, nivel de estudios, gustos, etc., hacen

que sean más propensos a pertenecer al grupo de tratamiento o al grupo de control. Por lo tanto, que un individuo esté en el grupo de tratamiento no es aleatorio.

Existen diferentes tipos de sesgos de selección, pudiéndose clasificarse en función del individuo que crea el sesgo y si es observado (o no) por el evaluador. Considerando al individuo que lo genera, se pueden considerar sesgos debido a la política, ya que están focalizadas en una parte de la población que presenta algún determinado problema –nivel de pobreza, tiempo en desempleo, tamaño de empresa–, o sesgo generado por los potenciales beneficiarios de la política (autoselección), dado que en la mayoría de intervenciones el Estado ofrece la oportunidad a los individuos de participar en la política, de convertirse en beneficiarios efectivos, pero son las personas las que finalmente deciden si reciben la intervención pública o no, es decir, una autoselección. Otra clasificación del sesgo de selección es en función de si la característica que afecta a que un individuo esté en el grupo de tratamiento o control (no) se encuentra en el fichero de datos que posee el evaluador. Entonces, el sesgo es en variable (no) observada.

### **3. DISEÑO DE LA EVALUACIÓN DE IMPACTO: EXPERIMENTAL Y CUASIEXPERIMENTAL**

Para que la estimación del evaluador sea igual al impacto verdadero del programa, según (6), existen dos posibilidades. Por un lado, la utilización de «diseños experimentales», donde la asignación de los individuos al grupo de control y tratamiento se realiza de forma aleatoria. En estos casos el sesgo de selección es cero, por construcción del diseño de evaluación. La otra alternativa son los «diseños cuasimexperimentales», situación en la que no hay asignación aleatoria de los individuos al grupo de tratamiento y control, lo que supone la aparición de un sesgo de selección diferente de cero, siendo necesario corregirlo mediante la aplicación de técnicas econométricas y estadísticas (Imbens y Rubin, 2015).

Dentro de los diseños experimentales, el método de estimación empleado es la técnica de diferencia de medias, mientras que en los diseños cuasimexperimentales las metodologías de estimación más habituales son diferencias en diferencias, técnicas de emparejamiento, destacando el Propensity Score Matching, variables instrumentales y regresión en discontinuidad.

Pasamos a continuación a describir cada una de ellas, indicando los supuestos empleados para su utilización, la estructura de la base de datos necesaria, opciones de estimación del impacto, y posibles riesgos en su utilización.

### **4. DISEÑO EXPERIMENTAL: TÉCNICA DE DIFERENCIA DE MEDIAS**

Cuando la intervención pública (IP) se encuentra en la fase de diseño, se puede proponer un diseño experimental de la evaluación de impacto (EI) (Banarjee y

Duflo, 2017). Las características necesarias para que exista un diseño aleatorio o experimental son: a) la evaluación se diseña a la vez que la política; b) el evaluador determina al «azar» qué individuos estarán en control y tratamiento, utilizando, por ejemplo, el lanzamiento de un dado; c) una vez que un individuo ha sido seleccionado en un grupo, no puede cambiarse del grupo donde ha caído.

Después de implementar la intervención pública en los individuos asignados aleatoriamente, se recopila información de la variable de resultado Y, tanto en los individuos de control como en los de tratamiento. En la tabla nº 1 se presenta un ejemplo de base de datos que contiene la información mínima para poder estimar el impacto utilizando este método, y la posibilidad de incorporar información adicional (color gris) que mejore la calidad estadística de las estimaciones.

*Tabla nº 1. INFORMACIÓN NECESARIA Y EXTRA PARA REALIZAR UNA EVALUACIÓN DE IMPACTO CON UN DISEÑO EXPERIMENTAL*

INFORMACIÓN NECESARIA			INFORMACIÓN EXTRA			
Momento tiempo	Línea base (t=0)	Post tratamiento (t=1)	Línea base (t=0)			
Individuo	Tratamiento (D)	Resultado (Y)	Var X <sub>1</sub>	Var X <sub>2</sub>	Y	...
1	0	87	1	23	60	
2	1	56	0	25	21	
...	...	...	...	....	...	..
i	D <sub>i</sub>	Y <sub>i</sub>	X <sub>1i</sub>	X <sub>2i</sub>	Y <sub>0i</sub>	

Fuente: Elaboración propia.

Supongamos que se desea evaluar el efecto que un programa de asistencia a búsqueda de empleo (D) tiene sobre los días trabajados al año siguiente de los individuos que lo reciben (Y). Para estimar el impacto se propone un diseño experimental, con una asignación aleatoria de personas al grupo de tratamiento o control. Si se cumple esta asignación aleatoria, solo es necesario disponer de la información de la variable de resultado –días trabajados en el año siguiente– un periodo posterior a implementarse el programa. Si, además, hay información extra, especialmente en la línea de base (t=0), es posible contrastar si realmente tenemos un diseño experimental, ya que estas características (sexo- x<sub>1</sub>- , edad – x<sub>2</sub>-, etc.) deben ser parecidas en el grupo de control y tratamiento. Si no se cumple este supuesto, entonces existen sesgos de selección en la asignación de los individuos al grupo de tratamiento y control.

Una vez que se dispone de la base de datos como la dada en la tabla nº 1, para estimar el impacto se puede emplear un enfoque estadístico o econométrico (Glen-nester y Takavarasha, 2013).

Considerando la aproximación estadística, después de implementarse el programa, en el momento t=1 se observa la variable de resultado Y en todos los individuos estudiados y se generan dos medias, el valor promedio de la variable de interés en el grupo de tratamiento, dado por  $E(Y | D = 1)$ , y el valor promedio de la variable de interés en el grupo de control,  $E(Y | D = 0)$ . Entonces, el impacto del programa viene dado por:

$$\alpha = E(Y|D = 1) - E(Y|D = 0) \quad (7)$$

Esta cuantía calculada no permite determinar al evaluador si el valor obtenido es lo suficientemente grande para considerarse distinta de cero en términos estadísticos. Para determinar cuánto es suficiente desde una óptica estadística, es necesario realizar un contraste estadístico, valorando si la diferencia entre medias de dos poblaciones –tratados y controles– es estadísticamente distinta de cero:

Hipótesis nula: El programa no tiene impacto  $H_0: E(Y|D = 1) = E(Y|D = 0)$ .

Hipótesis alternativa: Programa si tiene impacto  $H_A: E(Y|D = 1) \neq E(Y|D = 0)$ .

En la lectura de la hipótesis nula, dado que la media de la variable de resultado del grupo de tratados es (estadísticamente) similar a la media de la variable en el grupo de control, implica que no se observan diferencias entre los dos grupos en la variable de interés, por lo que el programa no es capaz de influir en el valor de Y. No tiene ningún impacto sobre la variable de resultados. Para contrastar estas hipótesis será necesario realizar un test de diferencia de medias.

En la aproximación econométrica se puede especificar una ecuación que capture la relación entre la variable recibir programa y la variable resultado dada por:

$$Y_i = \alpha_0 + \alpha_1 D_i + U_i \quad (8)$$

Donde Y es la variable de resultado, mientras que D es el único regresor de la ecuación y captura si ese individuo recibe el programa, pudiendo tomar solo dos valores:

$$D_i = \begin{cases} 1 & \text{si recibe tratamiento} \\ 0 & \text{si no recibe tratamiento} \end{cases} \quad (9)$$

La variable U es el término de error y contiene el resto de factores que afectan a Y, de los que no se dispone de información.

Recordando que el impacto de un programa D se obtiene como la diferencia de  $E(Y | D = 1) - E(Y | D = 0)$ , se puede calcular el valor esperado de la variable Y para el grupo de tratamiento (cuando D=1) y para el grupo de control (cuando D=0) por separado, sustituyendo en (8). Entonces, el impacto del programa se obtiene como:

$$impactprog = E(Y|D = 1) - E(Y|D = 0) = \frac{\alpha_0 + \alpha_1}{E(Y|D=1)} - \frac{\alpha_0}{E(Y|D=0)} = \alpha_1 \quad (10)$$

Por lo tanto, en el enfoque econométrico, para obtener el impacto de un programa hay que prestar atención al valor de la estimación del parámetro  $\alpha_1$  de la ecuación anterior.

#### 4.1. Ventajas, limitaciones y potenciales riesgos

Sin lugar a dudas, los diseños experimentales permiten una comparación sólida entre grupos, minimizando el sesgo en la selección de la muestra. Sin embargo, también hay una serie de limitaciones y riesgos. A continuación, destacamos los más relevantes.

No todas las políticas son adecuadas para un diseño experimental. Este método funciona mejor cuando: a) se conocen bien los mecanismos por los que se espera que funcione la intervención, lo que no siempre es el caso; b) existe poca variación en la ejecución de la intervención, ya que estos diseños necesitan una ejecución rigurosa y uniforme; c) se cumplen ciertos requisitos de información, no solo de la variable de resultados sino también datos de otras características de los grupos de tratamiento y de comparación para comprobar el equilibrio en la asignación (Berger, 2005).

Además, pueden presentar problemas de validez externa, debido a lo no representatividad de la muestra empleada en el análisis, o problemas de validez interna, al producirse fallos en el mecanismo de asignación aleatoria, desgaste de la muestra y problemas de abandono del programa por parte, especialmente, de los individuos de control.

Otra limitación son las relacionadas con las externalidades, ya que no siempre es posible garantizar que el grupo de control no se vea afectado por el programa. El «cegamiento» clásico en el área médica, en el que tanto los participantes y gestores del programa no saben quién está en el grupo de tratamiento o en el de control, rara vez es factible en las intervenciones sociales, lo que conlleva la existencia de externalidades del programa y efectos de comportamiento.

Finalmente, también es necesario destacar los posibles altos costes de este tipo de diseño, así como en algunos casos, la aparición de problemas éticos.

### 5. DISEÑO CUASIEXPERIMENTAL: MÉTODO DE EMPAREJAMIENTO

Cuando alguna de las tres condiciones necesarias para tener un diseño experimental no se cumple, nos encontramos ante un diseño cuasiexperimental. En este caso, dado que la asignación de los individuos a los grupos de tratamiento y control no se ha producido de manera aleatoria, entonces existen sesgos de selección y será necesario utilizar técnicas cuantitativas que logren minimizar lo máximo posible este sesgo (Barnow, Cain y Goldberger, 1981).

Cuando el sesgo de selección se debe a características de los individuos observadas por el evaluador, para estimar el efecto del programa sobre la variable de resultado se utilizan técnicas de emparejamiento o *matching* (Todd, 2007), tratando de agrupar individuos que presentan características observadas similares<sup>2</sup>. En estos casos, la base de datos siguiente muestra la información necesaria para aplicar el método de emparejamiento:

Tabla nº 2. INFORMACIÓN PARA REALIZAR UNA EVALUACIÓN DE IMPACTO DE EMPAREJAMIENTO

Momento	Línea de base t=0			t=1		
	Individuo	Var X <sub>1</sub>	Var X <sub>2</sub>	.....	Tratamiento (D)	Resultado (Y)
1	0	1			1	21
2	2	1			0	23
3	10	0			1	41
...	...	...			...	..
I	X <sub>1i</sub>	X <sub>2i</sub>	...	D <sub>i</sub>	Y <sub>i</sub>	

Fuente: Elaboración propia.

### 5.1. Estimación del impacto aplicando el Método de **matching** o emparejamiento

La idea que subyace en esta metodología es lograr individuos del grupo de control lo más similares posible al grupo de tratamiento en términos de las características observables, siendo necesario realizar una serie de fases.

En la primera etapa, se estratifican los datos de la muestra en función de los diferentes valores de las características observadas x, por ejemplo, sexo, tramos de edad, etc.

A continuación, dentro de cada estrato (o celda) se computa la diferencia en la variable de resultado promedio entre el grupo de tratamiento y el de control, es decir:

$$\alpha_{X=x} = E(Y^D | D = 1, X = x) - E(Y^C | D = 0, X = x) \quad (11)$$

<sup>2</sup> El supuesto para poder emplear esta metodología es que la selección de los individuos al grupo de tratamiento (o control) se deba a variables observadas por el evaluador.

Donde  $E(Y^D|D=1, X=x)$ , es la media de la variable de resultado en el grupo de tratados cuando tienen las características  $X=x$ , mientras que  $E(Y^C|D=0, X=x)$  es el valor promedio de individuos en el grupo de control con esas mismas características  $X=x$ .

Finalmente, se obtiene el promedio de estas diferencias que se han calculado para cada valor  $X=x$ , con respecto a la distribución de  $X$ , en la población de unidades tratadas, dada por  $N_T$ , obteniendo:

$$\alpha = \frac{1}{N_T} \sum_{X=x} (E(Y^D|D=1, X=x) - E(Y^C|D=0, X=x)) \quad (12)$$

Aunque esta técnica de emparejamiento es muy intuitiva, presenta ciertas limitaciones cuando hay muchas variables de estratificación, o alguna de las características empleadas en la estratificación es continua, como la edad, o la muestra de datos disponible es pequeña. Estas situaciones pueden suponer la ausencia de soporte común también conocido como maldición de la dimensionalidad, es decir, celdas que contienen solo tratamientos, sin controles, o vice-versa, generándose una pérdida de información, ya que en esos casos no se pueden emparejar individuos. Como solución, se desarrolla el método Propensity Score Matching que se describe a continuación.

## 5.2. Estimación del impacto con el Método de Propensity Score Matching (PSM)

Este método consiste en colapsar todas las características de los individuos en un único valor, de tal forma que a la hora de emparejar ya no se tenga que utilizar variables, sino que, mediante la aplicación de un modelo de regresión, todas estas posibles combinaciones de características se colapsen en un valor que indique la probabilidad que un individuo tiene de pertenecer al grupo de tratamiento ( $D=1$ ) de acuerdo a sus características observadas  $X$ , según Dehejia y Wahba (2002) y Rosenbaum y Rubin (1983) entre otros. Ahora el emparejamiento de individuos similares será en función de esta propensión o probabilidad que tengan para ser beneficiarios (Caliendo y Kopeinig, 2008).

El primer paso consiste en estimar el *propensity score*. Para ello se realiza la estimación de una regresión econométrica en el que la variable dependiente es participar o no en el programa,  $D_i$ , en función de las características de los individuos,  $X_i$ . Como  $D_i$  solo puede tomar dos valores (0,1) se emplea un modelo Probit o Logit. Una vez estimados los parámetros de la ecuación, a continuación se calcula la probabilidad estimada de un individuo de participar en el programa ( $D=1$ ), de acuerdo a sus características observadas, es decir  $\hat{P}(D=1|X=x)$ .

Tras comprobar la calidad de la estimación previa, analizando el soporte común de  $\hat{P}$  donde hay tratados y controles para poder realizar los emparejamientos, y el contraste de equilibrado de muestras, se procede a emparejar a cada individuo trata-

do con aquel individuo de control cuyo valor estimado del *propensity score* esté más próximo, descartándose las observaciones de control que queden sin emparejar, y se compara la variable de resultado Y entre el tratado y (los) control(es) que se considera(n) similar(es).

$$Y_i^D - \sum_{j \in C} w(i, j) Y_j^C$$

Donde  $Y_i^D$ ,  $Y_j^C$  es la variable de resultado del tratado y control(es) respectivamente, mientras que  $w(i, j)$  es una función de ponderación que determina cuántos controles, y en qué forma, se usarán para comparar contra un determinado tratado (vecino más cercano, radio, kernel, etc.). Así se calcula el impacto del programa en un tratado «i». Finalmente se obtiene la media de estas diferencias individuales, para obtener el impacto promedio general.

$$\hat{\alpha} = \frac{1}{N_D} \sum_{i \in D} [Y_i^D - \sum_{j \in C} w(i, j) Y_j^C] \quad (13)$$

siendo  $N_D$  el número total de tratados empleados en el análisis (Guo y Fraser, 2019).

### 5.3. Ventajas, limitaciones y potenciales riesgos

Los métodos de emparejamiento permiten estimar el impacto de un programa correctamente cuando no hay asignación aleatoria calculando el efecto medio del tratamiento sobre todos los tratados. Sin embargo, esta aproximación no es ajena a ciertas limitaciones, que pasamos a describir.

Por un lado, el emparejamiento sólo debe utilizarse en los casos en los que el evaluador tenga un conocimiento claro y detallado de los criterios de elegibilidad de una intervención (reglas claras y explícitas de selección) y estas variables estén en la base de datos. Dado que el emparejamiento sólo puede realizarse sobre las características que no se han visto afectadas por la intervención pública, estos datos deben ser invariables en el tiempo, como el año de nacimiento, o recogidos antes de producirse la IP. Además, en caso de existir sesgos debidos a características no observadas, la evolución de impacto no es robusta, mientras que, por lo que respecta al emparejamiento utilizando el PSM, pueden existir dudas por la utilización de esa probabilidad a la hora de emparejar, ya que se puede considerar una especie de una caja negra.

## 6. MÉTODO DE DIFERENCIAS EN DIFERENCIAS

Cuando el sesgo de selección de los individuos depende de variables no observadas, el método de diferencias en diferencias, también conocido como «Dif-in-Dif», permite corregir esta desviación, generando estimaciones robustas del impacto de una IP<sup>3</sup>.

<sup>3</sup> El principal supuesto para aplicar el método «Dif-in-Dif» es que la selección de los individuos al grupo de tratamiento o control se basa en variables no observadas, también denominada heterogeneidad no observada, y que esta heterogeneidad no observada es constante en el tiempo. Esto implica que, en

Por lo que respecta a las necesidades de información, para poder utilizar el método «Dif-in-Dif» es necesario disponer de información de la variable de resultado Y en dos momentos de tiempo: la línea de base (momento t=0) y posterior a tratamiento (momento t=1). La tabla nº 3 ofrece un ejemplo de la información necesaria para utilizar este método.

*Tabla nº 3. INFORMACIÓN NECESARIA Y ADICIONAL PARA UNA EVALUACIÓN DE IMPACTO CON «DIF-IN-DIF»*

Momento	INFORMACIÓN NECESARIA			INFORMACIÓN EXTRA		
	t=1	t=0 (línea base)	t=1	t=0	t=1	...
Individuo	Tratamiento (D)	Resultado (Y)	Resultado (Y)	Var X <sub>1</sub>	Var X <sub>1</sub>	
1	0	57	82	7		...
2	1	40	56	2		...
3	0	55	67	9		...
...	...	...	...	...	.	...
i	D <sub>i</sub>	Y <sub>0,i</sub>	Y <sub>1,i</sub>	X <sub>10i</sub>	X <sub>11i</sub>	...

Fuente: Elaboración propia.

Donde  $Y_{0,i}$  hace referencia al valor de la variable de resultado Y en el individuo «i» en línea de base, momento de tiempo t=0, mientras que  $Y_{1,i}$  es el valor de la variable de resultado Y en el individuo «i» después del tratamiento, t<sub>1</sub>.

Con la base de datos anterior, la estimación del impacto de la IP se puede realizar mediante una aproximación estadística o econométrica.

La aproximación estadística implica el cálculo de cuatro esperanzas condicionadas de la variable de resultado Y, en función del momento de tiempo y del grupo, dadas por:

$$\begin{aligned} E(Y|\text{tratado, antes}) &= \bar{Y}_{D0} & E(Y|\text{tratado, despues}) &= \bar{Y}_{D1} \\ E(Y|\text{control, antes}) &= \bar{Y}_{C0} & E(Y|\text{control, despues}) &= \bar{Y}_{C1} \end{aligned}$$

A continuación, se obtiene la estimación de la diferencia temporal para cada grupo, como:

$$\text{Grupo de tratamiento: } \Delta Y_D = Y_{D1} - Y_{D0} \quad (14)$$

---

ausencia de tratamiento, los dos grupos se comportan de manera similar a lo largo del tiempo (hipótesis de caminos paralelos).

$$\text{Grupo de control: } \Delta Y_C = Y_{C1} - Y_{C0} \quad (15)$$

Estas diferencias capturan cuánto ha crecido, en promedio, la variable de interés durante los dos períodos de tiempo (de  $t_0$  a  $t_1$ ) para cada uno de los grupos analizados, control y tratamiento por separado.

Finalmente se calcula la diferencia de la diferencia, estimando el diferencial entre el incremento temporal de los dos grupos obtenidos en la etapa anterior:

$$\alpha = \Delta Y_D - \Delta Y_c = (\bar{Y}_{D1} - \bar{Y}_{D0}) - (\bar{Y}_{C1} - \bar{Y}_{C0}) \quad (16)$$

El impacto del programa captura la diferencia existente entre el crecimiento del grupo de tratamiento respecto al grupo de control. El razonamiento que subyace detrás de (16) es que, para el grupo de control, la primera diferencia que captura el crecimiento entre los dos períodos de tiempo,  $(\bar{Y}_{C1} - \bar{Y}_{C0})$ , es debida a muchos motivos, de los que no se dispone de información, variables no observadas para el evaluador, y que se pueden denominar «entorno». Para el grupo de tratamiento, la primera diferencia, que es el crecimiento experimentado en ese periodo,  $(\bar{Y}_{D1} - \bar{Y}_{D0})$ , es debido a los mismos motivos que el grupo de control, dado que el «entorno» afecta de manera similar a todos los individuos, y además este grupo recibió tratamiento. Por lo tanto:

$$\alpha = \Delta Y_D - \Delta Y_c = \left( \frac{\text{entorno} + \text{prog}}{\bar{Y}_{D1} - \bar{Y}_{D0}} \right) - \left( \frac{\text{entorno}}{\bar{Y}_{C1} - \bar{Y}_{C0}} \right) = (\text{entorno} + \text{prog}) - \text{entorno} = \text{prog}.$$

La ecuación anterior nos ofrece la cuantía del impacto, pero por si sola no es capaz de informar si este valor del impacto es estadísticamente significativo, por lo que será necesario realizar un contraste de hipótesis del siguiente modo:

Hipótesis nula: el programa no tiene impacto.  $H_0 : \Delta \bar{Y}_D = \Delta \bar{Y}_C$

Hipótesis alternativa: el programa sí tiene impacto.  $H_A : \Delta \bar{Y}_D \neq \Delta \bar{Y}_C$

Donde  $\Delta \bar{Y}_D = (\bar{Y}_{D1} - \bar{Y}_{D0})$  y  $\Delta \bar{Y}_C = (\bar{Y}_{C1} - \bar{Y}_{C0})$  respectivamente. Para realizar el contraste se realiza un test de diferencia de medias entre dos poblaciones (tratados y controles).

En la aproximación econométrica, se puede especificar una ecuación que capture la relación entre la variable recibir programa y la variable resultado donde se ve el impacto del programa. La ecuación viene dada por:

$$Y_i = \alpha_0 + \alpha_1 D_i + \alpha_2 M_t + \alpha_3 (D_i \times M_t) + U_i \quad (17)$$

Donde  $Y_i$  es la variable de resultado, mientras que  $D_i$  captura si el individuo pertenece al grupo de tratamiento o no, y la variable  $M_t$  determina si la información es de línea de base o ex-post. Ambas son variables dicotómicas con dos posibles valores:

$$D_i = \begin{cases} 1 & \text{si recibe tratamiento} \\ 0 & \text{si no recibe tratamiento} \end{cases} \quad M_t = \begin{cases} 1 & \text{si informacion post tratamiento} \\ 0 & \text{si informacion en linea base} \end{cases}$$

Siendo  $(D_i \times M_t)$  la interacción de ambas. La variable U es el término de error, y captura el resto de factores que afectan a la variable de interés Y, y de los cuales no se dispone de información.

74

Considerando (17), y calculando las esperanzas condicionadas asociadas al grupo de control ( $Y|D = 0, M = 0$ ),  $E(Y|D = 0, M = 1)$  y para el grupo de tratados  $E(Y|D = 1, M = 0)$ ,  $E(Y|D = 1, M = 1)$  se obtiene que el impacto del programa está capturado por el parámetro  $\alpha_3$ , que estima si el crecimiento de la variable Y a lo largo del tiempo entre ambos grupos es lo suficientemente distinto.

### 6.1. Ventajas, limitaciones y potenciales riesgos

El método «Dif-in-Dif» es intuitivamente sencillo, además de corregir sesgos en características no observadas para el evaluador, pero hay una serie de limitaciones y riesgos a la hora de implementarlo (Bertrand, Duflo y Mullainathan, 2004). La principal limitación está en los requisitos de información, ya que para aplicar este método es necesario disponer de datos antes y después del tratamiento, tanto de la variable de resultados como otras características de los tratados y controles, por lo que, si las necesidades de información no han sido detectadas con antelación, es posible que no dispongamos de datos en la línea de base. También es posible que las estimaciones de «Dif-in-Dif» no presenten buenas propiedades si a lo largo del tiempo se producen cambios en uno de los grupos, pero no en el otro, o si los dos grupos tenían tendencias diferentes antes de la intervención, no cumpliendo el supuesto de caminos paralelos. Además, si se espera que el efecto se manifieste después de unos años, existe la posibilidad de que otros eventos contaminen los resultados, unido a que pueda producirse cierta pérdida de información debido al desgaste, lo que condicionaría los resultados obtenidos.

## 7. VARIABLES INSTRUMENTALES

Cuando existen características del individuo, tanto observadas y no observadas, que influyen en que sea beneficiario del programa, la utilización de un instrumento o variable instrumental (VI) puede ser una buena alternativa para solucionar el problema de endogeneidad en la participación del programa (Angrist y Krueger, 2001)<sup>4</sup>. En estos casos, las necesidades de información para utilizar este método se presentan en la tabla nº 4.

---

<sup>4</sup> Los supuestos necesarios para aplicar el método de variables instrumentales son: (1) existe una variable, llamada instrumento, que permite corregir el problema de endogeneidad en la variable de tratamiento. Además, este instrumento tiene que cumplir dos condiciones adicionales: (2) el

**Tabla nº 4. INFORMACIÓN NECESARIA Y ADICIONAL PARA UNA EVALUACIÓN DE IMPACTO CON VARIABLE INSTRUMENTAL**

Momento	Línea base t=0			t=0	t=1	t=1
Individuo	Var X <sub>1</sub>	Var X <sub>2</sub>	.....	Instrumento (Z)	Tratamiento (D)	Resultado (Y)
1	0	0		1	0	211
2	2	0		0	0	230
3	10	0		1	1	341
4	3	1		0	1	199
...	...	...			...	..
i	X <sub>1i</sub>	X <sub>2i</sub>	...	Z <sub>i</sub>	D <sub>i</sub>	Y <sub>i</sub>

Fuente: Elaboración propia.

Para poder corregir la endogeneidad de la participación en el programa, D<sub>i</sub>, es necesario disponer de una nueva variable, Z<sub>i</sub>, denominada variable instrumental (VI) o instrumento (Heckman, 1997). Por lo tanto, para aplicar este enfoque es necesario una base de datos con al menos 3 variables: el resultado (Y), el tratamiento (D) y un instrumento (Z).

Dependiendo de las características del instrumento, es posible considerar diferentes tipos de estimaciones de V.I. Por un lado, existen casos en los que el evaluador tiene cierta capacidad de manipulación del experimento de evaluación, logrando que Z se genere mediante un procedimiento aleatorio, como lanzar una moneda, garantizando que Cov(Z, U) =0, siendo U las características no observadas de los individuos. Estas situaciones se producen cuando hay una aceptación parcial por parte del beneficiario, y en los procesos de fomento de participación aleatoria. El estimador a emplear en estos casos es el Estimador de Wald, que pasaremos a describir a continuación. También hay situaciones en las que el evaluador no tiene capacidad de influir en Z, pero el instrumento aún cumple la condición de exogeneidad Cov(Z,U)=0. En estos casos se utilizará el método de estimación de Mínimos Cuadrados en Dos Etapas, con una primera etapa que predice la probabilidad de ser tratado en función del instrumento y de otros factores exógenos, mientras que en la segunda etapa se estima la ecuación del efecto que el programa tiene sobre el resultado, utilizando la probabilidad de tratamiento estimada previamente.

---

instrumento influye sobre la participación; y (3) el instrumento no tiene relación con las variables no observadas de los individuos.

## 7.1. El Estimador de Wald

En ciertas situaciones, el evaluador tiene la capacidad para generar de forma aleatoria el instrumento, Z, pudiendo tomar dos valores (1 si se selecciona aleatoriamente y 0 si no), y este instrumento afecta a que los individuos sean beneficiarios de la política D, con valores (1, si es beneficiario y 0 si no), entonces la relación existente entre las 3 variables es la siguiente:

$$\text{Variable aleatoria } Z \text{ (0,1)} \rightarrow \text{Tratamiento } D \text{ (0,1)} \rightarrow \text{Resultado (Y)}$$

Siendo recomendable utilizar el estimador de Wald para estimar el impacto del programa analizado, que viene dada por:

$$\alpha = \frac{E(Y|Z=1) - E(Y|Z=0)}{E(D=1|Z=1) - E(D=1|Z=0)} \quad (18)$$

Donde  $E(Y|Z=1)$  es el valor esperado de Y en el grupo de unidades seleccionadas para tratamiento,  $E(Y|Z=0)$  es el valor esperado de Y en el grupo de individuos seleccionados en el control, mientras que  $E(D=1|Z=1)$  y  $E(D=1|Z=0)$  hace referencia a la participación en el tratamiento de los individuos seleccionados para tratamiento y control respectivamente.

Analizando en detalle (18), se observa que en el numerador se estima el impacto cuando estamos en un diseño experimental, ya que Z se diseña de manera aleatoria:

$$E(Y|Z = 1) - E(Y|Z = 0)$$

Mientras que en el denominador se estima la diferencia de participación en el programa ( $D=1$ ) dependiendo del grupo en el que esos individuos fueron seleccionados inicialmente de forma aleatoria.

$$E(D = 1|Z = 1) - E(D = 1|Z = 0)$$

Por lo tanto, este estimador calcula el efecto en la variable de resultado de ser seleccionado (de manera aleatoria), escalado por el diferencial de participación en el programa dependiendo de haber sido seleccionado.

## 7.2. Ventajas, limitaciones y potenciales riesgos

Este método de variable instrumental (VI) es apropiado cuando se evalúan intervenciones en las que los individuos pueden autoseleccionarse sugiriendo que tienen características que los hacen más propensos a ser tratados.

Las principales limitaciones de esta metodología son, por un lado, encontrar un instrumento válido si no se planifica con antelación. Además, es difícil localizar una VI que satisfaga los dos supuestos cruciales: que influya fuertemente sobre la participación en el programa, pero que no esté correlacionada con ninguna característica que afecte a los resultados. Finalmente destacar que cuando se utiliza el método de

VI sólo se estiman efectos locales, es decir, el impacto sobre los que están en el margen de la participación.

## 8. REGRESIÓN EN DISCONTINUIDAD

Este método se aplica cuando ser beneficiario del programa depende de una característica observada y continua (Cattaneo, Idrobo y Titiunik, 2019). Ahora, la participación en la intervención pública (IP) se determina por una regla que utiliza una determinada característica continua que tienen los individuos, como la edad, nota académica o número de trabajadores. Si el valor de la variable está por debajo (arriba) de un determinado valor crítico fijado por la política evaluada, denominado umbral o frontera, ese individuo resulta elegible para recibir el programa, mientras que si queda por encima (debajo) de ese valor no será elegible<sup>5</sup>.

La utilización de este método de Regresión en Discontinuidad (RD) implica disponer de información de 3 variables, en color beige, con la opción de información extra en gris, tal y como se señala en la tabla nº 5.

*Tabla nº 5. INFORMACIÓN NECESARIA Y ADICIONAL PARA UNA EVALUACIÓN DE IMPACTO CON REGRESIÓN EN DISCONTINUIDAD*

Momento	INF. ADICIONAL			INFORMACIÓN NECESARIA		
	Línea base t=0	t=0	t=1	t=1		
Individuo	Var X <sub>1</sub>	Var X <sub>2</sub>	.....	Instrumento (Z)	Tratamiento (D)	Resultado (Y)
1	1	17		24	1	37
2	0	34		25	1	56
3	1	23		26	0	78
4	0	34		27	0	36
...	...	...	...		...	...
i	X <sub>1i</sub>	X <sub>2i</sub>		Z <sub>i</sub>	D <sub>i</sub>	Y <sub>i</sub>

Fuente: Elaboración propia.

La información necesaria para aplicar el método de Regresión en Discontinuidad implica poder disponer de una característica continua de los individuos, deno-

<sup>5</sup> Los supuestos para poder aplicar el método de regresión en discontinuidad implican que la asignación a tratamiento depende del valor que toma una característica observada y continua previo a que se produzca el tratamiento, y que influye en pertenecer al grupo de tratamiento o control. Además, la probabilidad de ser tratado debe cambiar de forma discontinua en función de esta característica continua.

minada Z, que influye en que pertenezcan al grupo de tratamiento, D, además de la variable de resultado Y después de aplicar la intervención pública. En esta situación, la relación existente entre las 3 variables es la siguiente:

$$\text{Variable continua } Z \rightarrow \text{Tratamiento (D)} \rightarrow \text{Resultado (Y)}$$

Considerando este diagrama, es fundamental que la selección de un individuo al grupo de tratamiento o control se determine por la posición respecto al valor frontera, llamado  $Z^*$ , definido a lo largo de la variable continua Z.

Cuando se desea realizar una regresión en discontinuidad (RD), se pueden dar dos alternativas: la «RD estricta», en la que si el individuo evaluado tiene un Z inferior (o superior) al valor critico  $Z^*$  entonces es tratado, y si está por encima (debajo) del control. Por lo tanto, se produce una obligatoriedad de recibir tratamiento a partir de cierto valor de Z, cumpliéndose:

$$D_i = \begin{cases} 1 & \text{si } Z_i < Z^* \\ 0 & \text{si } Z_i \geq Z^* \end{cases} \quad (19)$$

También existe la situación de «RD difusa». En este caso, que un individuo tenga un Z inferior (superior) al valor critico  $Z^*$  no implica que obligatoriamente sea tratado (control). Ahora, esta situación solo afecta a que la probabilidad de ser tratado es más alta en un lado de la frontera que en el otro. Entonces:

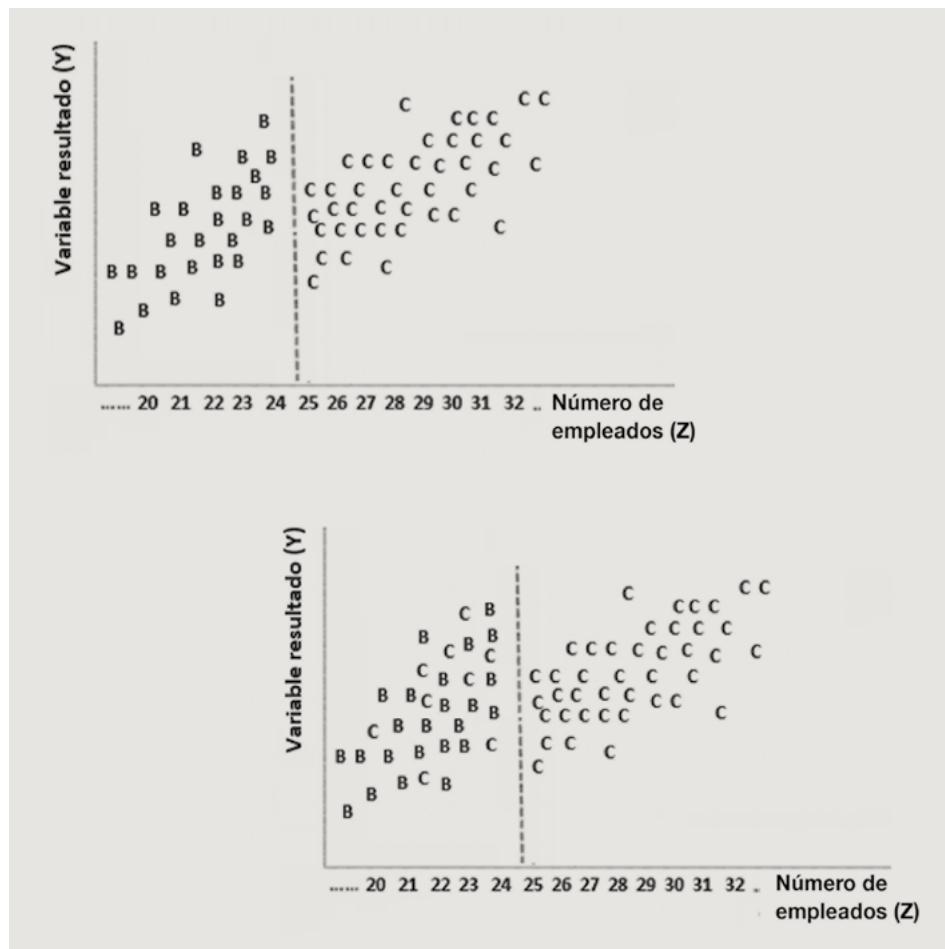
$$D_i = \begin{cases} P^-(D = 1) & \text{si } Z_i < Z^* \\ P^+(D = 1) & \text{si } Z_i \geq Z^* \end{cases} \quad (20)$$

Donde  $P^-(D = 1)$  indica la probabilidad de ser beneficiario del programa a la izquierda del valor  $Z^*$ , mientras que  $P^+(D = 1)$  es la probabilidad de ser tratado en la derecha del valor frontera.

Desde un punto de vista gráfico, la diferencia entre la RD estricta y la RD difusa consiste en detectar si en uno (o en los dos) lados de la frontera  $Z^*$  existen simultáneamente beneficiarios (B) y controles (C), situación que se da en la RD difusa, mientras que si a un lado solo hay tratados y al otro controles estaremos ante un enfoque estricto. Supongamos como ejemplo que se desea evaluar el impacto de una política destinada a empresas con menos de 25 trabajadores, pudiéndose dar las situaciones propuestas en la figura nº 1.

En el gráfico de la izquierda, la política obliga que, para empresas con menos de 25 trabajadores, todas reciben el programa, son tratadas, mientras que las que superan ese valor no son beneficiarias, considerada una RD estricta. Sin embargo, en el gráfico de la derecha, la política se ofrece a aquellas empresas con menos de 25 trabajadores. Algunas de ellas deciden adherirse al programa y convertirse en beneficiarios de este, B, mientras que otras declinan, C. Aquellas con más de 25 trabajadores se las considera no elegibles para este programa dado que no cumplen la condición. Ahora estaremos ante una RD difusa.

Figura nº 1. REPRESENTACIÓN GRÁFICA DE RD ESTRICTA (IZQUIERDA) Y RD DIFUSA (DERECHA) EMPLEANDO INFORMACIÓN DE EMPRESAS



Nota: número de empleados (Z), recibir programa (B: beneficiario, C: control), nivel de ventas (Y).

Fuente: Elaboración propia.

Dado que en la evaluación de impacto se comparan individuos tratados y no tratados que tienen características similares, utilizar este diseño de forma correcta implica considerar para el análisis, exclusivamente, individuos alrededor de la frontera 25, donde existen empresas, más o menos parecidas en cuanto a número de trabajadores (entre 24 y 26 trabajadores), donde algunas han recibido ayuda, mientras que otras no. En ese área se puede realizar una evaluación de impacto porque se aísla el efecto de la intervención pública analizada (Hahn, Todd y Van der Klaauw, 2001).

En el caso de tener una regresión en discontinuidad estricta, el impacto del programa se puede estimar mediante una aproximación estadística o econométrica. En la aproximación estadística la evaluación realiza un contraste de diferencia de medias de la variable de resultados Y, comparando los individuos que están justo debajo  $Z^*$ , dado por  $E(Y|Z^-)$  y encima del valor  $Z^*$ ,  $E(Y|Z^+)$ , calculando la diferencia de medias de la variable de interés alrededor de la frontera.

80

$$\alpha = E(Y|Z^+) - E(Y|Z^-) \quad (21)$$

Por lo que respecta a la aproximación econométrica, la evaluación consiste en proponer la siguiente regresión:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 Z_i + U_i \quad (22)$$

donde la variable endógena es la variable de resultados Y, y hay dos regresores, la variable continua Z, y D, recibir programa, que viene dado por (19). Entonces, calculando el valor esperado de Y para cada uno de los dos grupos ( $D=1$  y  $D=0$ ), el impacto del programa es:

$$E(Y|D=1) - E(Y|D=0) = \frac{\beta_0 + \beta_1 + \beta_2}{E(Y|D=1)} - \frac{\beta_0 - \beta_2}{E(Y|D=0)} = \beta_1$$

El impacto del programa se localiza en el parámetro  $\beta_1$ , asociado al regresor ser beneficiario del programa D.

### 8.1. Ventajas, limitaciones y potenciales riesgos

La regresión discontinua puede ser un método útil cuando existe una regla clara de elegibilidad para ser beneficiario de un programa en función de una característica continua, situación muy habitual en nuestro entorno. Sin embargo, también presenta ciertas limitaciones. Por ejemplo, al ser una evaluación que se realiza alrededor de una frontera, se pueden dar problemas de potencia estadística. También se pueden dar situaciones en las que los individuos pueden manipular su elegibilidad en el programa, decidiendo en qué lado de la frontera situarse, y sesgando de este modo las estimaciones.

Otro riesgo en la aplicación de esta metodología está en los problemas asociados a la detección correcta de la forma funcional existente entre la característica continua Z y el resultado Y. Además, esta aproximación sólo permite estimar el impacto de la intervención para los individuos cercanos al punto de corte, mientras que el efecto de la intervención puede ser diferente para los individuos más alejados de este umbral.

## 9. CONCLUSIONES

La evaluación de impacto (EI) se presenta como una herramienta para la mejora de las intervenciones públicas, focalizada en la generación de evidencia para una

mejor toma de decisiones, con análisis que están orientados a resultados. En este trabajo se presenta la importancia de este tipo de evaluaciones, mostrando los diferentes tipos de preguntas que el evaluador se puede hacer cuando realiza este tipo de evaluación. A continuación, se describen los principales elementos de la EI, como el contrafactual y sesgo de selección. Las principales metodologías de evaluación existentes se diferencian en función de si la asignación de los individuos al grupo de tratamiento y control se realiza de manera aleatoria, denominadas «diseños experimentales», en los que el sesgo de selección es cero; o no hay asignación al azar, llamados «diseños cuasiexperimentales», donde será necesario corregir ese sesgo mediante la utilización de técnicas cuantitativas. En las siguientes secciones de este trabajo se ha presentado, de forma resumida, el contenido de estos métodos, prestando especial atención a los supuestos necesarios para su aplicación, necesidades de información y limitaciones y riesgos en su aplicación.

## REFERENCIAS BIBLIOGRAFICAS

- ANGRIST, J.D.; KRUEGER, A. (2001): «Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments», *The Journal of Economic Perspectives*, 15 (4): 69-85.
- ARTES, J.; RODRIGUEZ-SANCHEZ, B. (2022): «Métodos de evaluación de políticas públicas. Evaluación de Políticas Públicas». *Papeles de Economía Española*. 172: 18-29.
- BENARJEE, A.; M.; DUFLO, E. (2017); *Handbook of Field Experiments*. Vol2. Handbooks in Economics. North Holland. Amsterdam. Holanda.
- BARNOW, B.S.; CAIN, G.G.; GOLDBERGER, A. (1981): «Selection on Observables», *Evaluation Studies Review Annual*, 5: 43 - 59.
- BERGER, V.W. (2005): *Selection Bias and covariate imbalances in randomized clinical trials*, N.J. Wiley, Hoboken.
- BERTRAND, M.; DUFLO, E.; MULLAINATHAN, S. (2004): «How Much Should We Trust Differences-in-Differences Estimates?», *The Quarterly Journal of Economics*, 119(1): 249-275.
- CALIENDO, M.; KOPEINIG, S. (2008): «Some Practical Guidance for the Implementation of Propensity Score Matching», *Journal of Economic Surveys*, 22(1): 31-72.
- CATTANEO, M.; IDROBO, N.; TITIUNIK, R. (2019): *A practical introduction to regression discontinuity designs: foundations*. Cambridge University press. Cambridge.
- COMISION EUROPEA (2013): *Evalsed Sourcebook: Method and Techniques*. Comision Europea. Bruselas.
- CUNNINGHAM, S. (2021): *Causal inference. The mixtape*. Yale University Press. New Haven.
- DEHEJIA, R.; WAHBA, S. (2002): «Propensity Score-Matching Methods For Nonexperimental Causal Studies», *The Review of Economics and Statistics*, 84(1): 151-161.
- FEDEA (2021): «La evaluación de políticas públicas en España: antecedentes, situación actual y propuestas para una reforma». FEDEA Policy Papers, 2021/09. [https://documentos.fedea.net/pubs/fpp/2021/10/FPP2021-09.pdf?utm\\_source=wordpress &utm\\_medium=actualidad&utm\\_campaign=estudio](https://documentos.fedea.net/pubs/fpp/2021/10/FPP2021-09.pdf?utm_source=wordpress &utm_medium=actualidad&utm_campaign=estudio)
- FEINSTEIN, O. (2007): «Evaluación pragmática de políticas públicas» *Información Comercial Española, ICE: Revista de economía*, 836: 19-32.
- (2012): «Evaluación y gestión pública innovadora», *Ekonomiaz*, 80: 138-155.
- FEINSTEIN, O.; ZAPICO, E. (2010): «Evaluation of Government Performance and Public Policies in Spain» IEG World Bank ECD Working Papers, No.22. <https://openknowledge.worldbank.org/handle/10986/27913>
- FRÖLICH, M.; SPERLICH, S. (2019): *Impact evaluation, treatment effects and Causal analysis*. Cambridge University Press. Cambridge.
- GERTLER, P.J.; MARTINEZ, S.; PREMAND, P.; RAWLINGS L.B.; VERMEERSCH, C.M. (2016): *Impact evaluation in practice. 2nd edition*. Banco Mundial. Whasington D.C.
- GLENNERSTER, R.; TAKAVARASHA, K. (2013): *Running Randomized Evaluations: A Practical Guide*. Princeton University Press. Princeton, New Jersey.
- GUO, S.; FRASER, M.W. (2019): *Propensity score analysis. Statistical methods and applications*. SAGE publications. Londres.
- HAHN, J.; TODD, P.; VAN DER KLAUW, W. (2001): «Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design», *Econometrica*, 69(1): 201-209.
- HECKMAN, J. (1997): «Instrumental variables, a study of implicit behavioral assumptions used in making program evaluations». *Journal of human resources*. 32 (3): 441-462.
- IMBENS, G.W.; RUBIN, R. (2015): *Causal Inference in Statistics, Social and Biomedical Sciences*. Cambridge University Press. Cambridge.
- JIMENO, J.F. (ED.) (2021): *La evaluación de las políticas públicas en España: recursos, metodologías y algunos ejemplos*. Número monográfico de Cuadernos Económicos (ICE).
- KHANDKER, S.; KOOLWAL, G.; SAMAD, H. (2010): *Handbook in impact evaluation: Quantitative methods and practices*. Banco Mundial, Washington D.C.
- LEEUW, F.; VAESSEN, J. (2009): «Impact evaluations and development. NONIE guidance on impact evaluation». Working paper. Banco Mundial. Washington, D.C. <https://www.re>

- searchgate.net/profile/Frans-Leeuw/publication/264869403\_Impact\_Evaluations\_and\_Development\_NONIE\_Guidance\_on\_Impact\_Evaluation/links/55aecb8308ae98\_e661a6f1c1/Impact-Evaluations-and-Development-NONIE-Guidance-on-Impact-Evaluation.pdf
- OECD (2006): *Outline of principles of impact evaluation*. OECD publishing. Paris.
- ONRUBIA, J.; SÁNCHEZ-FUENTES, A.J. (2019): «Evaluación de políticas públicas: una tarea urgente e imprescindible», en A. Cuenca y S. Lago (coord.), *El sector público español: reformas pendientes*, 211-241. FUNCAS. Madrid.
- PLANAS, I. (2005): «Principales mecanismos de evaluación económica de políticas públicas». *EconomiaZ*, 60: 98-121.
- ROGERS, P.J. (2012): «Introduction to impact evaluation». Inter Action y Rockefeller Foundation. Impact evaluation notes. <https://www.interaction.org/wp-content/uploads/2019/03/1-Introduction-to-Impact-Evaluation.pdf>
- ROSENBAUM, P.; RUBIN, D. (1983): «The Central Role of the Propensity Score in Observational Studies for Causal Effects», *Biometrika*, 70(1): 41-50.
- STERN, E. (2015): «Impact Evaluation: a guide for commissioners and managers, departments for international development». Bond International development.
- TODD, P. (2007): «Evaluating Social Programs with Endogenous Program Placement and Selection of the Treated», en *Handbook of Development Economics*, vol. 4, Paul y Strauss eds, Noth-Holland Pub. 3847-3894. Ámsterdam.
- UNEG (2020) «Compendium of evaluation methods reviewed – Volume 1». Grupo de evaluación de Naciones Unidas. Documento de trabajo. <http://www.unevaluation.org/document/detail/2939>
- ZAPICO, E. (2022): *Presupuesto y evaluación del gasto público: De la competición a la colaboración*. Instituto de estudios Fiscales. Madrid.