

# Evaluación de políticas públicas sobre poblaciones heterogéneas. ¿Pueden los órganos de control externo contribuir a su avance?

---

José Manuel Cansino Muñoz-Repiso\*  
Rocío Román Collado\*\*

## RESUMEN

A partir de las aportaciones de Heckman *et al.* (1997), Manski (2001) e Imbens (2004), así como Durán (2004) en la literatura española, este artículo aborda la cuestión de si los órganos de control externo pueden contribuir eficazmente al desarrollo de la investigación evaluatoria en poblaciones heterogéneas, particularmente cuando aquélla se basa en la aplicación de métodos de inferencia causal estadística.

El artículo sostiene que cuando las características que motivan las diferentes respuestas de los individuos a los tratamientos resultan observables al investigador, el control de las mismas mediante el método de *selección sobre observables* resulta adecuado.

La evidencia encontrada en estos métodos evaluatorios debe, no obstante, contrastarse con el arrojado por otros estimadores, particularmente en el caso de existir características no observables. El estimador de diferencias en diferencias puede resultar, en este caso, de ayuda.

\* Universidad de Sevilla. Dirección para comentarios: [jmcansino@us.es](mailto:jmcansino@us.es)

\*\* Universidad de Sevilla.

Los órganos de control externo españoles y, de manera especial, el Tribunal de Cuentas cuentan con las habilitaciones legales necesarias para poder recopilar el soporte empírico necesario para este tipo de evaluaciones. La implementación de su tratamiento se beneficiaría de los paquetes informáticos disponibles.

Sobre los resultados de los análisis evaluatorios en poblaciones heterogéneas, los decisores públicos podrían desarrollar reglas de asignación que permitan –bien sobre un criterio de eficiencia, bien sobre un criterio de equidad– optimizar los efectos de los tratamientos o políticas públicas.

## 1. INTRODUCCIÓN<sup>1</sup>

Aunque la evaluación económica de políticas públicas cuenta con una notable tradición a lo largo del siglo XX, recientemente y de manera progresiva sus resultados son tomados en consideración con el fin de mejorar la calidad de un número importante de administraciones.

En un sector económico como el público, donde el criterio de éxito para las acciones es más difuso que en el privado, las investigaciones evaluatorias a menudo encuentran evidencias que, de ser tomadas en consideración por los decisores públicos, las convertirían en un instrumento potencialmente muy útil para la buena gestión de los recursos públicos.

Uno de los desarrollos recientes en las investigaciones sobre evaluación económica de políticas públicas se ubica en la búsqueda de respuestas al problema que supone el que los individuos respondan de forma desigual a una misma política pública –Manski (2001)–.

Pacientes que reaccionan desigualmente a un mismo tratamiento, condenados con actitudes de reinserción diferentes a pesar de recibir una misma sentencia judicial, o desempleados con distinta vida laboral tras recibir un mismo programa de formación, son algunos ejemplos de estas respuestas heterogéneas.

Dado que la misma política puede alcanzar resultados dispares según los individuos beneficiarios, su efectividad depende no solo de su propio diseño, sino también de la adecuada asignación de partici-

---

<sup>1</sup> Los autores agradecen a Antonio SÁNCHEZ (Universidad de Sevilla) los comentarios a una versión preliminar de este artículo. Los posibles errores son, no obstante, responsabilidad de los autores.

pantes a las diferentes políticas alternativas. De esta forma, la mejora en el proceso de asignación supondría una mejora en la efectividad.

El decisor público tiene entonces que elegir la política pública óptima para cada grupo de población según el criterio de elección que aplique y, para esa decisión, sería muy útil conocer los resultados de los métodos evaluatorios.

Aunque en España la legislación fundamental impide las prácticas discriminatorias en el diseño de las políticas públicas, es frecuente que en la definición de los grupos beneficiarios de las mismas se discrimine según características personales. El nivel de ingresos de los destinatarios, educación, grado de minusvalía, etc., son solo algunos ejemplos de definición de beneficiarios específicos de las políticas públicas.

A partir de las aportaciones de Heckman *et al.* (1997), Manski (2001) e Imbens (2004), así como Durán (2004) en la literatura española, este artículo aborda la cuestión de si los órganos de control externo pueden contribuir eficazmente a esta línea de desarrollo de la investigación evaluatoria en España, particularmente cuando ésta se basa en la aplicación de métodos de inferencia causal estadística, y la población beneficiaria es heterogénea.

El artículo está estructurado en los siguientes apartados. Tras la introducción, que también especifica el objetivo del artículo, en el apartado 2 se justifica el uso del término «tratamientos» para referirnos a las políticas públicas. En el apartado 3 se subrayan las limitaciones de las evaluaciones basadas en la estimación de los efectos medios del tratamiento cuando se aplican a poblaciones heterogéneas. El apartado 4 está destinado a definir y justificar el uso del método de «selección sobre observables». El apartado 5 está dedicado a abordar la asignación de tratamientos rivales en una población heterogénea; mientras que en el apartado 6 se hace una valoración de las posibilidades reales que tiene esta línea de desarrollo de la evaluación de políticas públicas, ubicando la cuestión en el papel que pueden desempeñar los órganos de control externo y, de manera especial, el Tribunal de Cuentas. Las conclusiones se resumen al final en el apartado 7.

## 2. LAS POLÍTICAS PÚBLICAS COMO TRATAMIENTO

En la literatura anglosajona especializada en evaluación de políticas públicas, a menudo se denomina a éstas «tratamientos». La explicación del uso generalizado del término probablemente se encuentre en la pionera aplicación del análisis evaluatorio a los resultados de tratamientos médicos sobre una población de pacientes. Trabajos se-

minales en la línea anterior, basados en el uso de la inferencia causal estadística, fueron los de Billewicz (1965), Cochran (1968) y Cameron y Pauling (1976). En este artículo también nos centramos en la evaluación de políticas públicas a partir de la aplicación de métodos de inferencia causal estadística.

Con la aplicación de la inferencia causal estadística, la efectividad de la política pública o tratamiento (término que utilizaremos en adelante) se determina según el efecto que provoca (efecto causal) sobre la población tratada. Para un individuo cualquiera de la población estudiada este efecto se mide como la diferencia entre el valor alcanzado por la variable respuesta –*outcome*– tras ser expuesto al tratamiento y el valor del resultado contrafactual de la misma variable, en la situación hipotética de que el individuo no hubiese participado en el tratamiento. Dado que el resultado contrafactual no puede ser observado por el investigador, resulta necesario proceder a su estimación. Para ello se recurre a métodos econométricos y estadísticos.

Tras el estudio<sup>2</sup> de Lalonde (1986), bien sobre experimentos sociales de tipo aleatorio o a partir de métodos cuasi-experimentales, los individuos son asignados a un grupo de tratamiento o a un grupo de control, lo que determina dos estatus potenciales para cualquier individuo; el de tratado (cuando pertenece al primer grupo y se expone de manera efectiva al tratamiento) y el de no tratado (cuando pertenece al segundo, manteniéndose al margen del tratamiento).

### **3. LIMITACIONES DE LAS EVALUACIONES BASADAS EN EL EFECTO MEDIO DEL TRATAMIENTO. RESPUESTAS UNIFORMES VERSUS HETEROGÉNEAS**

Si todos los individuos respondieran a un tratamiento de la misma manera, la decisión política óptima sería tratar a todas las personas uniformemente con la mejor política pública o tratamiento disponible. La bondad del tratamiento sería una bondad relativa establecida tras comparar sus resultados con los que arrojan políticas alternativas aplicadas sobre la misma población.

El supuesto de la respuesta homogénea de la población a un mismo tratamiento ha sido habitual en los métodos evaluatorios. Par-

---

<sup>2</sup> El trabajo de LALONDE (1986) desarrollaba una evaluación del efecto medio de unos programas públicos de formación. Este tipo de políticas públicas ha sido objeto de numerosos estudios evaluatorios basados en los métodos de inferencia causal estadística.

tualmente para el caso de los métodos cuasi-experimentales<sup>3</sup>, la práctica común de los investigadores ha sido asumir que todas las personas responden a un mismo tratamiento de manera similar. Consecuencia de lo anterior es que la mayoría de las evaluaciones basadas en el uso de la inferencia causal estadística se centren exclusivamente en la estimación del efecto medio que provoca un determinado tratamiento sobre la población, analizando cómo el tratamiento impacta sobre una variable que se considera relevante para el propósito evaluatorio y que habitualmente se denomina en la literatura evaluatoria variable respuesta (*outcome*). Este efecto medio puede estimarse para el conjunto de la población (PATE) o solo para el grupo de individuos seleccionados para exponerse al tratamiento<sup>4</sup> (PATT).

El criterio de elegir el efecto medio de un tratamiento sobre la población como aquel sobre el que se basa la evaluación se ve respaldado, a menudo y además, por dos supuestos:

- a) Aumentos en el *outcome* total aumentan el bienestar social<sup>5</sup>.
- b) Los aspectos no deseables del tratamiento son bien aspectos sin importancia, o bien se compensan mediante transferencias del gobierno a través de una función de bienestar social.

---

<sup>3</sup> En CANSINO y SÁNCHEZ (2004) argumentamos que en el ámbito de las ciencias sociales los experimentos aleatorios no resultan habitualmente factibles pese a las ventajas técnicas que facilitan a las investigaciones evaluatorias. Consecuentemente, los investigadores deben buscar alternativas para obtener el soporte empírico necesario para realizar la evaluación. Esto es lo que pretenden los métodos cuasi-experimentales tratando de reproducir escenarios experimentales. Estos métodos aplicados para la determinación de la inferencia causal permiten realizar la evaluación de los efectos de un tratamiento, teniendo en cuenta nuestro conocimiento institucional del problema. Asimismo, presentan la ventaja de permitir la consideración de muestras de mayor tamaño frente a los experimentos aleatorios, que trabajan con muestras relativamente pequeñas, y admiten también la posibilidad de aplicar las nuevas técnicas y los nuevos estimadores que se han ido desarrollando en el ámbito no experimental; así lo argumentan HECKMAN y SMITH (1995) en su réplica al trabajo de LALONDE (1986).

<sup>4</sup> En la revisión de IMBENS (2004) se resumen los tipos de efectos medios del tratamiento (ATE en acrónimo anglosajón) que han sido considerados por la literatura. Así, se refiere en primer lugar al PATE como el efecto medio que el tratamiento evaluado causa sobre la población, y al PATT como el efecto medio del tratamiento cuando solo se evalúa el efecto que provoca sobre la población que ha sido expuesta al tratamiento. En segundo lugar, el SATE representaría el efecto medio del tratamiento cuando para la evaluación se toma una muestra de la población, y el SATT cuando la muestra es de solo la parte de la población efectivamente tratada. Finalmente, el CATT estimaría el efecto medio del tratamiento condicionado a la distribución de covariables, y el CATE haría lo propio, aunque solo considerando la población que ha se ha beneficiado del tratamiento.

<sup>5</sup> Así, por ejemplo, un aumento en el nivel salarial, educativo o de la salubridad tendría un efecto positivo sobre el bienestar social.

Sin embargo, es habitual que las respuestas a un mismo tratamiento varíen entre la población. Con frecuencia, en el origen de esa variación deben considerarse determinadas características observables de las personas tratadas, como, por ejemplo, su nivel educativo, experiencia profesional, edad, ingresos, etc. Esta falta de homogeneidad es la responsable, habitualmente, del sesgo de selección muestral en aquellas evaluaciones que se realizan a partir de muestras<sup>6</sup>. La situación es diferente cuando estamos ante características personales como la motivación, la habilidad o la capacidad física o intelectual; todas ellas son características no directamente observables para el investigador y que pueden introducir sesgo de selección muestral.

Lo anterior permite concluir que, para la investigación evaluatoria, la media no responde a muchas cuestiones interesantes si las personas responden de manera diferente a un mismo tratamiento.

Si no tuviéramos esto en cuenta podría darse el resultado de eliminar del menú del decisor público un determinado programa, porque la investigación evaluatoria haya encontrado evidencia de que su efecto medio es negativo o, simplemente, no significativo. Sin embargo, los resultados hubiesen sido diferentes si a este tratamiento hubieran sido expuestos solo aquellos individuos con las características adecuadas, mientras que el resto de la población se hubiese beneficiado de tratamientos alternativos.

En definitiva, existen cuestiones relevantes sobre el análisis evaluatorio relacionadas con la distribución de las ganancias causadas por un tratamiento.

Así, Heckman *et al.* (1997: 490) subrayan que para el decisor público pueden resultar relevantes, entre otras cuestiones, las siguientes:

a) La parte de la población que, siguiendo un determinado tratamiento, se beneficia de él<sup>7</sup>.

---

<sup>6</sup> Para el caso de las políticas activas de empleo, DURÁN (2004: 108) ejemplifica el problema cuando escribe que si aquellos individuos que muestran generalmente unos mejores resultados en su búsqueda de empleo tienen también una mayor probabilidad de participar en el programa, una comparación de los resultados post-programa entre el grupo de participantes y de control tenderá a sobreestimar el efecto del programa.

<sup>7</sup> Así, en un sistema democrático, el electorado no tiene por qué mostrar la misma preferencia ante dos tratamientos cuyo efecto medio sea similar, pero con resultados favorables para distintas subpoblaciones. La idea de equidad más generalmente asentada permitiría valorar positivamente un tratamiento muy efectivo para colectivos vulnerables, a pesar de que el efecto medio del mismo tratamiento sobre el conjunto de la población no sea significativo. Esto debe incorporarse en la agenda del decisor público.

b) La proporción del total de la población que se beneficia del tratamiento<sup>8</sup>.

c) Determinados cuantiles de la distribución de impactos<sup>9</sup>.

d) La distribución de ganancias entre unos valores de estado básico seleccionados.

Desde el punto de vista de la conformación de la acción colectiva, algunas personas podrían juzgar como exitosos tratamientos de los que solo determinados ciudadanos o colectivos reciben beneficios, incluso aunque el efecto medio del tratamiento sea no significativo o, en un caso extremo, reduzca el bienestar social<sup>10</sup>.

En el caso concreto de los programas públicos de formación, a la luz de los resultados obtenidos por Heckman, Smith y Clements (1997) sería posible incrementar los ingresos laborales de los beneficiarios mediante la asignación de unos beneficiarios a la formación profesional ocupacional, y a otros a la formación básica<sup>11</sup>.

En un escenario con población heterogénea, el bienestar social aumentaría si las políticas públicas se diseñaran de manera que el tratamiento asignado a cada individuo variase adecuadamente en función de esas características. Cada persona o grupo de personas recibiría el tratamiento con mayores posibilidades de éxito dadas sus características.

En otros términos, la existencia de programas alternativos confiere distintas opciones al decisor público y, para éste, es de interés establecer el valor de las opciones, tarea para la que recurre al análisis evaluatorio de sus efectos.

La posibilidad de que varíen las respuestas de los individuos a un mismo tratamiento hace que la investigación en la evaluación de programas pueda ayudar a informar a la política pública a través de los

---

<sup>8</sup> HECKMAN *et al.* (1997: 490) señalan que si los votantes tienen expectativas correctas sobre la distribución conjunta de resultados, resulta de interés para los políticos y para los estudiosos de la política económica determinar cómo de ampliamente se distribuyen los beneficios de un tratamiento. Relacionado también con la definición de la acción colectiva, si la mayor parte de los beneficios de un tratamiento los reciben unas pocas personas se hace más fácil organizar grupos de interés en apoyo del tratamiento que si las mismas ganancias se distribuyen más ampliamente.

<sup>9</sup> *Vid.* LEHMAN (1974) y DOCKSUM (1974).

<sup>10</sup> Éste puede ser el caso de tratamientos o programas que, teniendo un efecto medio no significativo, sí resultan eficaces para colectivos concretos, como personas marginadas, con adicciones, etc.

<sup>11</sup> En este sentido, HECKMAN, SMITH y CLEMENTS (1997) rechazan la hipótesis de que todos los tratados por un programa de formación en EE.UU. mismo resultado. Cfr. DURÁN (2004).

esfuerzos en aprender las consecuencias de reglas alternativas de tratamiento. De manera más precisa, la evaluación de las políticas públicas debería incorporar en su agenda la investigación de cómo las respuestas a un mismo tratamiento varían entre la población. En todo este planteamiento, *la llave del éxito consiste en determinar qué personas deben recibir qué tratamientos*<sup>12</sup>.

El resto es asignar óptimamente tratamientos rivales a las subpoblaciones relevantes.

En línea con lo anterior, el interés y utilidad de la investigación evaluatoria mejoraría si se prestase mayor atención a la variación observada en la respuesta de los tratados al tratamiento.

## 4. EL MÉTODO DE SELECCIÓN SOBRE OBSERVABLES

### 4.1. Las características observables

En buena medida, desde Lalonde (1986), la evaluación económica de políticas públicas a partir de la aplicación de métodos de inferencia causal estadística se desarrolla sobre experimentos en los que los individuos susceptibles de exponerse al tratamiento son asignados a un grupo de tratamiento o a un grupo de control, lo que determina los dos mencionados estatus potenciales para cualquier individuo: el de tratado y el de no tratado.

En los métodos cuasi-experimentales puede esperarse que los individuos beneficiarios y los pertenecientes al grupo de control difieran en características personales, distintas de las propias del tratamiento y que afectan a la variable respuesta o *outcome*. No obstante, si ambos tipos de individuos (tratados y de control) difieren solo en las características observables, es posible controlar estas diferencias a través de lo que se denomina *selección sobre observables* con el objetivo de desarrollar una investigación evaluatoria más robusta.

El método de selección sobre observables es un método de evaluación que se aplica a los métodos cuasi-experimentales<sup>13</sup> y que permite evaluar los tratamientos sobre individuos homogéneos en determinadas características que resultan observables para el investigador. Avanzando lo que luego se expondrá, esto permite, por ejemplo, definir grupos de pacientes con igual edad, sexo, grado de evolución de la patología, etc. En cada una de esas características

<sup>12</sup> Cf. MANSKI (2001: 103).

<sup>13</sup> Cf. HECKMAN *et al.* (1997: 488).



observables o para varias, los grupos de individuos resultan homogéneos.

El método de selección sobre observables permite abordar el problema de las respuestas heterogéneas a un mismo tratamiento cuando éstas están influidas por características observables de la población. En esencia, el método permite controlar la selección por estas características otorgando el estatus de tratado o no tratado, esto es, asignando a individuos al grupo de tratamiento o al de control que son homogéneos en alguna o algunas características observables.

Sin embargo, la selección sobre observables no resuelve el problema de la heterogeneidad de la población cuando ésta tiene su origen en características no observables de los individuos. En este sentido resultaría oportuno comparar los resultados con los que pudiera arrojar un estimador como el de diferencias en diferencias.

La literatura especializada establece que este control de los individuos por características observables puede realizarse aplicando tres métodos: el método de la subclasificación, el método de *matching*<sup>14</sup> y el método del *propensity score*<sup>15</sup>.

Para el propósito de este artículo mostramos las posibilidades que ofrece el método de la subclasificación por su operatividad para los órganos de control. No obstante, debe la literatura cuenta por una producción reciente<sup>16</sup> e importante en el desarrollo de los métodos de *matching* y de *propensity score*.

## 4.2. Consideraciones preliminares

Realicemos, siguiendo a Imbens (2004), algunas consideraciones preliminares, particularmente sobre la nomenclatura relevante. El origen de esta nomenclatura se encuentra<sup>17</sup> en las aportaciones ini-

---

<sup>14</sup> Para una revisión teórica de los estimadores *matching*, véanse los trabajos de HECKMAN, ICHIMURA, TODD (1997 y 1998) e IMBENS (2004), así como el desarrollo práctico planteado por ANGRIST (1998).

<sup>15</sup> Un desarrollo amplio y reciente del *propensity score* puede encontrarse en HIRANO, IMBENS y RIDDER (2003).

<sup>16</sup> *Vid.* IMBENS (2004).

<sup>17</sup> DURÁN (2004: 110) se refiere a este desarrollo de la literatura sobre la evaluación econométrica de políticas públicas como el modelo de resultados potenciales. Para la evaluación específica de programas de formación, DURÁN (2004) se refiere al modelo de resultados potenciales como Modelo de Roy-Rubin, añadiendo a las referencias que hace IMBENS (2004) en su revisión las de ROY (1951) y HECKMAN y VYTLACIL (2000).

ciales de Fisher (1935) y Neyman (1923 y 1935), luego generalizadas, fundamentalmente, por Rubin (1974, 1977 y 1978).

Conocer el efecto causal de un tratamiento exige, inicialmente, definir un indicador de tratamiento  $D_i$  en forma de variable binaria para cualquier individuo  $i$ -ésimo de los que, potencialmente, pueden beneficiarse del mismo.

Así,  $D_i = 1$  indicará que el individuo  $i$  ha participado en el tratamiento, y  $D_i = 0$  indicará que no lo ha hecho.

Definido el indicador  $D_i$ , los métodos de inferencia causal utilizan la noción de *respuestas potenciales* de los tratados para definir la(s) variable(s) respuesta(s) –*outcome* (s)– de cada individuo al tratamiento.

De esta forma, si consideramos una única variable resultado  $Y_i$ , que representa la respuesta del individuo  $i$ -ésimo (reacción a una medicación, actitud de reinserción tras el cumplimiento de una condena, ingresos laborales obtenidos tras recibir un programa de formación, etc.), las respuestas potenciales de los individuos pueden definirse como:

$Y_{0i}$  = Valor de la variable respuesta en el caso de que el individuo  $i$ -ésimo no se beneficie del tratamiento.

$Y_{1i}$  = Valor de la variable respuesta del individuo  $i$ -ésimo en caso de beneficiarse de la política pública o tratamiento.

Definidas las respuestas potenciales, el efecto causal del tratamiento sobre la respuesta potencial del individuo  $i$ -ésimo vendría determinado por la diferencia « $Y_{1i} - Y_{0i}$ », a partir de cuya magnitud podría evaluarse la efectividad del tratamiento.

Naturalmente, no podemos observar simultáneamente  $Y_{0i}$  e  $Y_{1i}$ , ya que se trata de sucesos contrafactuales. Para cada individuo solo puede observarse la respuesta realizada  $Y_i$  definida como<sup>18</sup>:

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i} \quad (1)$$

Este problema, conocido como el problema fundamental de identificación en los métodos de inferencia causal, impide calcular los efectos causales individuales del tratamiento.

Lo anterior obliga a buscar soluciones de *second best* que podrían derivarse de la búsqueda de efectos causales agregados. Así, un elevado grado de homogeneidad entre los beneficiarios potenciales del

<sup>18</sup> Véase ANGRIST e IMBENS (1991).

tratamiento que implicase que los mismos tuviesen idénticas respuestas potenciales permitiría resolver el problema. De esta forma, para algunos individuos «j» (pertenecientes al grupo de tratados) observaríamos « $Y_{1j}$ », mientras que para otros individuos «k» –idénticos a los anteriores pero pertenecientes al grupo de control– observaríamos « $Y_{0k}$ ».

La homogeneidad permitiría calcular el efecto causal del tratamiento para cada individuo «i» ( $Y_{1i}$ - $Y_{0i}$ ) mediante la comparación del valor de las variables respuesta entre un individuo que ha recibido el tratamiento y otro que no lo ha hecho.

No obstante, el elevado grado de homogeneidad en las respuestas es una característica habitual en experimentos realizados en disciplinas científicas como la Física o la Química. Sin embargo, en el ámbito de los experimentos sociales suele darse un amplio grado de heterogeneidad en las respuestas individuales a la participación (y no participación) en programas públicos. Abordar el problema que presentan las poblaciones heterogéneas es el tema central del presente artículo.

Por tanto, aunque no es posible calcular los efectos causales individuales de los tratamientos, puede resultar factible el cálculo de algún efecto promedio. En este sentido, teniendo en cuenta que la variable respuesta  $Y$ , para el individuo  $i$ -ésimo, queda definida como:

$$Y_i = \begin{cases} Y_{1i} & \text{si } D_i = 1 \\ Y_{0i} & \text{si } D_i = 0 \end{cases} \quad (2)$$

podemos definir, entonces, el efecto causal medio del tratamiento sobre la población (PATE) como:

$$\alpha_{\text{PATE}} = E[Y_1 - Y_0] \quad (3)$$

Alternativamente, es posible estimar el efecto medio únicamente sobre los individuos que han recibido el tratamiento (PATT), esto es, los individuos tratados. En este caso:

$$\alpha_{\text{PATT}} = E[Y_1 - Y_0 | D = 1] \quad (4)$$

Ahora, la estimación del efecto medio se restringe únicamente a los individuos seleccionados<sup>19</sup> para exponerse al tratamiento.

<sup>19</sup> Para algunas políticas públicas, como los programas de formación, el estimador  $\alpha_{\text{PATT}}$  puede contener algún error, como ha señalado ASHENFELTER (1978).

### 4.3. Formalización del método de la subclasificación

Como se ha señalado, la ausencia de homogeneidad en la población obliga a buscar soluciones que no invaliden los resultados de la investigación evaluatoria. Esto es lo que se pretende con el método de la subclasificación. Formalmente, el método se expone de la siguiente forma.

Para la evaluación basada en la inferencia causal, resulta esencial el cumplimiento del supuesto de independencia. Este supuesto exigiría que beneficiarse o no de un tratamiento no estuviese relacionado con las respuestas potenciales de los individuos y quedaría formulado como

$$(Y_1, Y_0) \perp D \quad (5)$$

Sin embargo, en ausencia de datos propios de un experimento social aleatorizado, la condición de independencia recogida en la Expresión 5 es difícilmente plausible. Ésta es la situación en la que se encuentran los métodos cuasi-experimentales.

Para este tipo de escenarios introducimos el uso de covariables. Decimos que una variable  $X$  está predeterminada con respecto al tratamiento,  $D$ , si para cada individuo  $i$ ,  $X_{1i} = X_{0i}$ . Esto es, el valor de  $X_i$  no depende del valor de  $D_i$ . En la medida en que estas características ajenas al tratamiento puedan ser observadas y que los individuos beneficiarios y los del grupo de control presenten diferencias solo en estas características que pueden ser observadas se podrán controlar tales diferencias. Sobre esta base se fundamenta el método de *selección sobre observables*<sup>20</sup>.

Frente a este tipo de variables, seguiremos llamando variables respuesta a cualquier variable  $Y$  cuyo valor puede depender del valor de  $D_i$ . Para las variables respuesta y para cualquier individuo  $i$ ,  $Y_{1i} \neq Y_{0i}$ .

Con frecuencia, las variables predeterminadas se denominan covariables<sup>21</sup>. No obstante, el hecho de que  $X$  esté predeterminada no implica que  $X$  y  $D$  sean independientes en la población<sup>22</sup>. En lo sucesivo,  $X$  representará normalmente un vector de covariables predeterminadas.

<sup>20</sup> Para mayor abundamiento nos remitimos a BARNOW, CAIN y GOLDBERGER (1980).

<sup>21</sup> Para una exposición más detallada sobre la naturaleza de las variables contaminantes o covariables, véase RUBIN (1978). Asimismo remitimos a una revisión posterior en IMBENS y RUBIN (1997).

<sup>22</sup> Por ejemplo, podemos estudiar los efectos de una titulación universitaria sobre los salarios. Es posible definir  $X =$  «raza» como una covariable (en el lenguaje estadístico, las covariables se denominan variables de confusión o contaminantes). Obviamente el título universitario no afecta a la raza. Sin embargo, la raza y los resultados académicos de los individuos no son independientes en la población.

Como ya hemos avanzado, de entre los métodos de evaluación aplicados a los métodos cuasi-experimentales nos referimos al método de *selección sobre variables observables* o *selección sobre observables*<sup>23</sup>.

Decimos que hay selección sobre observables cuando existe un vector de covariables predeterminadas observadas,  $X$ , tal que las respuestas potenciales son independientes del tratamiento condicionado por las covariables, cumpliendo así con la condición de independencia de la expresión 5, que ahora se formula como

$$(Y_1, Y_0) \perp D | X \quad (6)$$

Para un mismo valor de  $X$  tendríamos una especie de experimento aleatorio. La condición anterior permite comparar a individuos que pertenecen al grupo de tratados y al grupo de control para un mismo valor de  $X$ , como puede ser su educación escolar, experiencia profesional, edad o nivel de ingresos, etc.

Como exponen Heckman y Hotz (1989: 865), la selección sobre observables se da cuando la dependencia entre  $Y$  y  $D$  se debe a la covariable  $X$ , que influye en el proceso de selección de los individuos. En este caso, controlando la variable  $X$  se soluciona el problema del sesgo en la selección, esto es, se elimina la dependencia entre  $Y$  y  $D$ .

Bajo la selección sobre observables, el efecto causal del tratamiento puede expresarse, siguiendo a Dehejia y Wahba (1999: 1057), como:

$$\begin{aligned} E[Y_1 - Y_0 | X] &= E[Y_1 - Y_0 | X, D = 1] = \\ &= E[Y | X, D = 1] - E[Y | X, D = 0] \end{aligned} \quad (7)$$

Por tanto, el efecto causal medio del tratamiento (PATE) puede expresarse como:

$$\begin{aligned} \alpha_{\text{PATE}} &= E[Y_1 - Y_0] = \int E[Y_1 - Y_0 | X] dP(X) = \\ &= \int (E[Y | X, D = 1] - E[Y | X, D = 0]) dP(X) \end{aligned} \quad (8)$$

De manera similar, el efecto causal medio del tratamiento sobre los individuos tratados (PATT) resulta:

$$\begin{aligned} \alpha_{\text{PATT}} &= E[Y_1 - Y_0 | D = 1] = \int (E[Y | X, D = 1] - \\ &\quad - E[Y | X, D = 0]) dP(X | D = 1) \end{aligned} \quad (9)$$

<sup>23</sup> El concepto de *selección sobre observables* hace referencia al procedimiento mediante el cual se eligen las variables predeterminadas que, además, son observables. Este procedimiento se contrapone a los casos en los que las variables predeterminadas no son observables, aunque se tenga constancia de su existencia.

La expresión anterior muestra que el efecto causal medio del tratamiento sobre los tratados es igual a la diferencia entre los valores promedio observados de las respuestas potenciales de los tratados y de los individuos del grupo de control para cada uno de los valores posibles de  $X$ , cuando  $D = 1$ .

Realizada la selección sobre observables, desaparecen las diferencias sistemáticas previas al tratamiento entre el grupo de individuos tratados y el grupo de control<sup>24</sup>. Este resultado permitirá estimar los efectos causales del programa mediante la construcción de estimadores de  $\alpha_{\text{PATE}}$  y de  $\alpha_{\text{PATT}}$  a partir de las expresiones 7 y 8.

Identificada la variable o vector  $X$ , habrán de especificarse los valores que puede tomar. En función de los valores posibles de  $X$  se procederá a la subclasificación de la población diana. Lo anterior permite dividir los grupos de tratamiento y control en diferentes subclases.

Los valores que pueda tomar  $X$  conducen a dos situaciones:

- Si  $X$  es una variable discreta, existirá una subclase de individuos para cada valor que pueda tomar  $X$ . El número total de subclases es finito.
- Si  $X$  es una variable continua, se tendrá que construir un número finito de intervalos que incluyan todo el rango de valores posibles de  $X$ . Cada uno de estos intervalos constituirá una subclase de individuos.

La subclasificación de la población origina un conjunto  $(X^1, X^2, \dots, X^k)$  de  $k$  subclases de individuos<sup>25</sup>. Así, podemos denotar por  $n$  el número de individuos que participan en la evaluación del tratamiento, de los cuales  $n_1$  se benefician del programa ( $D = 1$ ) y los  $n_0$  restantes se incluyen en el grupo de control ( $D = 0$ ). Siendo  $k$  el número de subclases en que se ha dividido la población, habrá que determinar para cada una el número de individuos que la componen. De manera genérica tendremos:

$n_j$ : Número total de individuos que componen la subclase  $j$ .

$n_1^j$ : Número total de individuos tratados que pertenecen a la subclase  $j$ .

<sup>24</sup> Véase, en este sentido, RUBIN (1977).

<sup>25</sup> Ejemplos de subclasificaciones de la población diana pueden encontrarse en COCHRAN (1968), LALONDE (1986) y KRUEGER (1999).

$n_0^j$ : Número total de individuos que integran el grupo de control en la subclase  $j$ .

En cada momento se cumplirá que:

$$\begin{aligned} n_1^j + n_0^j &= n^j \\ n^1 + n^2 + \dots + n^k &= n \end{aligned}$$

$n^j/n$  será la ponderación atribuible a la subclase  $j$  sobre el total de la población, y  $n_1^j/n^j$  el peso que los individuos tratados tienen sobre el conjunto de individuos de cada subclase.

Al mismo tiempo  $\bar{Y}$  medirá la respuesta media esperada, distinguiendo entre  $\bar{Y}_1^j$ , que representa el valor medio de la variable respuesta para los tratados pertenecientes a la subclase  $j$ , y  $\bar{Y}_0^j$ , que representa las respuestas medias de los individuos del grupo de control pertenecientes a la subclase  $j$ .

Dentro de cada subclase, el grupo de tratados y de control estarán formados por individuos homogéneos en características observables. Es posible estimar los efectos causales  $\alpha_{\text{PATE}}$  y  $\alpha_{\text{PATT}}$  mediante la suma, para cada una de las subclases construidas, de las diferencias entre los valores medios observados en las variables respuesta de los tratados y de los individuos de control. Estas diferencias estarán ponderadas por el peso que corresponde a cada subclase. Analíticamente:

$$\alpha_{\text{PATE}} = \sum_{j=1}^K (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n^j}{n}\right) \quad (10)$$

$$\alpha_{\text{PATT}} = \sum_{j=1}^k (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n_1^j}{n_1}\right) \quad (11)$$

## 5. ASIGNACIÓN DE TRATAMIENTOS A LA LUZ DE LA VARIACIÓN DE LAS RESPUESTAS

### 5.1. Buscando una regla de asignación

El decisor público debe estar interesado en la variación cuantitativa de las respuestas según los tratamientos y las características de los individuos. En otros términos, contrastar simplemente la hipótesis nula del efecto medio del tratamiento no ubica correctamente la cuestión evaluatoria.

La estimación de  $\alpha_{\text{PATE}}$  y  $\alpha_{\text{PATT}}$  siguiendo el método de subclasificación permite al decisor político conocer cómo el resultado medio del tratamiento varía dependiendo de las covariables observadas en

la población. La selección sobre observables permite evaluar la efectividad del tratamiento sobre poblaciones heterogéneas en características observables que logran controlarse por el método de subclasificación, de manera que, finalmente, es posible evaluar la efectividad del tratamiento sobre subpoblaciones homogéneas en esa o esas características. El valor de los estimadores permitirá conocer el rango de variación de las respuestas a un mismo tratamiento.

Una vez que la investigación evaluatoria incorpora el análisis de la variación de las respuestas entre la población y si éste encuentra evidencia significativa, Manski (2001) concluye que el planificador social debe entonces solucionar el problema de elegir un tratamiento específico para cada miembro de la población (o subpoblación similar en sus características). En este estadio es factible asignar cada individuo o colectivo al mejor tratamiento disponible.

Con un planteamiento optimizador concreto, la elección del planificador podría consistir en elegir una regla de asignación de tratamientos que optimice la respuesta media de la población al tratamiento, siguiendo para ello un criterio de eficiencia económica. Para Durán (2004) debe precisarse el concepto de optimalidad aplicado, por lo que debe recurrirse a algún criterio normativo para poder valorar las distintas asignaciones posibles. Manski (2001) analiza la forma en que un planificador benevolente, que pretende maximizar el bienestar social, asigna los individuos a los distintos programas. En esto, el planificador únicamente puede discriminar entre los individuos a partir de las características observables  $X$ .

Alternativamente, es posible que la actuación racional del decisor público se traduzca en una regla de asignación de tratamientos informada por el criterio de equidad. Así lo hacen, por ejemplo, Heckman *et al.* (1997).

## 5.2. Análisis formal de la asignación óptima de tratamientos

Manski (2001) formaliza el comportamiento de un planificador benevolente que asigna tratamientos rivales entre la población a la luz del rango de variación de las respuestas a un mismo tratamiento. El comportamiento racional del planificador benevolente lo exponemos formalmente a continuación siguiendo, en lo esencial, a este autor<sup>26</sup>.

---

<sup>26</sup> MANSKI (2001) sostiene que, al margen del análisis formal, decisores públicos, como jueces, médicos y otros, aun sin disponer de información sobre cómo varían los resultados de los tratamientos entre la población, abordan el problema de la elección entre tratamientos rivales, como describe.



Supongamos que existe un conjunto finito  $T$  de tratamientos mutuamente excluyentes<sup>27</sup>. El planificador debe elegir una regla que permita asignar un tratamiento específico del conjunto  $T$  a cada miembro de la población total  $J$ .

Las reacciones de cada individuo  $j \in J$  al tratamiento al que se le somete se representan por su función de respuesta,

$$y_j(\cdot) : T \rightarrow Y \quad (12)$$

con  $y_j(\tau) \in Y$ . En definitiva, una regla de tratamiento es una función  $\tau(\cdot) : J \rightarrow T$  que especifica qué tratamiento recibe cada persona. Por tanto, la respuesta del individuo  $j$ -ésimo bajo la regla  $\tau(\cdot)$  es  $y_j[\tau(j)]$ .

El planificador observa ciertas covariables/características  $x_j \in X$  para cada miembro de la población.  $Z$  denota el espacio de todo el mapa de funciones,  $X \rightarrow T$ . Entonces, las reglas de tratamiento factibles son

$$\{\tau(j) = z(x_j), j \in J\}, z(\cdot) \in Z \quad (13)$$

Ahora suponemos que el planificador quiere maximizar el resultado medio de la población. Una regla como  $z^*(\cdot)$  puede mostrarse como óptima si

$$z^*(x) \in \arg \max_{\tau \in T} E[y(\tau)|x], x \in X \quad (14)$$

En otros términos, la regla de asignación será óptima si el bienestar social resultante es máximo, lo que permitiría:

$$E_x \{ \max_{\tau \in T} E[y(\tau)|x] \} \quad (15)$$

Por último, el valor de la información que proporcionan las características observables resulta apropiadamente medido por la diferencia entre el mejor resultado medio logrado con y sin el uso de esta información.

De esta forma, si no se usa la información de las características de la población  $y$ , por tanto, se utiliza un tratamiento uniforme para todos los individuos, el mejor resultado medio logrado sería:

$$\max_{\tau \in T} E[y(\tau)] \quad (16)$$

Consecuentemente, el valor de la información sobre las características observables puede cuantificarse por la diferencia

$$V(x) = E_x \{ \max_{\tau \in T} E[y(\tau)|x] \} - \max_{\tau \in T} E[y(\tau)] \quad (17)$$

---

<sup>27</sup> En el caso de las políticas públicas es razonable suponer que el número de tratamientos rivales es finito.

Analizando la expresión anterior se demuestra que la información sobre las características observables de los individuos no tiene valor si la respuesta al tratamiento es suficientemente homogénea, de manera que exista un tratamiento óptimo común, esto es,  $\tau^* \in T$  tal que  $z^*(x) = t^*$  para casi cualquier caso del conjunto de características  $X$ .

Cuanto más varíe la respuesta al tratamiento con  $X$  más valiosa resulta la información de las características para el decisor público.

## 6. ¿ES POSIBLE AVANZAR?

Algunos órganos de control externo españoles han aplicado técnicas de evaluación económica de políticas públicas recientemente, obteniendo resultados esperanzadores. La aplicación del DEA a la evaluación del servicio de radiodiagnóstico del Servicio Andaluz de Salud, como revela Colinas (2006), es un ejemplo de esta línea de desarrollo.

La evaluación de políticas públicas en poblaciones heterogéneas, cuando se enmarca en los métodos cuasi-experimentales y aborda el problema de la heterogeneidad, según hemos expuesto, a partir del método de la subclasificación, necesita de la información empírica adecuada para poder controlar por características observables y para conocer los valores de la variable respuesta. Solo así es posible obtener el valor de los estimadores  $\alpha_{\text{PATE}}$  y  $\alpha_{\text{PATT}}$ , o bien  $\alpha_{\text{SATE}}$  y  $\alpha_{\text{SATT}}$ , conocer el rango de variación de los mismos entre las subpoblaciones y estar en condiciones de abordar finalmente la asignación óptima de tratamientos.

En evaluaciones de programas de formación que toman como variable respuesta los ingresos laborales, la información procedente de organismos de gestión y recaudación tributaria puede ofrecer un soporte empírico adecuado. En el caso español, no obstante, el artículo 95 de la Ley General Tributaria<sup>28</sup> impide esta posibilidad al establecer que «*los datos, informes o antecedentes obtenidos por la Administración tributaria en el desempeño de sus funciones tienen carácter reservado*»<sup>29</sup>.

Con objeto de favorecer este tipo de investigaciones evaluatorias, una opción a considerar es incluir en la normativa reguladora de los

<sup>28</sup> Ley 58/2003, de 17 de diciembre.

<sup>29</sup> Ninguna de las excepciones incluidas en el artículo 95.1 resultan relevantes en nuestra argumentación. Tampoco en el apartado g) del mismo artículo, que autoriza la cesión de datos al Tribunal de Cuentas, pero solo «*en el ejercicio de sus funciones de fiscalización de la Agencia Estatal de Administración Tributaria*».

programas de formación la obligación de facilitar los datos relevantes a los responsables de la evaluación. Ésta fue la vía seguida por el JTPA<sup>30</sup> en EE.UU.

Sin circunscribirnos a la información tributaria, la Ley Orgánica de Protección de datos de carácter personal<sup>31</sup> ofrece más amplias posibilidades a la investigación evaluatoria cuando ésta, además, la realizan los órganos de control externo. En este sentido, aunque su artículo 11.1 exige que los datos de carácter personal –base empírica de la investigación– sean cedidos *con el previo consentimiento del interesado*, lo que en la práctica dificultaría disponer de muestras amplias, tal consentimiento no será preciso cuando la comunicación tenga por destinatario el Tribunal de Cuentas en el ejercicio de las funciones que tiene atribuidas. Además, el apartado *d)* del artículo 11.2 amplía la excepción anterior a las instituciones autonómicas con funciones análogas al Tribunal de Cuentas.

Las posibilidades abiertas al soporte empírico de las investigaciones todavía pueden aumentarse considerando el apartado *e)* del precitado artículo. Este apartado también hace innecesario el consentimiento del interesado *«cuando la cesión se produzca entre Administraciones Públicas (es el caso que aquí se considera) y tenga por objeto el tratamiento posterior de los datos con fines (...) científicos»*<sup>32 y 33</sup>.

Analizada la normativa relevante en materia de protección de datos, puede concluirse que particularmente el Tribunal de Cuentas y, en general, los órganos de control externos españoles están en condiciones de impulsar las investigaciones evaluatorias en el marco de sus funciones. El reforzado carácter empírico de estas investigaciones, cuando se evalúan tratamientos sobre poblaciones heterogéneas, también hace de los órganos de control instituciones que pueden muy bien impulsarlas.

---

<sup>30</sup> El programa asociado a la *Job Training Partnership Act –JTPA–* estadounidense fue objeto de muy influyentes estudios evaluatorios, de entre los que destacan los de HECKMAN, ICHIMURA y TODD (1997, 1998), y el de HECKMAN, ICHIMURA, SMITH y TODD (1998). El JTPA entró en vigor en 1983.

<sup>31</sup> Ley Orgánica 15/1999, de 13 de diciembre, modificada por la Ley 62/2003, de 30 de diciembre.

<sup>32</sup> El texto entre paréntesis es nuestro.

<sup>33</sup> Dada la temprana aplicación de la inferencia causal estadística a la investigación médica, es importante tomar en consideración el apartado *f)* del artículo 11.2, que facilita la investigación evaluatoria al considerar que tampoco es necesario el consentimiento *«cuando la cesión de datos (...) relativos a la salud sea necesaria para (...) realizar los estudios epidemiológicos en los términos establecidos en la legislación sobre sanidad estatal o autonómica»*.

La operatividad de este planteamiento se beneficia del hecho de que el desarrollo de las evaluaciones, el tratamiento de las bases de datos y el cálculo de los estimadores son factibles a través del uso de paquetes informáticos como Stata, E-views o Matlab.

## 7. CONCLUSIONES

La evaluación de políticas públicas mediante técnicas de inferencia causal estadística debe resolver el problema que plantea la heterogeneidad de la población cuando el escenario en el que se desarrolla la evaluación no es el de un experimento social. Este problema afecta de manera crucial a la relevancia de las evaluaciones basadas en la estimación de los efectos medios del tratamiento.

Cuando las características que motivan las diferentes respuestas de los individuos a los tratamientos resultan observables al investigador, el control de las mismas mediante el método de *selección sobre observables* resulta adecuado; particularmente, en este artículo se ha propuesto el método de la subclasificación, aunque se apuntan otros desarrollos.

La evidencia encontrada en estos estudios evaluatorios debe, no obstante, contrastarse con el arrojado por otros estimadores, particularmente en el caso de existir características no observables. El estimador de diferencias en diferencias puede resultar en este caso de ayuda.

Los órganos de control externo españoles y, de manera especial, el Tribunal de Cuentas cuentan con las habilitaciones legales necesarias para poder recopilar el soporte empírico necesario para este tipo de evaluaciones. La implementación de su tratamiento se beneficiaría de los paquetes informáticos disponibles.

Sobre los resultados de los análisis evaluatorios en poblaciones heterogéneas, los decisores públicos podrían desarrollar reglas de asignación que permitan –bien sobre un criterio de eficiencia, bien sobre un criterio de equidad– optimizar los efectos de los tratamientos o políticas públicas.

## BIBLIOGRAFÍA

ANGRIST, J. D. (1998): «Estimating the Labor Market Impact of Voluntary Military Service Using Security Data on Military Applicants», *Econometrica*, 66 (2): 249-288.

- ANGRIST, J. D., y IMBENS, G. W. (1991): «Sources of Identifying Information in Evaluation Models». *Harvard Institute of Economic Research Working Papers* 1568, Harvard-Institute of Economic Research.
- ANGRIST, J.; IMBENS, W., y RUBIN, B. (1996): «Identification of casual effects using instrumental variables: Rejoinder». *Journal of the American Statistical Association*, 91 (434): 468-472.
- ASHENFELTER, O. (1978): «Estimating the Effect of Training Programs on Earnings». *Review of Economics and Statistics*, 60 (1): 47-57.
- BARNOW, B.; CAIN, G., y GOLDBERGER, A. (1980): «Selection on Observables», en Stromsdorfer, E. W., y Farkas, G. (eds.), *Evaluation Studies Review Annual*, vol. 5, 43-59, Beverly Hills, California, Ed.: Sage Publications.
- BILLEWICZ, W. Z. (1965): «The Efficiency of Matched Samples: an Empirical Investigation». *Biometrics*, 21: 623-643.
- CAMERON, E., y PAULING, L. (1976): «Supplemental Ascorbate in the Supportive Treatment of Cancer: Prolongation of Survival Times in Terminal Human Cancer». *Proceedings of the National Academy of Sciences (USA)*, 73: 3685-3689.
- CANSINO, J. M. (2004): «Una nota sobre los principios de eficiencia y eficacia en la programación y gestión presupuestaria a la luz de la nueva Ley General Presupuestaria». *REVISTA ESPAÑOLA DE CONTROL EXTERNO*, nº 18, Vol. VI: 221-234.
- CANSINO, J. M., y SÁNCHEZ, A. (2004): «Propuestas metodológicas para la evaluación de políticas públicas de formación. Una revisión crítica». *Documento de trabajo E2004/*. Fundación Centro de Estudios Andaluces.
- COLINAS, F. (2006), «Aplicación del método DEA a la evaluación de la eficiencia de los servicios de radiodiagnóstico hospitalario del SAS», en Cansino, J. M.: *La calidad del gobierno. Evaluación económica de políticas públicas*. Ed. Delta Publicaciones.
- COCHRAN, W. G. (1968): «The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies». *Biometrics*, 24: 295-313.
- DEHEJIA, R. H., y WAHBA, S. (1999): «Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs». *Journal of the American Statistical Association*, 94 (448): 1053-1062.

- DOCKSUM, K., (1974): «Empirical probability plots and statistical inference for nonlinear models in the two-sample case». *Annals of Statistics* 2: 267-277.
- DURÁN, C. (2004): «Evaluación microeconómica de las políticas públicas de empleo: Aspectos metodológicos». *Hacienda Pública Española/Revista de Economía Pública*, 170-(3/2004): 107-133.
- FISHER, R. A. (1935): *The design of experiments*. Boyd. Londres.
- FITZGERALD, J.; GOTTSCHALK, P., y MOFFITT, R. (1998): «An analysis of sample attrition in panel data: The Michigan panel study of income dynamics». *Journal of Human Resources*, 33, 251-299.
- HECKMAN, J. J., y HOTZ (1989): «Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training». *Journal of the American Statistical Association*, 84 (408): 862-874.
- HECKMAN, J. J.; ICHIMURA, H., y TODD, P. E. (1997): «Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme». *Review of Economics Studies*, 64 (4): 605-654.
- (1998): «Matching As an Econometric Evaluation Estimator», *Review of Economics Studies*, 65 (2): 261-294.
- HECKMAN, J. J.; ICHIMURA, H.; SMITH, J., y TODD, P. E. (1998): «Characterizing selection bias using experimental data». *Econometrica*, 66: 1017-1098.
- HECKMAN, J. J., y SMITH, J. A. (1995): «Assessing the Case for Social Experiments». *Journal of Economic Perspectives*, 9 (2), 85-110.
- HECKMAN J. J.; SMITH, J., y CLEMENTS, N. (1997): «Making the most out of programs evaluations and social experiments: Accounting for the heterogeneity in programs impacts». *Review of Economics Studies*, 64: 421-471.
- HECKMAN, J. J., y VITLACIL, E. (1999): «Local instruments variables and latent models for identifying and bounding treatments effects». *Proceedings of the National Academy of Science USA, Economic Sciences*, 96: 4730-4734.
- HIRANO, K.; IMBENS, G. W., y RIDDER, G. (2003): «Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score». *Econometrica*, 71(4): 1161-1189.
- IMBENS, G. W. (2004): «Nonparametric estimation of average treatment effects under exogeneity: A review». *The Review of Economics and Statistics*, 86 (1): 4-29.

- KRUEGER, A. (1999): «Experimental estimates of education production functions». *Quarterly Journal of Economics*, 114: 497-532.
- LALONDE, R. J. (1986): «Evaluating the Econometric Evaluation of Training Programas with Experimental Data». *American Economic Review*, 76 (4): 604-620.
- LEHMAN, E. (1974): *Nomparametrics: Statistical methods based on ranks*. Holden-Day. San Francisco.
- MANSKI, C. F. (2001): «Designing programs for heterogeneous populations: The value of covariate information». *The American Economic Review*, 91 (2): 103-106.
- NEYMAN, J. (1923): «On the application of probability theory to agricultural experiments». Essay on principles. Section 9. Traducido (con discusión) en *Statistical Science* 5:4 (1990), 465-480.
- (1935): «Statistical problems in agricultural experimentation». *Supplement to the Journal of The Royal Statistical Society*, 2: 107-180.
- ROY, A. D. (1951): «Some thoughts on the distribution of earnings». *Oxford Economics Papers*, 3: 33-38.
- RUBIN, D. B. (1974): «Estimating causal effects of treatments in randomized and non-randomized studies». *Journal of Educational Psychology*, 66, 688-701.
- (1977): «Assignment to a Treatment Group on the Basis of a Covariate». *Journal of Educational Statistics*, 2 (1): 1-26.
- (1978): «Bayesian Inference for Causal Effects the Role of Randomization». *Annals of Statistics*, 6 (1), 34-58.

