

Traducción de

BENJAMÍN GONZÁLEZ RODRÍGUEZ

CIS

Centro de
Investigaciones
Sociológicas

Diseño estadístico para la investigación

Leslie Kish

XI
Siglo veintiuno
de España
Editores, sa

plazados por los efectos secundarios. La evaluación se vuelve más complicada; la pérdida de efectos beneficiosos varía entre pacientes y con el tiempo, los efectos secundarios varían igualmente con los pacientes y con el tiempo; el pronóstico resulta difícil, tanto para los efectos secundarios como para los nocivos; una evaluación general debe necesariamente incluir pronósticos que tengan en cuenta todos estos factores (Sacks, 1973).

Debemos reconocer que en muchas o en la mayoría de las situaciones, no es posible obtener mediciones abundantes y prolongadas por lo que hemos de basarnos en modelos y en otro tipo de indicadores de cambio de los efectos. En muchos casos, por ejemplo, si un único tratamiento X tiene un impacto igual a cero, tras un intervalo corto pero apropiado, podemos suponer que el efecto cero se está produciendo en el futuro como resultado de ese tratamiento. Son posibles todos los efectos temporales que aparecen en la figura 3.5, pero es de esperar que, en la mayoría de las situaciones, la decisión se plantee casi siempre entre $A1$ y $B1$, o entre $A2$ y $B2$. Por supuesto, hay otras causas, además de la de tratamientos repetidos, que pueden producir efectos futuros. Sin embargo, la combinación de varios tratamientos es un tema demasiado amplio como para incluirlo en el presente análisis aquí planteado.

La importancia de tener en cuenta los factores estacionales de las mediciones se pone de relieve en

los datos de Marshall y Swan (1971)... sobre el crecimiento medio de la altura de dos "cohortes" cuyas edades son exactamente, 7,0 y 7,5 años a comienzos del año. Se supone que, a lo largo del período de un año, ambos crecen en 6,0 cm, de manera que no habrá habido una tendencia secular operando sobre el medio año que separa las cohortes. Vemos también que la tasa de crecimiento más rápido se produce en la primavera. De esta manera, una estimación de la tasa media de crecimiento que se base en un período inferior a un año dará una estimación sesgada. Por ejemplo, al utilizar cualquiera de las dos cohortes, si estimamos una tasa de crecimiento basada en el período de seis meses que va de enero a julio, obtenemos un valor de 7,2 cm al año, un 20 % más de lo debido. De igual modo, si basamos una estimación en la diferencia entre la cohorte de más edad en abril y la de menor edad en julio, obtenemos un valor de 5,4 cm al año, que se queda corto en un 11 % [Goldstein, 1979].

No sólo las variaciones estacionales, sino incluso las variaciones diarias, pueden afectar a muchas mediciones corporales. Varios investigadores han demostrado que las mediciones de la altura arrojan por

un aumento de 1,27 cm más por la mañana, después de haber estado acostado en la cama, que al anochecer, debido a la compresión de la columna producida por la postura erecta durante el día. En este campo es necesario poseer conocimientos sustantivos, a fin de poder planificar el diseño estadístico de estudios a lo largo del tiempo. El dilema que esta breve sección puede hacer es llamar la atención sobre este problema, y no sólo contradecir la primera cita sobre los efectos inmediatos, sino también alertar frente a la conclusión absoluta en las meras diferencias entre pretest y postest.

INVESTIGACIÓN EVALUATIVA

Constituye un reto el definir, describir y delimitar la investigación evaluativa (IE), al objeto de distinguirla de la investigación social y de la investigación en general, cubriendo, a pesar de todo, los diversos aspectos especiales de las situaciones concretas. Dado además que la IE es un campo relativamente nuevo, los lectores llegarán a él desde horizontes profesionales muy dispares. Aunque existieron actividades previas en este terreno, 1955 puede ser el año del nacimiento formal de la IE (en el *International Social Science Bulletin* de 1955). En 1967 se podían encontrar algunos informes concretos y unos pocos manuales, pero ningún libro de texto (Wright, 1967). Hacia 1985 se había convertido en una industria floreciente, que contaba con libros de texto, cursos, una revista (*Evaluation*) y una asociación, la Evaluation Research Society.

Dos tendencias destacan en la actitud moderna frente a la evaluación. De acuerdo con la primera, la evaluación ha llegado a ser el acompañamiento habitual de cualquier programa racional de acción social. Según la segunda, ha habido un movimiento en demanda de pruebas del éxito más sistemáticas, rigurosas y objetivas. A la aplicación de técnicas de las ciencias sociales para la valoración de los programas de acción social se la ha denominado investigación evaluativa.

Un enfoque científico de la valoración de los logros de un programa constituye el sello distintivo de la moderna investigación evaluativa. A este respecto, la investigación evaluativa se parece a otros tipos de investigación social en su preocupación por la objetividad, la fiabilidad y la validez en la recogida, análisis e interpretación de los datos. Pero tiene sus rasgos propios en cuanto que constituye una forma especial de investigación social debido a

sus objetivos y a las condiciones en que se deberá realizar la investigación. Estos dos factores afectan a componentes del proceso de investigación tales como el diseño del estudio y su puesta en práctica, la asignación del momento de investigación y otros recursos, y el valor o mérito que se debe otorgar a los hallazgos empíricos [Wright, 1967].

La investigación social incluye, desde este punto de vista, a la epidemiología y a las ciencias de la salud (Susser, 1975), pero no a la investigación de laboratorio o investigación biológica sobre animales.

Por lo tanto, la IE constituye un tipo especial de investigación social que tiene relación con los programas sociales y que implica la toma de decisiones.

La evaluación existe [...] para facilitar la toma de decisiones inteligente [...] Si no mejora la base para la toma de decisiones sobre el programa y sus competidores, pierde entonces su carácter peculiar como investigación evaluativa y se convierte simplemente en investigación social. Consideramos que los programas más significativos son evaluados porque alguien con poder de toma de decisiones desea obtener ayuda para tener una idea de qué hacer... Quizás el tema a decidir con mayor frecuencia sea: ¿Debemos continuar haciendo esto o deberíamos intentar otra cosa, incluso no hacer nada? [Edwards, Guttentag y Snapper, 1975].

Este capítulo es el mejor lugar para discutir la IE, porque los problemas de diseño (y también de análisis) que encierra son, en esencia, parecidos a los de los diseños para comparaciones en otros tipos de investigación social. Sin embargo, la IE merece una mención independiente, porque sus objetivos y alcance son más específicos, restringidos y claramente definidos que los de la investigación social en general. Presento a continuación mi propia lista de características que ayuda a distinguir la IE de los restantes tipos de investigación social, aunque no existe ninguna línea fina y nítida que separe a ambas, en ninguna de las características concretas.

1. La IE implica programas concretos de acción social. Estos programas podrán ser de alcance local, nacional o incluso internacional, así, por ejemplo, algunos programas de control de natalidad o contra el analfabetismo disponen de organizaciones internacionales. En general, los programas han de ser de suficiente envergadura como para afrontar los costes de diseños IE que sean lo suficientemente buenos y de un alcance tal que proporcionen precisión y capacidad

para detectar los cambios, aunque sean modestos. La mayoría de las mejoras logradas en los programas son moderadas, pero pueden tener amplias consecuencias de carácter social o financiero. Por otra parte, aquellos programas que tengan un éxito evidente pueden no tener necesidad de la IE.

2. La IE comprende, como norma, la cooperación entre alguna agencia (departamento) responsable del programa de acción social y un equipo de investigación (instituto) que lleve a cabo la IE. Se necesita cierto grado de separación entre el equipo y la agencia: para facilitar, en primer lugar, la objetividad de la investigación, y, en segundo término, para reforzar la percepción y aceptación pública de la objetividad de los resultados de la IE. A veces se necesita un patrocinador o fuente de fondos económicos para la IE, fuente que debe ser independiente tanto de la agencia como del equipo de investigación.

La cooperación activa entre las dos (o tres) partes suele ser fundamental. La agencia tiene la responsabilidad principal de los objetivos de la investigación, pero los investigadores tienen que tomar la iniciativa en la determinación de los métodos y en la ejecución de la investigación. Al seleccionar las áreas de investigación, y al asegurar su cooperación, suelen ser necesarias tanto la libertad del equipo investigador como la colaboración por parte de la agencia.

3. Tanto la elección de los tratamientos como la observación de los efectos se ven restringidas por las exigencias prácticas de evaluar la eficacia del programa en estudio.

En consecuencia, [el equipo] tiene menos libertad para seleccionar o rechazar determinadas variables independientes, dependientes e intervinientes de la que tendría en estudios diseñados para responder a [sus] propias preguntas formuladas teóricamente, como las que podrían plantearse en la investigación social básica. Los conceptos utilizados y su traducción en variables medibles deben seleccionarse con imaginación, pero dentro del marco general establecido por la naturaleza del programa que se está evaluando y por sus objetivos [Wright, 1967].

Estos objetivos restringidos conllevan varias consecuencias importantes. En primer lugar, el programa conjunto concreto se convierte en la variable de tratamiento (independiente, predictora), y sus diversos componentes pueden quedar desenmarañados. Por ejemplo, un nuevo programa de enseñanza escolar podrá ser una combinación

de métodos, profesores y ambientes. En segundo lugar, los efectos medidos deben concentrarse en los objetivos propuestos como objetivos de respuesta (variable dependiente, pronosticada) principal. No obstante, puede ser conveniente, y a veces posible, añadir efectos a largo plazo a los objetivos marcados a corto plazo. Por otro lado, se deberían anticipar y vigilar, además de los objetivos planeados, determinadas consecuencias no programadas, tanto las beneficiosas como, especialmente, las perniciosas. Más aún, el equipo debería estar alerta para detectar algunas consecuencias totalmente imprevistas. La "serendipidad" puede a veces revelar resultados más interesantes que los efectos principales esperados. En tercer lugar, el programa debe evaluarse en el marco de su entorno natural, con sus propias condiciones operativas, y para su propia población objetivo. En relación con todos estos aspectos no necesita ser sometido a otras pruebas de falsabilidad más severas, y que vayan más allá de los confines estrechos de las condiciones estipuladas del programa (7.6). Por supuesto, la superación de tales pruebas contribuiría a fortalecer la credibilidad, estabilidad y carácter generalizable de los resultados de la IE.

4. La elección del momento apropiado para la IE entraña a menudo problemas especiales. Resulta difícil concluir la investigación, que va seguida de decisiones de política social basadas en sus resultados, antes de comenzar un programa, a pesar que esto sería lo preferible, por varias razones. La modificación o desmantelamiento de un programa en curso podría encontrar obstáculos, incluso aunque no haya dado resultados, porque los defectos y críticas son una amarga medicina para la gente que tiene que ver con su funcionamiento. Además, cuando un programa en curso sirve a toda la población, se plantean asimismo problemas para encontrar lugares adecuados para las comparaciones de "control".

Por otra parte, completar la IE y tomar seguidamente decisiones sobre los nuevos programas propuestos, antes de su implantación, implica retrasos y problemas de financiación. Además, los obstáculos administrativos pueden dificultar la introducción del programa propuesto para una muestra objetivamente aleatorizada, lo cual trae consigo los problemas relativos a los tratamientos aleatorizados (1.4, 3.1).

Cuando el programa se introduce de modo parcial, el equipo investigador debería convencer a la agencia para que implante un buen diseño de IE en la selección de áreas y del calendario de implantación.

¿Cualquier programa que se propone para todas las áreas, puede pedirse una cierta prórroga para hacer una muestra de control (aleatorizada) de áreas.

La necesidad de tomar decisiones rápidas en el proceso de la IE se contrapone a los ideales de la investigación académica referentes a la neutralidad, a leyes universales y a la falsabilidad (7.6). Al enfrentarse con estos problemas e incertidumbres, el enfoque estadístico aconseja confiar en las probabilidades y en la comprobación repetida mediante la IE, a fin de comprobar la efectividad continuada del programa, tras el cambio del entorno y de las "condiciones circundantes".

5. «Las evaluaciones están ahí para facilitar la toma de decisiones inteligente». Estimar la *efectividad* de un programa va más allá de medir sus efectos. No basta, en primer lugar, con demostrar la mera existencia de efectos frente a la hipótesis nula de efectos cero, mediante los valores P de las pruebas de "significación". Tampoco es la capacidad explicativa (R^2) de la variable de tratamiento el principal problema, sino que, por el contrario, es la magnitud del efecto lo que se convierte en el problema principal. Además, a menudo es preciso medir varios efectos, combinados de alguna manera —formalmente a través de un índice o de un proceso de toma de decisión o de modo informal. Con frecuencia, hay que sopesar los efectos negativos frente a los efectos positivos. En todo caso, hay que calibrar el efecto del programa comparándolo con una medida realista y aceptable de su verdadero *coste* total. El *coste/efectividad* del programa, que constituye una función de uso múltiple, deberá ser evaluado por parte de varias instancias decisorias, lo que puede implicar no sólo a la agencia responsable del programa, sino también a instituciones públicas, expertos, usuarios del programa (presentes y futuros) y público en general (Edwards, Guttentag y Snapper, 1975).

6. También pueden ser diferentes de las normas aceptadas en la investigación social académica, las normas de publicación de los resultados de la IE. Los informes elaborados para su entrega a la agencia deben elaborarse generalmente con premura, a fin de encajar en unos calendarios preestablecidos y fijos. Por el contrario, los informes científicos para el público pueden verse aplazados, obstaculizados o retenidos por razones de confidencialidad, de falta de fondos, o de ausencia de incentivos positivos.

7. Los requisitos precedentes de la IE en cuanto a restricciones, decisiones y rapidez, suelen entrar en conflicto con las normas usuales de los investigadores académicos. Sin embargo, los equipos que llevan a cabo la IE necesitan (según se apuntó en el apartado 1) independencia y objetividad, tanto real como aparente. Una forma deseable de organización podría constar de equipos independientes de investigación dentro de las agencias públicas, pero con independencia y reputación garantizadas. Un ejemplo precoz (1937) lo tenemos en la *Division of Program Surveys*, establecida en forma independiente para evaluar los múltiples programas del Departamento de Agricultura de Estados Unidos (Converse, 1986, cap. 5). Es posible hallar algunos buenos ejemplos de este tipo en Estados Unidos y en otras partes, pero no los suficientes.

El progreso de la IE en Estados Unidos se ha reflejado también en el desarrollo de institutos y centros que se han dedicado muy activamente a la IE y a investigaciones parecidas, estando algunos de estos centros vinculados con universidades, y otros trabajando de manera independiente.

CONTROLES PARA LAS VARIABLES PERTURBADORAS

«Sólo cuando la atención haya sido sustituida por la negligencia, como práctica generalizada, empezará a demostrar la estadística las ventajas de cada método de atención concreto que se adopte.»

G. B. SHAW, en el prefacio de *The Doctors Dilemma*.

«[...] nada mejora tanto el rendimiento de una innovación como la falta de controles.»

Postulado de Muench en *Experimentation and Innovations*, de MOSTELLER (1977).

4.1. ESTRATEGIAS DE CONTROL

4.1A. Cuatro controles para el sesgo

Los controles son necesarios en los diseños de investigación por dos razones esenciales. Una de ellas se refiere a la eficiencia y economía de los proyectos de investigación, y se puede expresar en términos de reducir ("minimizar") ya sean las varianzas ya sean los costes de los proyectos. La segunda razón tiene que ver con los sesgos que tienen origen en las variables perturbadoras en los diseños no aleatorizados. Es en este segundo aspecto en el que insistiré en este capítulo. En los experimentos auténticos (ideales) estas variables perturbadoras quedan todas ellas eliminadas por el diseño, convirtiéndose o bien en variables controladas *C*, o bien en variables aleatorizadas *A*. Los controles en el diseño de experimentos tienen como objetivo principal la reducción de los efectos causados por las variables aleatorizadas *A* a través del aumento de los efectos de las variables controladas *C*. Además hay que añadir, sin embargo, que en los estudios observacionales no experimentales ni aleatorizados, algunas variables perturbadoras *P* pueden permanecer como factores que sesgan las observaciones y las comparaciones de las variables explicativas (sección 1.2). Pondremos aquí el acento sobre la reducción de estos sesgos, al ir analizando las medidas para el control de las variables perturbadoras.

Para manejar las variables perturbadoras, es posible recurrir a varios métodos alternativos, dependiendo de las situaciones y recursos