

Valorando la efectividad de las intervenciones en salud pública: la fuerza de los «diseños evaluativos débiles»

Manel Nebot^{a,b,c,d,*} y César Victora^{d,e,f}

^a Agencia de Salud Pública de Barcelona, Barcelona, España

^b CIBER de Epidemiología y Salud Pública (CIBERESP), España

^c Departamento de Ciencias Experimentales y de la Salud, Universitat Pompeu Fabra, Barcelona, España

^d Institut d'Investigacions Biomèdiques Sant Pau (IIB Sant Pau), Barcelona, España

^e Programa de Pos-Grado en Epidemiología, Universidad Federal de Pelotas, Pelotas, Brasil

^f Department of International Health, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Estados Unidos

Desde hace algunos años asistimos a un creciente debate conceptual y metodológico sobre la necesidad (y sobre todo la forma) de superar el paradigma del diseño experimental en la evaluación de resultados en salud pública^{1,2}. Para algunos autores no tiene sentido empeñarse en buscar indicadores objetivos de efectividad, que sólo podrían valorarse en el contexto específico en que se produce la intervención, por lo que debemos concentrarnos en entender «desde dentro» los mecanismos del cambio que sigue a un programa³. Efectivamente, la imposibilidad de controlar todos los factores relevantes que influyen en la intervención, especialmente los contextuales, y de encontrar grupos de control equivalentes para valorar el efecto contrafactual, nos obligan a buscar métodos y abordajes distintos.

En una posición pragmática, podemos tratar de buscar evidencias de efecto combinando la identificación de indicadores de cambio con la valoración de la atribución del cambio a la intervención⁴. El primer paso sería la demostración de que se ha producido un efecto o cambio (lo que se ha llamado *adecuación* de la intervención), y a continuación deberíamos investigar hasta qué punto podemos atribuir con un cierto grado de certeza el cambio observado a la intervención, lo que podríamos llamar *plausibilidad*. Así, en la evaluación de algunas políticas públicas como la ley de tabaquismo o las normativas de seguridad vial, deberíamos constatar un cambio de los indicadores seleccionados y, además, relacionarlos con las medidas adoptadas en la implementación de las políticas, sean radares de control de velocidad, inspecciones en locales, campañas de sensibilización o cualquier otra medida encaminada a alcanzar los objetivos. Siempre que sea posible, es importante también demostrar que ha habido cambios en indicadores intermedios de las cadenas causales, por ejemplo en ventas de cigarrillos o en la velocidad de los vehículos.

En cualquier caso, no hay duda de que los planteamientos apropiados para la evaluación de la efectividad de intervenciones terapéuticas, que permiten un alto grado de control de las covariables y de los potenciales confusores, no son adecuados para evaluar intervenciones complejas, con componentes múltiples y dependientes del marco social, cultural o económico. Se puede afirmar que si en la investigación clínica la modificación de efectos (o la heterogeneidad entre los resultados de distintos estudios) es la excepción, en salud pública es la regla.

De hecho, las metodologías de estimación de impacto en salud, que no miden directamente los efectos de las intervenciones sino que tratan de estimarlos de forma indirecta a partir de la combinación de estudios epidemiológicos y datos de exposición, se están utilizando cada vez más para valorar consecuencias en salud de políticas públicas⁵. ¿Supone todo esto que debemos renunciar a la medida del efecto con los diseños evaluativos clásicos? Ciertamente no, pues sería absurdo dejar de utilizar un grupo de comparación si está disponible,

pero debemos ser muy conscientes de las limitaciones a la inferencia causal de efectividad que se pueden derivar de las diferencias entre los grupos. En este sentido, cabe señalar que hay ciertas metodologías (*propensity score*, *regression discontinuity*) que pueden mejorar el tratamiento de los sesgos de selección. Sin embargo, en muchas ocasiones no podremos utilizar un grupo de comparación, por lo que será aún más importante utilizar criterios como la consistencia de resultados con métodos distintos y la comprensión del proceso. La formulación de un modelo conceptual o marco lógico nos permitirá definir los indicadores a corto, medio y largo plazo, de proceso y de resultados, y garantizar que nuestro modelo de recolección de datos y de análisis estadístico tiene en cuenta tales indicadores.

Obviamente queda mucho camino por recorrer, pero empezamos a disponer de propuestas de partida, que deberán ser elaboradas en un largo y esperamos que rico y fructífero proceso de reflexión, debate y consenso. A esta reflexión esperamos que pueda contribuir la presente monografía, en la cual se abordan algunos de los aspectos principales de la evaluación de la efectividad en salud pública, en particular los relativos a los diseños cuasi-experimentales y no experimentales, en ocasiones llamados «diseños evaluativos débiles».

En el primer artículo se describen y analizan las bases conceptuales y metodológicas de la evaluación en salud pública, y se debate sobre las limitaciones de los abordajes convencionales, revisando algunas de las propuestas orientadas a superarlas. En el segundo se abordan algunos de los principales aspectos metodológicos de la evaluación en salud pública. Se identifican y explican los principales diseños evaluativos, analizando sus ventajas y limitaciones. Asimismo, se describe brevemente el análisis estadístico más utilizado para cada diseño. El tercer capítulo ilustra con diversos ejemplos los retos en la evaluación de políticas y planes de salud, y las formas de abordarlos para obtener indicadores que sean de ayuda en la toma de decisiones en planificación y gestión. En el cuarto se esbozan e ilustran con algunos ejemplos los principales métodos utilizados en la evaluación económica, de gran relevancia actualmente en los sistemas sanitarios y de salud. En el quinto artículo se presentan algunos ejemplos de evaluación de programas de promoción de la salud en la escuela, que permiten identificar aspectos relevantes del proceso a tener en cuenta en la evaluación, ya que permiten comprender los mecanismos implicados en el efecto de las intervenciones. El sexto trata de los conceptos y métodos en la evaluación de los programas vacunales, con ejemplos procedentes de programas aplicados en nuestro entorno. Finalmente, en el séptimo capítulo se analiza el papel de la evaluación de tecnologías en el marco general de la evaluación de la efectividad en salud pública.

En conjunto, la monografía no puede (ni se ha planteado) cubrir todos los aspectos del complejo y variado mundo de la evaluación.

Sin embargo, deseamos que al menos permita avanzar en la clarificación de los grandes retos y aportar propuestas concretas y elementos de reflexión. Esperamos, por lo tanto, que sea de utilidad.

Contribuciones de autoría

M. Nebot escribió el borrador y C. Victora lo revisó y editó.

Financiación

Este artículo ha sido elaborado con apoyo del Comissionat per a Universitats i Recerca del DIUE de la Generalitat de Catalunya (AGAUR SGR 2009-1345).

Conflicto de intereses

Los autores declaran no tener ningún conflicto de intereses.

Bibliografía

1. Nebot M. Evaluación en salud pública: ¿todo vale? *Gac Sanit.* 2007;21:95-6.
2. Victora CG, Habicht JP, Bryce J. Evidence-based public health: moving beyond randomized trials. *Am Journal P Health.* 2004;94:400-5.
3. Pawson R, Tilley N. Realistic evaluation. Sage: London (UK); 1997.
4. Habicht JP, Victora CG, Vaughan JP. Evaluation designs for adequacy, plausibility and probability of public health performance and impact. *IJE.* 1999;28:10-8.
5. Wismar M, Blau J, Ernst K, et al. The effectiveness of health impact assessment: scope and limitations of supporting decision-making in Europe. Copenhagen: World Health Organization; 2007.