

18. Cómo planificar el monitoreo y la evaluación de las políticas

Atle Fretheim¹, Andrew D Oxman², John N Lavis³ y Simon Lewin⁴

Fretheim A, Oxman AD, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 18. Planning monitoring and evaluation of policies.** Health Research Policy and Systems; 2009, **7(Suppl 1):S18** doi:10.1186/1478-4505-7-S1-S18.

<http://www.health-policy-systems.com/content/pdf/1478-4505-7-S1-s18.pdf>

1. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Section for International Health, Institute of General Practice and Community Medicine, Faculty of Medicine, University of Oslo, Norway
2. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway
3. Centre for Health Economics and Policy Analysis, Department of Clinical Epidemiology and Biostatistics, and Department of Political Science, McMaster University, 1200 Main St. West, HSC-2D3, Hamilton, ON, Canada, L8N 3Z5
4. Norwegian Knowledge Centre for the Health Services, P.O. Box 7004, St. Olavs plass, N-0130 Oslo, Norway; Health Systems Research Unit, Medical Research Council of South Africa

* Autor para correspondencia (atle.fretheim@nokc.no)

El presente documento es una traducción del artículo publicado en Health Research Policy and Systems, 2009; 7:Supplement 1 (www.health-policy-systems.com/supplements/7/S1). Está permitido su uso, distribución y reproducción ilimitada en cualquier medio, siempre y cuando la fuente esté debidamente citada. En el sitio web de SUPPORT (www.support-collaboration.org) se pueden encontrar enlaces a las traducciones de esta serie al chino, español, francés y portugués. Las sugerencias para correcciones y mejoras de las herramientas de esta serie, que son muy bienvenidas, deben enviarse a: STP@nokc.no.

Esta serie de artículos fue preparada como parte del Proyecto SUPPORT, respaldado por el programa INCO del 6º Programa Marco de la Comisión Europea, contrato 031939. Las siguientes organizaciones financiaron una reunión de revisión por pares donde se discutió una versión anterior de la serie: Norwegian Agency for Development Cooperation (Norad), Alliance for Health Policy and Systems Research (AHPSR) y Milbank Memorial Fund. John Lavis recibió apoyo salarial por su condición de Jefe de Investigación de Canadá en Intercambio y Transferencia de Conocimiento (Canada Research Chair in Knowledge Transfer and Exchange). Las siguientes organizaciones respaldaron la traducción y la disseminación de los artículos: Norad, la sede noruega del Grupo de Revisión Cochrane de Prácticas Efectivas y Organización de la Atención de la Salud (Cochrane Effective Practice and Organisation of Care, EPOC); Norwegian Knowledge Centre for the Health Services, AHPSR, Canadian Health Services Research Foundation (CHSRF), Evidence-Informed Policy Network (EVIPNet) y la Organización Panamericana de la Salud (OPS). Ninguno de los financiadores participaron en la preparación del borrador, la revisión o la aprobación del contenido.

Este artículo fue traducido al español por el Centro Rosarino de Estudios Perinatales (CREP) con el apoyo de Canadian Health Services Research Foundation (CHSRF) www.chsrf.ca, Centro Rosarino de Estudios Perinatales (CREP) www.crep.org.ar, y Organización Panamericana de la Salud (OPS) (www.paho.org/researchportal).



Resumen

El presente artículo forma parte de una serie de artículos destinada a los responsables de la toma de decisiones sobre políticas y programas de salud y a quienes los respaldan.

El término *monitoreo* se utiliza normalmente para describir el proceso por el cual se recolectan de manera sistemática datos para informar a responsables de la toma de decisiones en políticas, gestores y otras partes interesadas (*stakeholders*) si una nueva política o programa se está implementando según sus expectativas. Se utilizan indicadores con fines de monitoreo para juzgar, por ejemplo, si se están cumpliendo los objetivos o si se están administrando adecuadamente los fondos asignados. Algunas veces, los términos *monitoreo* y *evaluación* se utilizan indistintamente pero, por lo general, el segundo sugiere un enfoque más sólido en el logro de los resultados. Cuando se utiliza el término *evaluación del impacto*, generalmente implica que existe un intento específico para tratar de determinar si los cambios observados en los resultados pueden atribuirse a una política o un programa específico. En este artículo, sugerimos cuatro preguntas que pueden utilizarse para guiar el monitoreo y la evaluación de las opciones de políticas o programas. Éstas son: 1. ¿Es necesario el monitoreo? 2. ¿Qué debe medirse? 3. ¿Debe realizarse una evaluación del impacto? 4. ¿Cómo debe realizarse la evaluación del impacto?

ACERCA DE LAS HERRAMIENTAS SUPPORT (STP)

El presente artículo forma parte de una serie de artículos destinada a los responsables de la toma de decisiones sobre políticas y programas de salud y a quienes los respaldan. El objetivo de esta serie de artículos es ayudar a que estas personas se aseguren de tomar las decisiones informándose por la mejor evidencia de la investigación disponible. En la Introducción de esta serie [1], se describen con más detalle las Herramientas SUPPORT y las maneras en las cuales pueden ser utilizadas. Adjunto a cada artículo se encuentra un glosario que se aplica a toda la serie (ver Archivo adicional 1). En el sitio web de SUPPORT (www.support-collaboration.org) se pueden encontrar enlaces a las traducciones de esta serie al español, portugués, francés y chino. Las sugerencias para correcciones y mejoras de las herramientas de esta serie, que son muy bienvenidas, deben enviarse a: STP@nokc.no.

ESCENARIOS

Escenario 1: Usted es un funcionario público de alto rango responsable de varios programas de salud y quiere asegurarse de tener la información necesaria para evaluar cómo se están llevando a cabo los diferentes programas y qué impacto están teniendo.

Escenario 2: Usted trabaja en el Ministerio de Salud y le solicitaron que prepare un memorando sobre varios asuntos que deben tenerse en cuenta cuando se evalúe el programa nacional de vacunación.

Escenario 3: Usted trabaja en una unidad que respalda al gobierno en el uso de evidencia en la toma de decisiones en políticas. Está elaborando un plan de monitoreo y evaluación para el programa nacional de control de la tuberculosis.

ANTECEDENTES

Para los responsables de la toma de decisiones en políticas (Escenario 1), este artículo sugiere una serie de preguntas que podrían tener en cuenta los miembros de su equipo al planificar el monitoreo y la evaluación de una nueva política.

Para quienes respaldan a los responsables en la toma de decisiones en políticas (Escenarios 2 y 3), este artículo sugiere una serie de preguntas que deben considerarse al planificar cómo monitorear la implementación de políticas y programas, y la evaluación de sus impactos.

Los responsables de la toma de decisiones en políticas y otras partes interesadas (*stakeholders*) necesitarán saber, a menudo, si una nueva política o programa ha sido

implementada según sus expectativas. ¿Está progresando el lanzamiento del programa según lo planificado? ¿Se están logrando los objetivos? ¿Se están administrando adecuadamente los fondos asignados? *Monitoreo* es el término comúnmente utilizado para describir el proceso de recolección sistemática de datos para poder responder dichas preguntas [2]. El término *monitoreo del desempeño* se utiliza normalmente cuando el enfoque principal de una evaluación compara “qué tan bien que se está implementando un proyecto, programa o política en comparación con los resultados esperados” [2].

Con frecuencia, los *indicadores* se utilizan como parte del proceso de monitoreo. Se define como indicador a una “variable o factor cualitativo o cuantitativo que provea un medio simple y confiable para medir el logro, reflejar los cambios relacionados con una intervención o ayudar a la evaluación del desempeño” [2]. Un indicador puede ser un simple conteo de eventos (por ejemplo, la cantidad de vacunaciones realizadas dentro de un período de tiempo) o un constructo basado en varias fuentes de datos (por ejemplo, la proporción de todos los niños inmunizados antes de su primer cumpleaños).

Algunas veces, los términos *monitoreo* y *evaluación* se utilizan indistintamente pero, por lo general, el segundo sugiere un enfoque más sólido en el logro de los resultados. Estos términos no se utilizan de manera consistente y pueden tener diferentes significados para diferentes personas. El término *evaluación del impacto* se utiliza con frecuencia cuando se hace un intento de evaluar si los cambios observados en los resultados (o “impactos”) pueden atribuirse a una política o programa específico.

PREGUNTAS A TENER EN CUENTA

1. ¿Es necesario el monitoreo?
2. ¿Qué debe medirse?
3. ¿Debe realizarse una evaluación del impacto?
4. ¿Cómo debe realizarse la evaluación del impacto?

1. ¿Es necesario el monitoreo?

La importancia del monitoreo depende de la necesidad percibida entre las partes interesadas relevantes de querer saber más sobre lo que está pasando “sobre el terreno”.

Determinar si debe establecerse un sistema para monitorear una política o un programa puede depender de varios factores, entre los que se incluyen:

- Si ya existe un sistema de monitoreo que incluya los indicadores deseados o si se requiere un nuevo grupo de indicadores

- Los posibles costos para establecer el sistema solicitado. Por ejemplo, ¿se podrían agregar nuevos ítems a los procedimientos de recolección de datos ya existentes o es necesario realizar encuestas adicionales sobre familias a gran escala o desarrollar una herramienta completamente nueva?
- Si los hallazgos podrían ser útiles. ¿Qué acciones deben tomarse si el monitoreo revela que las cosas no funcionan según lo planificado?

El monitoreo no vale la pena si los datos no se utilizan. Los datos son especialmente útiles si se toman acciones correctivas cuando se identifica una brecha entre los resultados esperados y los reales. Dichos hallazgos pueden determinar la reconsideración de las expectativas. Esto puede tomar la forma de evaluaciones de, por ejemplo, si los planes iniciales eran demasiado ambiciosos o si una nueva política no ha funcionado tan eficazmente como se esperaba.

La Tabla 1 muestra dos ejemplos ilustrativos de los sistemas de monitoreo que se han puesto en funcionamiento dentro de los sistemas de salud [3,4].

2. ¿Qué debe medirse?

Los indicadores que se centran en varias partes de la “cadena de resultados” (es decir, en aportes, actividades, resultados o impactos – ver Figura 1) se utilizan normalmente para monitorear la implementación de una opción de política o programa. En algunos casos, puede considerarse suficiente para monitorear los aportes (es decir, la provisión de recursos, como por ejemplo personal y equipamiento). En otros, puede ser importante monitorear las actividades del programa o sus resultados (como la cantidad de niños totalmente inmunizados).

Cuando se selecciona qué indicador(es) debe(n) utilizarse, es necesario tener en cuenta una serie de factores [5,6]:

- *Validez*: hasta dónde el indicador mide de manera precisa lo que pretende medir
- *Aceptabilidad*: hasta dónde el indicador es aceptable para aquellos que están siendo evaluados y aquellos que realizan la evaluación
- *Viabilidad*: hasta dónde los datos consistentes, confiables y válidos se encuentran disponibles para su recolección
- *Confiablez*: hasta dónde existe un mínimo error de medida o hasta dónde los hallazgos son reproducibles en caso de ser recolectados de nuevo por otra organización
- *Sensibilidad para el cambio*: hasta dónde el indicador tiene la capacidad de detectar cambios en la unidad de medida
- *Validez predictiva*: hasta dónde el indicador tiene la capacidad de predecir de manera precisa resultados relevantes

Los costos relacionados con la recolección de datos y a la capacidad de análisis y devolución (*feedback*) de los datos para gestores y prestadores también pueden limitar

la elección de los indicadores. En lugares donde los recursos analíticos son escasos, puede ser preferible seleccionar un simple indicador, incluso si no tiene la mejor validez predictiva, en lugar de un indicador que requiere manipulación estadística.

Con frecuencia, existe un balance entre querer utilizar los indicadores deseados y óptimos por un lado y tener que utilizar aquellos indicadores que están basados en los datos existentes, por el otro. Existen buenas razones para no seleccionar más indicadores que los que son absolutamente esenciales. Estas razones incluyen la necesidad de limitar la carga de recolección de datos dentro de un sistema de salud, evitar la recolección de datos que no son utilizados y centrarse en la recolección de datos de mejor calidad, incluso si esto significa recolectar menos datos en total [7].

La información recolectada de rutina en los sistemas de salud puede proveer datos valiosos que pueden utilizarse como fuente de datos para los fines del monitoreo. Los datos también pueden recolectarse específicamente con el fin de monitorear (por ejemplo, mediante encuestas o entrevistas). Se debe tener en cuenta el nivel de motivación de las personas que se espera que recolecten datos. En muchos casos, el personal de la salud debe integrar la recolección de datos en un esquema diario atareado. Por lo tanto, si la información que está siendo recolectada no tiene valor o tiene poco valor obvio para ellos, su motivación para llevar a cabo dichas tareas puede ser baja. De igual manera, si los incentivos o las penalidades están asociados con los hallazgos del proceso de monitoreo (por ejemplo, donde el pago de los prestadores está relacionado con los indicadores de desempeño), debe tenerse en cuenta el riesgo de la manipulación de datos o jugarretas del sistema.

3. ¿Debe realizarse una evaluación del impacto?

Una de las limitaciones de las actividades de monitoreo, como se describe anteriormente, es el hecho de que dichas actividades no indican necesariamente si una política o programa ha tenido el impacto en los indicadores que han sido medidos. Esto se debe a que los indicadores utilizados para monitorear casi siempre serán influenciados por factores que no son aquellos relacionados con las intervenciones específicas, lo que dificulta considerablemente determinar qué factores causaron los cambios observados. Si el monitoreo revela que el desempeño está mejorando, esto no significa necesariamente que la intervención es el (único) factor causal. Es posible que los indicadores hubieran mejorado de todas maneras, incluso sin la intervención (ver Figura 2).

El establecimiento de una relación causal entre un programa o política y los cambios en los resultados es el objetivo central de la evaluación del impacto. Qué les habría pasado a aquellos que recibieron una intervención si de hecho no la hubieran recibido, es *la* pregunta principal de la evaluación del impacto, según el Banco Mundial [8].

Pueden existir razones convincentes para esperar resultados positivos basados en documentación sólida de, por ejemplo, evaluaciones previas. Sin embargo, con mucha frecuencia esta evidencia puede ser escasa. O bien la evidencia disponible no es aplicable al lugar en cuestión. Por lo tanto, existe un riesgo real de que un nuevo programa no sea efectivo, o incluso peor, ocasione más daños que beneficios. Es importante que los responsables en la toma de decisiones en políticas definan este tema cuando implementan nuevos programas. También es importante debido al beneficio que dicho conocimiento podría brindar a futuras tomas de decisiones en políticas de salud, tanto en el lugar del programa como en otras jurisdicciones.

La realización de evaluaciones de impactos puede ser costosa. Se puede determinar si esos estudios tienen valor al comparar las consecuencias de llevar a cabo una evaluación con las consecuencias de no llevar a cabo una evaluación. Por ejemplo, ¿es posible que un programa sea interrumpido o modificado si los resultados demuestran que es negativo? Si la respuesta es “no”, el valor de llevar a cabo una evaluación del impacto está claramente limitado.

En general, es más probable que una evaluación del impacto sea valiosa cuando los resultados se pueden obtener a medida que se implementa la intervención. En dichos casos, existe una oportunidad de mejorar o interrumpir el lanzamiento en función de los resultados de una evaluación del impacto realizada en las primeras etapas de la implementación. Esto sería valioso en dos circunstancias: en primer lugar, cuando no es posible un estudio piloto y, en segundo lugar, cuando fuera posible y práctico modificar o interrumpir el lanzamiento (de ser necesario) en función de los resultados.

El esquema de seguro de salud del gobierno mexicano, el Seguro Popular, es un ejemplo de una evaluación de impacto insertada en el lanzamiento de un programa [9-11]. Implementado en 2001, el esquema fue establecido con el fin de extender la cobertura del seguro de salud a casi 50 millones de mexicanos que aún no estaban cubiertos por los programas existentes. Y tomando ventaja del cronograma de introducción progresiva, el gobierno estableció una evaluación que comparó los resultados para aquellas comunidades que recibían el esquema con aquellas que aún lo estaban esperando. Además de evaluar si la reforma lograba los resultados deseados y no tenía efectos adversos no deseados, la evaluación también brinda el aprendizaje compartido.

Una evaluación del impacto también puede ser útil después de que un programa se haya implementado completamente, como cuando existe incertidumbre acerca de continuar con un programa. Por ejemplo, el esquema de transferencia de efectivo condicional, Progresá (después conocido como Oportunidades), que fue introducido a mediados de los 90, ofrecía dinero en efectivo “con la condición de que las familias cumplieran con elementos específicos de co-responsabilidad, como enviar a sus hijos a la escuela en lugar de hacerlos trabajar, brindándoles un suplemento nutricional especialmente elaborado y asistiendo a una clínica para recibir un paquete específico de

intervenciones para la promoción de la salud y la prevención de enfermedades” [12]. Con fines evaluativos, 506 comunidades fueron asignadas aleatoriamente para ingresar al programa inmediatamente o 2 años después [13]. Los hallazgos de esta evaluación del impacto informaron de manera directa las decisiones en políticas de México, persuadiendo al gobierno “no sólo a continuar con el programa sino también a ampliarlo” [12].

4. ¿Cómo debe realizarse la evaluación del impacto?

Atribuir un cambio observado a un programa o política requiere una comparación entre los individuos o grupos expuestos al mismo, y otros que no lo están. También es importante que los grupos comparados sean lo más similares posibles con el fin de descartar las influencias que no sean el programa en sí mismo. Esto puede realizarse de manera efectiva asignando aleatoriamente individuos o grupos de personas (por ejemplo, dentro de áreas geográficas) para que reciban el programa o no, en lo que se denomina un *estudio clínico aleatorizado*. Por lo general, dichos estudios clínicos son realizados como proyectos piloto antes de que un programa sea introducido a nivel nacional. Pero también pueden llevarse a cabo en paralelo con la implementación a escala global, según lo ilustran los ejemplos mexicanos mencionados anteriormente.

Sin embargo, los estudios clínicos aleatorizados no siempre son factibles. Los enfoques alternativos incluyen la comparación de los cambios antes y después de la implementación, con los cambios observados durante el mismo período de tiempo en áreas donde el programa no se haya implementado (por ejemplo, en países o distritos vecinos). Esto se conoce como *evaluación controlada antes y después*. Como alternativa, se puede utilizar una *serie temporal interrumpida* en la que los datos son recolectados desde múltiples puntos de tiempo antes, durante y después de la implementación del programa.

Comparar simplemente el valor de un indicador antes y después de la implementación del programa en general no es recomendable ya que es alto el riesgo de hallazgos confusos – los cambios observados, por ejemplo, en la incidencia de VIH pueden deberse a factores conocidos y desconocidos que no estén relacionados con el programa en sí (ver Figura 2) [14,15].

En el Archivo adicional 2 al final de este artículo, se brinda una descripción general de una serie de diseños de evaluación. Las fortalezas y debilidades de cada método descritas en el Archivo adicional 2 están explicadas en el Archivo adicional 3.

Las evaluaciones de impactos deben planificarse por adelantado a la implementación del programa, en conjunto con las partes interesadas relevantes, incluidos los responsables de la toma de decisiones en políticas. En general, después de que un programa se ha implementado extensamente es demasiado tarde para llevar a cabo medidas iniciales o para establecer los grupos de comparación adecuados. Por ejemplo,

el uso de asignaciones aleatorias para decidir si se incluirán comunidades en un programa o no, no puede realizarse después de que el programa se ha implementado nacionalmente. Es más probable que las evaluaciones de impactos que se incluyen en un programa desde el comienzo produzcan hallazgos válidos y no así aquellas evaluaciones realizadas tardíamente. Además, si las evaluaciones de impactos son consideradas como una parte integrada de la implementación del programa, los responsables de la toma de decisiones en políticas y otras personas pueden comprometerse más a tener en cuenta los hallazgos.

La cantidad de individuos o comunidades requerida para una evaluación de impacto también debe calcularse en una etapa temprana. Esto garantizará que exista un tamaño muestral lo suficientemente grande como para extraer conclusiones significativas de los hallazgos de la evaluación.

En la atención de la salud, como en la mayoría de las otras áreas, los programas deben ser efectivos y costo-efectivos. Para evaluar los aspectos económicos de un programa, se deben calcular los costos y el uso de recursos, basándose preferentemente en los datos recolectados de la implementación real [16]. Por lo tanto, las decisiones sobre qué datos económicos son necesarios recolectar deben tomarse en una etapa temprana, antes de que comience la evaluación.

Las evaluaciones de impactos pueden ser más informativas si se incluye una evaluación del proceso. Una evaluación del proceso puede determinar si la opción de política o programa fue implementada según lo deseado. También puede investigar los procesos de implementación y cambio, analizar las respuestas al programa y explorar las razones para los hallazgos de la evaluación [17].

La Tabla 2 ofrece ejemplos de evaluaciones de impactos.

Las limitaciones en los datos, el tiempo y el presupuesto pueden actuar como trabas para garantizar la implementación rigurosa. Dichas limitaciones pueden afectar la confiabilidad de las evaluaciones de impactos de varias maneras:

- Por ejemplo, al comprometer la validez global de los resultados, debido a la insuficiente planificación o seguimiento, o a través de una escasez de datos iniciales, una confianza en las fuentes de datos inadecuadas o la selección de grupos de comparación inadecuados
- Mediante el uso de muestras inadecuadas, por ejemplo, debido a la selección de muestras que son convenientes para el muestreo pero pueden no ser representativas como resultado de tamaños muestrales demasiado pequeños o por falta de atención suficiente a los factores contextuales

Dichas limitaciones pueden abordarse comenzando el proceso de planificación con anticipación o encontrando maneras para reducir los costos en la recolección de datos. Sin embargo, es importante garantizar que ni las posibles amenazas a la validez de los

resultados ni las limitaciones de la muestra serán tales que los resultados de la evaluación no puedan brindar información confiable. Por lo tanto, antes de realizar una evaluación, debe realizarse un análisis para ver si ésta es posible. De no ser así, se debe llevar a cabo una evaluación para saber si se debe implementar un programa sin evaluación previa, en medio de la incertidumbre sobre sus potenciales impactos [18].

Las evaluaciones de impactos no valen la pena si los hallazgos no son utilizados. Los resultados deben utilizarse para informar decisiones sobre si continuar, cambiar o interrumpir los programas existentes. Por supuesto, también deberán tenerse en cuenta otros intereses. Por ejemplo, los responsables de la toma de decisiones pueden elegir no enfatizar hallazgos específicos de ciertas evaluaciones cuando dichos hallazgos están en conflicto con otros intereses que se consideran más importantes [19]. Sin embargo, es importante evitar la supresión de los hallazgos de las evaluaciones de impactos, por ejemplo por razones políticas. No poder utilizar los hallazgos de la evaluación contradice uno de los principales objetivos de realizar dichas evaluaciones: aprender de la experiencia y compartir el conocimiento que se ha generado. El uso de partes independientes para realizar evaluaciones de impactos puede reducir el riesgo de tener que manipular los hallazgos u ocultárselos al público.

CONCLUSIÓN

Este artículo ha descrito una serie de aspectos relacionados con el monitoreo y la evaluación. Actualmente, muchos esfuerzos de monitoreo y evaluación de programas se hacen utilizando métodos que no producen evaluaciones válidas de la implementación de una política o un programa o estimaciones de efectos válidas. A veces dichas evaluaciones no se realizan. Al considerar los temas descritos en este artículo, los responsables de la toma de decisiones en políticas y quienes los respaldan deben poder desarrollar planes que generarán conocimientos nuevos y directamente útiles.

RECURSOS

Documentos útiles y lectura complementaria

Segone M (ed). Bridging the gap: The role of monitoring and evaluation in evidence-based policy making. UNICEF, the World Bank and the International Development Evaluation Association. www.unicef.org/ceecis/evidence_based_policy_making.pdf

MacKay K. How to Build M&E Systems to Support Better Government. 2007. Washington DC, The World Bank.
www.worldbank.org/ieg/ecd/docs/How_to_build_ME_gov.pdf

Monitoring and Evaluation (M&E): Some Tools, Methods and Approaches. 2004. Washington DC. The World Bank.

[http://lnweb90.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/\\$FILE/MandE_tools_methods_approaches.pdf](http://lnweb90.worldbank.org/oed/oeddoclib.nsf/24cc3bb1f94ae11c85256808006a0046/a5efbb5d776b67d285256b1e0079c9a3/$FILE/MandE_tools_methods_approaches.pdf)

Framework for Managing Programme Performance Information. 2007. National Treasury of South Africa. www.treasury.gov.za/publications/guidelines/FMPI.pdf

Barber S. Health system strengthening interventions: Making the case for impact evaluation. 2007. Geneva, Alliance for Health Policy and Systems Research.

www.who.int/alliance-hpsr/resources/Alliance%20%20HPSR%20-%20Briefing%20Note%202.pdf

Savedoff WD, Levine R, Birdsall N. When will we ever learn? Improving lives through impact evaluation. Report of the Evaluation Gap Working Group. 2006. Washington DC, Center for Global Development.

www.cgdev.org/content/publications/detail/7973/

Grimshaw J, Campbell M, Eccles M and Steen N. Experimental and quasi-experimental designs for evaluating guideline implementation strategies. Family Practice 2000; 17: S11-S18. http://fampra.oxfordjournals.org/cgi/reprint/17/suppl_1/S11

Enlaces a sitios web

Grupo de Evaluación Independiente (Independent Evaluation Group, IEG) del Banco Mundial: www.worldbank.org/ieg – El IEG es una unidad independiente dentro del Banco Mundial. El IEG evalúa qué es efectivo o no con respecto a las opciones de políticas, cómo un prestatario planifica llevar a cabo y mantener un proyecto y la contribución perdurable del Banco al desarrollo global de un país

International Initiative for Impact Evaluation (3ie): www.3ieimpact.org – El 3ie busca mejorar la vida de la gente pobre en países de medianos y bajos ingresos al proveer y resumir evidencia relacionada con qué opciones de políticas funcionan, y también cuándo y porqué, y los costos implicados

Red de la Métrica de Salud (Health Metrics Network, HMN):

www.who.int/healthmetrics/en – La HMN tiene el objetivo estratégico de aumentar la disponibilidad y el uso de la información de salud en tiempo y forma oportuna. Para lograr esto, la HMN identifica estrategias para el desarrollo y el fortalecimiento de HIS, respalda a países para que implementen la reforma de HIS, y aumenta el conocimiento sobre los bienes públicos globales mediante la investigación, la innovación técnica y compartir las lecciones aprendidas

NorthStar: www.rebeqi.org/?pageID=34&ItemID=35 – NorthStar es una herramienta para planificar, realizar y evaluar programas de mejora de calidad

CONFLICTOS DE INTERESES

Los autores declaran que no presentan conflictos de intereses.

CONTRIBUCIONES DE LOS AUTORES

AF preparó el primer borrador. ADO, JNL y SL contribuyeron en la elaboración y revisión de este artículo.

AGRADECIMIENTOS

Consulte la Introducción de esta serie para ver los agradecimientos a los financiadores y los colaboradores. Además, nos gustaría agradecerle a Ruth Levine por los comentarios útiles sobre una versión anterior de este artículo.

REFERENCIAS

1. Lavis JN, Oxman AD, Lewin S, Fretheim A: **SUPPORT Tools for evidence-informed health Policymaking (STP). Introduction.** *Health Res Policy Syst* 2009, **7 (Suppl 1:I1)**.
2. Development Assistance Committee Working Party on Aid Evaluation: *Glossary of Key Terms in Evaluation and Results Based Management*. Paris, OECD Publications. 2002.
3. Harries AD, Gomani P, Teck R, de Teck OA, Bakali E, Zachariah R, et al: **Monitoring the response to antiretroviral therapy in resource-poor settings: the Malawi model.** *Trans R Soc Trop Med Hyg* 2004, **98**:695-701.
4. Jakobsen E, Palshof T, Osterlind K, Pilegaard H: **Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results.** *Eur J Cardiothorac Surg* 2009, **35**:348-52.
5. Smith PC, Mossialos E, Papanicolas I: *Performance measurement for health system improvement: experiences, challenges and prospects*. Background Document for WHO European Ministerial Conference on Health Systems: "Health Systems, Health and Wealth". Copenhagen, World Health Organization, Europe. 2008.
6. Campbell SM, Braspenning J, Hutchinson A, Marshall M: **Research methods used in developing and applying quality indicators in primary care.** *Qual Saf Health Care* 2002, **11**:358-64.
7. MacKay K: *How to Build M&E Systems to Support Better Government*. Washington DC, The World Bank. 2007.

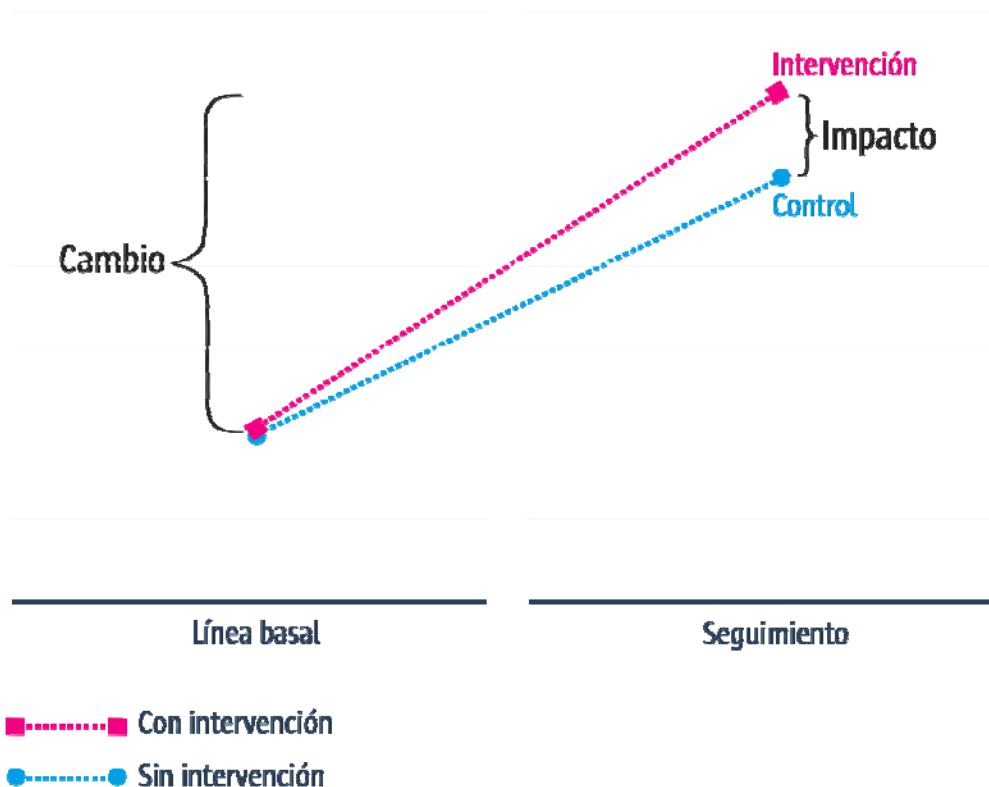
8. The World Bank: *Impact evaluation: Overview*. [<http://go.worldbank.org/2DHMCRFFT2>]. The World Bank. 2009.
9. Moynihan R, Oxman A, Lavis JN, Paulsen E: *Evidence-Informed Health Policy: Using Research to Make Health Systems Healthier*. Rapport nr. 1-2008. Oslo, Nasjonalt kunnskapssenter for helsetjenesten. 2008.
10. Frenk J, Gonzalez-Pier E, Gomez-Dantes O, Lezana MA, Knaul FM: **Comprehensive reform to improve health system performance in Mexico**. *Lancet* 2006, **368**:1524-34.
11. Gakidou E, Lozano R, Gonzalez-Pier E, Abbott-Klafter J, Barofsky JT, Bryson-Cahn C, et al: **Assessing the effect of the 2001-06 Mexican health reform: an interim report card**. *Lancet* 2006, **368**:1920-35.
12. Frenk J: **Bridging the divide: global lessons from evidence-based health policy in Mexico**. *Lancet* 2006, **368**:954-61.
13. Rivera JA, Sotres-Alvarez D, Habicht JP, Shamah T, Villalpando S: **Impact of the Mexican program for education, health, and nutrition (Progresa) on rates of growth and anemia in infants and young children: a randomized effectiveness study**. *JAMA* 2004, **291**:2563-70.
14. Savedoff WD, Levine R, Birdsall N: *When will we ever learn? Improving lives through impact evaluation*. Washington DC, Center for Global Development. 2006.
15. Shadish WR, Cook TD, Campbell DT: *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin; 2002.
16. Oxman AD, Fretheim A, Lavis JN, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 12. Finding and using research evidence about resource use and costs**. *Health Res Policy Syst* 2009, **7 (Suppl 1:S12)**.
17. Lewin S, Glenton C, Oxman AD: **Use of qualitative methods alongside randomised controlled trials of complex healthcare interventions: methodological study**. *BMJ* 2009, **339**:b3496.
18. Oxman AD, Lavis JN, Fretheim A, Lewin S: **SUPPORT Tools for evidence-informed health Policymaking (STP). 17. Dealing with insufficient research evidence**. *Health Res Policy Syst* 2009, **7 (Suppl 1:S17)**.
19. Scheel IB, Hagen KB, Oxman AD: **The unbearable lightness of healthcare policy making: a description of a process aimed at giving it some weight**. *J Epidemiol Community Health* 2003, **57**:483-7.
20. Amuron B, Coutinho A, Grosskurth H, Nabiryo C, Birungi J, Namara G, et al: **A cluster-randomised trial to compare home-based with health facility-based antiretroviral treatment in Uganda: study design and baseline findings**. *Open AIDS J* 2007, **1**:21-7.
21. Jaffar S, Amuron B, Birungi J, Namara G, Nabiryo C, Coutinho A, et al: **Integrating research into routine service delivery in an antiretroviral treatment programme: lessons learnt from a cluster randomized trial comparing strategies of HIV care in Jinja, Uganda**. *Trop Med Int Health* 2008, **13**:795-800.

22. London School of Hygiene and Tropical Medicine: *Home-based HIV care just as effective as clinic-based care in Sub-saharan Africa*. [www.lshtm.ac.uk/news/2009/homeHIVcare.html]. London School of Hygiene and Tropical Medicine, University of London. 2009.
23. Fretheim A, Havelrud K, MacLennan G, Kristoffersen DT, Oxman AD: **The effects of mandatory prescribing of thiazides for newly treated, uncomplicated hypertension: interrupted time-series analysis**. *PLoS Med* 2007, **4**:e232.

Figura 1. Modelo de la cadena de resultados (definiciones adaptadas de [2])



Figura 2. Comparación del cambio en el desempeño en dos áreas: una con intervención y otra sin intervención*



* La Figura muestra que es probable que sea confuso atribuir el cambio de "Línea basal" a "Seguimiento" en respuesta a una intervención. Esto se debe a que, en este caso, también existe una mejora en el "Control". Incluso en cuanto al Control, es incierto si la diferencia entre "Intervención" y "Control" (es decir, "Impacto") puede, de hecho, ser atribuible al programa o a la intervención. Pueden existir otras diferencias entre el lugar de la "Intervención" y del "Control" que podrían haber conducido a la diferencia observada en el indicador medido

Tabla 1. Ejemplos de sistemas de monitoreo en el sistema de atención de la salud

Aumento de la provisión de la terapia antirretroviral (ART) en Malawi [3]

Cuando las autoridades de la salud de Malawi decidieron que la ART estuviera disponible para una mayor parte de la población VIH positiva, se puso en funcionamiento un sistema para monitorear la implementación de esta nueva política. Los principios del sistema se basan en el enfoque de la OMS de monitorear los programas nacionales de tuberculosis. A cada paciente que comienza una ART se le otorga una tarjeta de identidad con un número de identidad único y ésta se conserva en la clínica. La información recolectada de los pacientes nuevos incluye su nombre, domicilio, edad, altura, el nombre del tutor y la razón para comenzar la ART. Durante esta visita se registra su peso y se le pregunta sobre su salud general, estado ambulatorio, trabajo y algún efecto secundario del medicamento. También se realiza y registra el conteo de los comprimidos como una forma de garantizar la adherencia al tratamiento. Además, se registran los siguientes resultados mensuales estandarizados usando estas categorías:

- *Vivo*: El paciente está vivo y ha recolectado su propio suministro de medicamentos de 30 días
- *Muerto*: El paciente ha muerto mientras estaba en la ART
- *Moroso*: El paciente no ha aparecido por un período de 3 meses
- *Interrumpido*: El paciente ha interrumpido por completo el tratamiento, sea por efectos secundarios o por otras razones
- *Transferido*: El paciente ha sido transferido de modo definitivo hacia otro tratamiento

Los datos recolectados como parte del sistema de monitoreo del lanzamiento de la ART pueden ser analizados y utilizados de varias maneras. Se pueden realizar comparaciones de los resultados del tratamiento para pacientes que fueron reclutados en diferentes momentos. Por ejemplo, si la tasa de intercambio de los tratamientos con antibióticos de primera y segunda línea aumenta, o las tasas de mortalidad hacen lo mismo, la causa podría ser un aumento en la resistencia del medicamento al tratamiento de primera línea. Si la tasa de muertes o incumplimientos decae, esto podría indicar que la administración del programa de tratamiento de la ART está mejorando. Si los resultados son particularmente pobres en algunas clínicas o áreas geográficas, deben tomarse acciones para abordar esto.

Cirugía de cáncer de pulmón en Dinamarca [4]

Las autoridades danesas presentaron unas guías de práctica clínica para el manejo del cáncer de pulmón. Éstas fueron impulsadas por los pobres resultados que presentaron los pacientes a los que se les realizó cirugía de cáncer de pulmón. Para monitorear la implementación de las guías, se estableció un registro de pacientes con cáncer de pulmón que incluía información específica sobre aquellos pacientes a los que se les

había realizado una cirugía. Los indicadores seleccionados por el Registro Danés de Cáncer de Pulmón incluyen el alcance (o “etapa”) del cáncer en el cuerpo, el procedimiento quirúrgico utilizado, cualquier complicación que ocurra y el resultado de supervivencia.

Los datos del Registro Danés del Cáncer de Pulmón son utilizados, entre otras razones, para monitorear si se siguen las recomendaciones nacionales para la cirugía de cáncer de pulmón. Las auditorías locales, regionales y nacionales se realizan con el fin de identificar los problemas o las barreras que puedan impedir la adherencia a las guías nacionales. Basadas en estos hallazgos, se proponen estrategias específicas para mejorar la calidad.

Tabla 2. Ejemplos de las evaluaciones de impactos

Terapia antirretroviral (ART) domiciliaria en Uganda [20-22]

La escasez de personal clínico y las dificultades con el acceso a la atención debido a los costos del transporte son los principales obstáculos para aumentar la provisión de ART en países en vías de desarrollo. Una solución propuesta es la atención domiciliaria del VIH, donde la provisión de medicamentos, el monitoreo del estado de la salud y el apoyo a los pacientes es realizado por personal no calificado clínicamente en el domicilio del paciente. Sin embargo, es muy incierto si esta estrategia puede proveer atención de calidad suficiente, incluidas las derivaciones oportunas para atención médica, o si dicho sistema es costo-efectivo. Por lo tanto, antes de implementar ampliamente programas de atención domiciliaria es importante que sean evaluados por su (costo-) efectividad.

Para garantizar una comparación justa entre la ART domiciliaria y en centros de salud, los investigadores en Uganda realizaron un estudio clínico aleatorizado. El área del estudio fue dividido en 44 sub-áreas geográficas distintas. En algunas de éstas, se implementó la atención domiciliaria mientras que en otras se continuó utilizando un sistema convencional en centros de salud. La selección y asignación de las áreas que recibirían, o no, el sistema de atención domiciliario fueron determinadas aleatoriamente. Esto redujo la probabilidad de diferencias importantes entre los grupos de comparación que podrían, de otra manera, haber influenciado al estudio si, por ejemplo, los mismos distritos hubieran decidido si tenían que implementar la atención domiciliaria o si las decisiones se hubieran basado en un estado de preparación existente para implementar la atención domiciliaria. El sistema de asignación aleatoria utilizado también fue la manera más justa para decidir dónde comenzar con la atención domiciliaria, ya que cada distrito tenía una misma probabilidad de ser elegido.

Los investigadores encontraron que el modelo de atención domiciliaria con trabajadores legos capacitados fue tan efectivo como la atención en las clínicas conducida por médicos.

Uso obligatorio de tiazidas para la hipertensión en Noruega [23]

Como medida para la contención de los costos, los responsables para la toma de decisiones en políticas en Noruega decidieron que las tiazidas serían prescritas como medicamentos antihipertensivos, en lugar de otras alternativas más costosas, en aquellos casos en que los gastos por medicamentos fueran reembolsados. La política fue implementada a nivel nacional unos meses después de que se tomó la decisión. Debido a que la crítica continuó argumentando que no era probable que la nueva política condujera a los resultados esperados, el Ministerio de Salud financió un estudio para evaluar el impacto de la política que habían implementado.

La prescripción obligatoria de tiazidas para tratar la hipertensión fue implementada en Noruega con tal urgencia que hizo que no fuera posible realizar una evaluación del impacto planificada y rigurosa. Sin embargo, al acceder a los registros médicos electrónicos de 61 clínicas en una etapa posterior, los investigadores extrajeron los datos de la prescripción que abarcaba desde un año antes hasta un año después de que se introdujo la nueva política. Además, analizaron los datos con una serie de tiempo interrumpido. Se analizaron a través del tiempo las tasas mensuales de tiazidas prescritas y otros resultados de interés para ver si cualquier cambio significativo podría ser atribuido a la política implementada. Los análisis indicaron que existía un aumento definido en el uso de tiazidas (de un 10% a un 25% sobre un período de transición de tres meses pre-determinado), luego del cual el uso de tiazidas se estabilizó.

Archivo adicional 2. Diseños de evaluación (adaptado del Manual Cochrane de Revisiones Sistemáticas de Intervenciones*)

| | |
|--|---|
| Estudio clínico controlado aleatorizado | <ul style="list-style-type: none">• Estudio experimental en el que los individuos son asignados aleatoriamente para recibir diferentes intervenciones (por ejemplo, mediante el lanzamiento de una moneda o una lista de números aleatorios generados por computadora) |
| Estudio clínico aleatorizado por grupos (<i>clusters</i>) | <ul style="list-style-type: none">• Estudio experimental en el que los grupos de personas (por ejemplo, aulas de una escuela u hospitales) son asignados aleatoriamente a recibir diferentes intervenciones |
| Estudio clínico controlado no aleatorizado | <ul style="list-style-type: none">• Estudio experimental en el que las personas están asignadas a diferentes intervenciones usando métodos que no son aleatorios (por ejemplo, pacientes ingresados durante la Semana 1 reciben la Intervención A; aquellos ingresados durante la Semana 2 reciben la Intervención B; aquellos ingresados durante la Semana 3 reciben la Intervención C, y así sucesivamente) |
| Estudio controlado antes y después | <ul style="list-style-type: none">• Estudio en el cual se hacen observaciones antes y después de la implementación de una intervención, tanto en un grupo que recibe la intervención como en un grupo que no la recibe. La recolección de datos en general debería realizarse al mismo tiempo en los dos grupos |
| Estudio de series temporales interrumpidas | <ul style="list-style-type: none">• Estudio que utiliza observaciones en múltiples puntos de tiempo antes y después de una intervención. Las <i>mediciones</i> son <i>interrumpidas</i> por la intervención. El diseño intenta detectar si una intervención tuvo un efecto significativamente mayor que cualquier tendencia subyacente en el tiempo |
| Estudio controlado históricamente | <ul style="list-style-type: none">• Estudio que compara un grupo de participantes que reciben una intervención con un grupo similar del pasado que no la recibió |
| Estudio de cohorte | <ul style="list-style-type: none">• Estudio en el cual un grupo definido de personas (la cohorte) es seguido a lo largo del tiempo para examinar las asociaciones entre las diferentes intervenciones recibidas y resultados posteriores. Un estudio de cohorte <i>prospectivo</i> recluta participantes antes de cualquier intervención y los sigue en el futuro. Un estudio de cohorte <i>retrospectivo</i> identifica participantes de |

registros pasados describiendo las intervenciones recibidas y siguiéndolas desde el momento de dichos registros

Estudio de casos y controles

- Estudio que compara personas con resultados de interés específicos (*casos*) con personas de la misma población fuente pero sin dicho resultado (*controles*), para examinar la asociación entre el resultado y la exposición previa (por ejemplo, recibir una intervención). El diseño es particularmente útil cuando el resultado es poco frecuente

Estudio transversal

- Estudio que recoge información sobre intervenciones pasadas o presentes y resultados de salud actuales para un grupo de personas en un punto de tiempo particular. Este tipo de estudio examina asociaciones entre los resultados y la exposición a las intervenciones

Estudio cualitativo

- Estudio realizado en un entorno natural que en general está diseñado para interpretar o darle sentido a los fenómenos en términos de los significados que las personas les asignan. Típicamente en un estudio como este, los datos narrativos son recogidos a partir de individuos o grupos de “informantes” o de documentos. Luego, los investigadores los interpretan

* Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration and John Wiley & Sons Ltd.; 2008

Archivo adicional 3. Fortalezas y debilidades de los diseños de evaluación

| | Fortalezas | Debilidades |
|--|--|--|
| Estudio clínico controlado aleatorizado | <ul style="list-style-type: none"> Considerado ampliamente como el diseño más fuerte para establecer las relaciones causa-efecto, que es el centro de impacto clave de evaluación | <ul style="list-style-type: none"> Puede consumir demasiado tiempo y representar desafíos logísticos. Los resultados no son necesariamente transferibles a lugares fuera del lugar del estudio |
| Estudio clínico aleatorizado por grupos (<i>clusters</i>) | <ul style="list-style-type: none"> Las mismas fortalezas que para los estudios clínicos aleatorizados comunes. Además, el riesgo de la “contaminación” se reduce por ejemplo a que la intervención A puede ser recibida por, o afectar a, las personas asignadas para recibir intervención B solamente. Por ejemplo, si las enfermeras están asignadas aleatoriamente para implementar una nueva rutina, otras enfermeras pueden estar influenciadas por estos cambios y pueden comenzar a realizar las mismas actividades. Por lo tanto, puede ser mejor aleatorizar guardias, y todo el personal dentro de ellas, en lugar de enfermeras individuales | <ul style="list-style-type: none"> Las diferencias iniciales pueden ser un problema ya que el número de unidades (o <i>clusters</i>) que son aleatorizadas en general sería menor que en un estudio clínico donde se aleatorizan individuos Pueden consumir más tiempo y ser logísticamente más desafiantes, pero menos que un estudio clínico aleatorizado común |
| Estudio clínico controlado no aleatorizado | <ul style="list-style-type: none"> Puede ser más fácil y más práctico de realizar que un estudio clínico controlado aleatorizado | <ul style="list-style-type: none"> Cuando la asignación no se realiza a través de métodos de aleatorización, pueden ocurrir sesgos de selección, |

| Fortalezas | | Debilidades |
|---|--|--|
| | | <p>porque los pacientes y los trabajadores de la salud ajustan su comportamiento al procedimiento de asignación si prefieren una intervención sobre otra</p> |
| Estudio controlado antes y después | <ul style="list-style-type: none"> • Puede ser la única opción práctica, por ejemplo, para intervenciones a gran escala donde la aleatorización no es factible para razones políticas o prácticas | <ul style="list-style-type: none"> • Las diferencias conocidas o desconocidas entre los grupos que son comparados pueden ejercer mayor influencia sobre los hallazgos que el hecho de que reciban intervenciones diferentes. Consecuentemente, extraer conclusiones sobre relaciones de causa-efecto puede ser riesgoso • Requiere disponibilidad de los datos iniciales |
| Estudio de series temporales interrumpidas | <ul style="list-style-type: none"> • Puede ser factible y relativamente fácil realizar este estudio si se dispone de los datos necesarios. No es necesario un grupo control | <ul style="list-style-type: none"> • La magnitud del efecto siempre es difícil de estimar en tales análisis porque influencias distintas de la intervención que se investiga pueden impactar en los cambios observados |
| Estudio controlado históricamente | <ul style="list-style-type: none"> • Puede realizarse rápida y fácilmente si se cuenta con los datos necesarios | <ul style="list-style-type: none"> • Las diferencias conocidas o desconocidas entre los grupos que son comparados pueden ejercer mayor influencia sobre los hallazgos que el hecho de que reciban intervenciones diferentes. Consecuentemente, extraer conclusiones sobre relaciones de causa-efecto es riesgoso |
| Estudio de | <ul style="list-style-type: none"> • A menudo, son los | <ul style="list-style-type: none"> • Los estudios de cohorte son |

| | Fortalezas | Debilidades |
|-------------------------------------|--|--|
| cohorte | estudios grandes con alto nivel de validez externa (es decir, los hallazgos pueden ser generalizados). A menudo, se realizan durante varios años, lo que hace posible detectar efectos a largo plazo de una intervención | típicamente extensos y costosos, principalmente debido a la necesidad de seguimiento del (normalmente) alto número de participantes <ul style="list-style-type: none"> Las diferencias conocidas o desconocidas entre los grupos que son comparados pueden ejercer mayor influencia sobre los hallazgos que el hecho de que fueron expuestas a intervenciones diferentes. Consecuentemente, extraer conclusiones sobre relaciones de causa-efecto es riesgoso |
| Estudio de casos y controles | <ul style="list-style-type: none"> Se realiza más rápida y fácilmente que los estudios de cohortes | <ul style="list-style-type: none"> La naturaleza retrospectiva de tales estudios implica la recolección de información sobre eventos que ocurrieron con anterioridad. Estas demoras de tiempo pueden ser una fuente de error Las diferencias conocidas o desconocidas entre los grupos que son comparados pueden ejercer mayor influencia sobre los hallazgos que el hecho de que reciban intervenciones diferentes. Consecuentemente, extraer conclusiones sobre relaciones de causa-efecto es riesgoso |

| | Fortalezas | Debilidades |
|----------------------------|--|---|
| Estudio transversal | <ul style="list-style-type: none"> • No requiere tiempo de seguimiento y, por lo tanto, puede ser realizado rápidamente y, con frecuencia, a bajo costo | <ul style="list-style-type: none"> • Las diferencias conocidas o desconocidas entre los grupos que son comparados pueden ejercer mayor influencia sobre los hallazgos que el hecho de que reciban intervenciones diferentes. Consecuentemente, extraer conclusiones sobre relaciones de causa-efecto es riesgoso |
| Estudio cualitativo | <ul style="list-style-type: none"> • Permite la recolección de información más profunda que otros diseños cuantitativos. Permite un entendimiento de cómo funcionan (o no) las intervenciones y los programas | <ul style="list-style-type: none"> • No genera datos que pueden ser utilizados para estimar el efecto de una intervención que está más allá de la percepción de aquellos a quienes se les realiza la entrevista o la encuesta |