

Evaluación de las intervenciones sanitarias en salud global

Métodos avanzados

Bajo la dirección de
Valéry Ridde y Christian Dagenais



Traducido del francés por
Clara Bermúdez-Tamayo, Jaime Jiménez Pernet,
Olga Leralta Piñan y Alberto Fernández Ajuria

esbc
éditions science
et bien commun

Éditions

Evaluación de las intervenciones sanitarias en salud global

Evaluación de las intervenciones sanitarias en salud global

Métodos avanzados

BAJO LA DIRECCIÓN DE VALÉRY RIDDE Y
CHRISTIAN DAGENNAIS

ÉDITIONS SCIENCE ET BIEN COMMUN Y IRD ÉDITIONS



Excepto cuando se especifiquen otros términos, Evaluación de las intervenciones sanitarias en salud global por Valéry Ridde y Christian Dagenais se distribuye bajo una Licencia Creative Commons Atribución-CompartirIgual 4.0 Internacional.

Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados.

Autores: Bajo la dirección de Valéry Ridde y Christian Dagenais

Diseño de la portada: Kate McDonnell, fotografía de Christian Dagenais

Traducción: Clara Bermúdez-Tamayo, Jaime Jiménez Pernet, Olga Leralta Piñan y Alberto Fernández Ajuria, Según Afolabi por la traducción de los resúmenes en inglés.

ISBN PDF : 978-2-924661-98-7

ISBN Print : 978-2-925128-04-5

Depósito legal – Bibliothèque et Archives nationales du Québec 2020

Depósito legal – Biblioteca y Archivo Nacional de Canadá

Depósito legal: 4er trimestre de 2020

Este libro está bajo una licencia Creative Commons CC-BY-SA 4.0.

Éditions science et bien commun

<http://editionscienceetbiencommun.org>

1085 avenue de Bourlamaque Québec (Québec) G1R 2P4 Canada

IRD Éditions, 44 bd de Dunkerque, 13572 Marseille cedex 02 France

Difusión: info@editionscienceetbiencommun.org

Contenido

Introducción	1
Valéry Ridde y Christian Dagenais	
Parte I. La fase pre-evaluativa y la sostenibilidad	
1. El estudio de la evaluabilidad	13
<i>Una intervención de prevención de consumo de drogas en la escuela en Quebec</i>	
Biessé Diakaridja Soura, Jean-Sébastien Fallu, Robert Bastien y Frédéric N. Brière	
2. Evaluación de la sostenibilidad	43
<i>Una intervención de financiación basada en los resultados en Malí</i>	
Mathieu Seppey y Valéry Ridde	
Parte II. Las aproximaciones cualitativas y participativas	
3. Evaluación cualitativa, informatizada, participativa e inter-organizacional (EQUIPO)	69
<i>Ejemplo de un programa para mujeres víctimas de la violencia en Bolivia</i>	
Mathieu Bujold y Jean-Alexandre Fortin	
4. El método foto voz	95
<i>Una intervención con poblaciones marginadas para el acceso al agua potable, la higiene y el saneamiento en México</i>	
Lynda Rey, Wilfried Affodégon, Isabelle Viens, Hind Fathallah y Maria José Arauz	

5. El análisis de un proyecto de investigación-acción 135
Combinación de enfoques en el sector de la salud en Burkina Faso
Aka Bony Roger Sylvestre, Valéry Ridde y Ludovic Queuille

Parte III. Métodos mixtos

6. Revisiones sistemáticas mixtas 163
Un ejemplo sobre la financiación basada en los resultados
Quan Nha Hong, Anne-Marie Turcotte-Tremblay y Pierre Pluye
7. Integración en métodos mixtos 193
Marco conceptual para la integración de fases, resultados y datos cualitativos y cuantitativos
Pierre Pluye
8. La práctica de la integración en métodos mixtos 219
Las múltiples combinaciones de las estrategias de integración
Pierre Pluye, Enrique García Bengoechea, David Li Tang y Vera Granikov

Parte IV. Evaluación de la efectividad y la eficiencia

9. Métodos cuasi-experimentales 247
El efecto de la edad mínima legal sobre el consumo de alcohol entre los y las jóvenes en los Estados Unidos
Tarik Benmarhnia y Daniel Fuller
10. Ensayos aleatorizados por conglomerados 275
Ejemplo dirigido a la salud materna e infantil
Alexandre Dumont
11. Medición de equidad 295
Un ejemplo de intervención de gratuidad de la atención obstétrica
Tarik Benmarhnia y Britt McKinnon

12.	Análisis de coste-efectividad	323
	<i>Una intervención para descentralizar la atención del VIH/ SIDA en Shiselweni (Swazilandia)</i>	
	Guillaume Jouquet	
13.	Análisis espacial	345
	<i>Un caso de intervención comunitaria para controlar el mosquito Aedes aegypti en Burkina Faso</i>	
	Emmanuel Bonnet, Samiratu Ouédraogo y Diane Saré	
14.	Análisis coste efectividad	363
	<i>Evaluación de la eficiencia de un programa para incrementar el número de partos vaginales realizados en situaciones de feto en posición no cefálica</i>	
	Leticia Garcia Mochon y Clara Bermúdez-Tamayo	
15.	Ensayo clínico aleatorizado por conglomerados	381
	<i>Diseño para evaluar la efectividad de una intervención en población desfavorecida para mejorar el autocuidado en diabetes adaptado a un contexto de restricción de recursos</i>	
	Eva Martín Ruiz y Antonio Olry de Labry Lima	
Parte V. Evaluación de los procesos y fidelidad de implementación		
16.	Análisis de los procesos de implementacion	401
	<i>Una intervención compleja en Burkina Faso: financiación basada en los resultados</i>	
	Valéry Ridde y Anne-Marie Turcotte-Tremblay	
17.	Evaluación de la fidelidad de implantación	431
	<i>Un proyecto para distribuir el carbón ártico a las mujeres embarazadas en Nunavik</i>	
	Lara Gautier, Catherine M. Pirkle, Christopher Furgal y Michel Lucas	

18. Evaluación de la fidelidad y la adaptación	451
<i>Un ejemplo de la puesta en marcha de las intervenciones en salud global</i>	
Dennis Pérez, Marta Castro y Pierre Lefèvre	
19. Evaluación realista	479
<i>El ejemplo de la adopción de una política de salud pública en Benín</i>	
Jean-Paul Dossou y Bruno Marchal	
Acerca de la casa editorial	509

Introducción

VALÉRY RIDDE Y CHRISTIAN DAGENAI

Hace diez años, nos enfrentábamos al hecho de que a menudo teníamos la dificultad de encontrar textos para apoyar nuestros intercambios pedagógicos al enseñar la evaluación de programas a las personas francófonas. En esa época, las bibliotecas estaban llenas de obras de referencia en inglés que a menudo utilizábamos por sus bondades pedagógicas (Patton 1997; Weiss 1998; Chen 2005), pero siempre nos enfrentábamos a retos lingüísticos. Por eso nos embarcamos en nuestra primera aventura editorial con la publicación en 2009 de una obra pedagógica en francés para formar a estudiantes y profesionales en evaluación de programas. Más allá del lenguaje, la originalidad del libro fue su enfoque multidisciplinar e intersectorial. De hecho, tras presentar propuestas de síntesis conceptual (incluidos textos de importantes autores de habla inglesa que habíamos traducido), el libro ofrecía numerosos ejemplos de evaluación de intervenciones. Estos ejemplos se ubicaban en varios sectores (educación, salud, acción social, etc.) y mostraban de manera didáctica cómo se podían movilizar varios métodos (cuantitativos, cualitativos, mixtos) y disciplinas para responder a preguntas de evaluación útiles para los profesionales y los/as responsables de la toma de decisiones. Este primer libro fue apreciado por lectores/as y profesores/as y sentimos la necesidad de publicar una edición revisada y ampliada en 2012 (Ridde y Dagenais 2012). Además, fue traducido al español y distribuido en América Latina por la editorial de una universidad colombiana, que deseaba difundir mejor estos planteamientos en ese continente, confirmando la necesidad de obras pedagógicas en idiomas distintos del inglés.

Pero a lo largo de los años y de nuestra experiencia de investigación y de enseñanza en los países del Sur, hemos sentido una triple necesidad que está en el origen de este trabajo colectivo.

Por una parte, observamos la necesidad de seguir publicando en francés, dadas las continuas dificultades a las que se enfrentan los/as estudiantes y colaboradores que sólo trabajan en un entorno francófono y que no dominan el inglés. Si bien desde 2009 se han publicado algunos libros en francés en el ámbito de la evaluación (Larivière y Corbière 2014; Hurteau, Houle

y Guillemette 2012), ninguno de ellos tenía una especificidad propia del contexto del Sur, o más precisamente de los países de bajos ingresos, con el propósito de evitar toda confusión de uniformidad contextual. En 2016, cuando ya habíamos iniciado la producción de este libro, se publicó otro libro específico sobre la práctica de la evaluación en África, pero se centró menos en los métodos que en las prácticas reflexivas (Ridde *et al.* 2016). Volveremos a esta especificidad contextual, pero aquí queremos mostrar la importancia de poner a disposición escritos franceses que se refieren a un campo que todavía está dominado por el mundo anglosajón.

Por otra parte, si bien nuestro primer trabajo es, a nuestro juicio, una muy buena introducción a la evaluación de las intervenciones para los neófitos en este campo, nos quedó claro que necesitábamos profundizar en los límites de este trabajo y entrar en la complejidad de los métodos avanzados. Sin embargo, también nos quedó claro desde el comienzo que este tipo de trabajos interdisciplinarios en francés y en salud global no existían. Sin embargo, especialmente para quienes participan en procesos de evaluación rigurosos que van más allá de las consultorías para dar respuesta a solicitudes de una organización-cliente (que también es obviamente un mandato importante), se hace esencial dominar los métodos complejos para responder a preguntas de evaluación complejas que se dirigen a programas igualmente complejos.

Esta complejidad hace imposible en muchos casos establecer lo que algunos economistas del desarrollo consideran la norma del análisis de causalidad, es decir, estudios aleatorios controlados, que en la mayoría de los casos son insuficientes para responder a las preguntas planteadas en el contexto del Sur. Esta complejidad también plantea importantes problemas de evaluación en el caso de las intervenciones que reúnen características de complejidad (Robert y Ridde 2014). Sin embargo, es evidente que es necesario movilizar métodos avanzados para esos análisis y, por lo tanto, presentarlos. Además, creemos que el contexto de los países de bajos ingresos es favorable a la innovación metodológica y al uso de estos métodos.

No volveremos sobre los interminables debates en torno al concepto de salud global (Taylor 2018), ya que queremos centrarnos en el hecho de que forma parte de un contexto de menos recursos y fuertes desigualdades, pero también de cuestiones de poder exacerbadas por la fuerte presencia de asistencia pública al desarrollo por parte de países de altos ingresos y la colaboración internacional en la práctica de la evaluación. Dado que

los contextos son omnipresentes e influyen en todas las “etapas” de una evaluación (Craig *et al.* 2018), es esencial tenerlos en cuenta al seleccionar métodos científicamente, pero también contextualmente relevantes. Así pues, contrariamente a la creencia popular, es en un contexto de escasez de recursos donde los evaluadores deben ser más ágiles e innovadores, y tratar de desplegar métodos apropiados para lograr sus objetivos evaluativos. Los capítulos de este libro son ejemplos excelentes de esta agilidad y de cómo se pueden desplegar enfoques complejos para evaluar las intervenciones. No se ha esperado a los recientes intentos de crear un nuevo concepto en torno a la investigación de las intervenciones de salud pública para evaluar rigurosamente las intervenciones de salud global. Los autores y autoras que hemos movilizad para construir este libro demuestran este dinamismo metodológico, incluso en un contexto en el que los medios disponibles para la evaluación son muy inferiores a los que hay en el Norte. Pero la difusión de estas innovaciones, para parafrasear a Rogers, sigue estando demasiado a menudo confinada a unos pocos expertos... especialmente a los expertos anglófonos que tienen los medios para acceder a ellas.

Por lo tanto, y finalmente, la tercera necesidad que hemos identificado y que este libro intenta abordar, es la de la libre difusión del conocimiento, que es un tema importante en la salud global (Smith *et al.* 2017). ¿De qué sirve desarrollar todas estas innovaciones metodológicas si siguen siendo confinadas y monopolizadas por personas que dominan el inglés y tienen los medios para acceder a ellas? ¿Cómo podemos asegurarnos de que los estudiantes francófonos radicados en el Sur tengan acceso a conocimientos originales y útiles cuando no pueden pagar los libros, aunque puedan pagarlos por Internet? ¿Cómo podemos apoyar el desarrollo de la red de evaluadores francófonos emergentes? Aquí es donde nos atrajo nuestro encuentro con el proyecto de las *Éditions science et bien commun* y su deseo de democratización del conocimiento y justicia social del saber. Ya habíamos podido publicar uno de nuestros libros sobre nociones preconcebidas en materia de salud global en versión gratuita gracias a una iniciativa de las *Presses de l'Université de Montréal* (Ridde y Ouattara 2015). También habíamos subvencionado una reproducción de la primera edición de nuestro libro introductorio sobre la evaluación que se publicaría en el África occidental a tarifa reducida, en colaboración con *Éditions L'Harmattan Burkina Faso*. Pero estas ayudas son temporales y realmente queríamos que nuestro nuevo libro fuera lo más accesible posible. La comunidad de

personas que realizan evaluaciones en el mundo francófono está creciendo en todo el mundo, todos los países del África francófona ofrecen al menos una maestría en salud pública donde se está empezando a enseñar la evaluación... aquí hay tantas oportunidades como este libre acceso a métodos avanzados lo haga útil, creemos.

Así, este libro está destinado a todos los amantes de la evaluación (¡sí, sí, existen!) que deseen desarrollar sus conocimientos y habilidades en métodos avanzados. Está destinada principalmente, pero no exclusivamente, a las personas que trabajan en un contexto de salud global. En efecto, si los casos presentados se refieren al ámbito de la salud, en el sentido más amplio del término y en particular, a las poblaciones, es evidente que los métodos que presentamos pueden extenderse a otros ámbitos (educación, economía, etc.) y en contextos distintos de los del Sur. Por lo tanto, todos se beneficiarán y podrán utilizar su pericia para adaptar y trasladar estos métodos a otros lugares, en otros campos y en otras circunstancias.

El plan del libro

El libro se compone de cinco secciones, cada una de las cuales representa una dimensión metodológica relevante.

La primera sección es muy innovadora, ya que presenta métodos raramente discutidos en los libros y artículos de evaluación en francés. Nuestro enfoque general forma parte de un enfoque de evaluación basado en la teoría de las intervenciones (Chen 2005; Weiss 1998). En este sentido, se hace imprescindible referir lo que Beaudry y Gauthier llaman la fase de pre-evaluación (Beaudry y Gauthier, 1992), es decir, el análisis de la “evaluabilidad” de las intervenciones. En el capítulo 1 (Soura y colaboradores) se presentan las diversas etapas del proceso de un estudio de “evaluabilidad”. El caso presentado trata de un programa de prevención en la escuela y tiene como objetivo determinar el grado en que este programa es evaluable.

La sostenibilidad de las intervenciones es también un tema relativamente poco abordado en la literatura de lengua francesa. Cuando se trata, a menudo se trata como un tema marginal. Sin embargo, las investigaciones nos han dicho desde hace mucho tiempo que la sostenibilidad debería ser la primera, y no la última preocupación de los profesionales al planificar una intervención (Pluye y otros, 2005). Por lo

tanto, decidimos incluirlo en la primera parte, ya que el primer autor de este capítulo sobre una intervención en Malí se quejó de que quedaba relegado, al igual que la sostenibilidad, al final del libro. Tenía razón y le agradecemos su petición. La sostenibilidad de las intervenciones en África está muy poco estudiada. Este capítulo presenta una metodología para hacerlo (Seppey y Ridde, Capítulo 2).

La segunda sección consta de tres capítulos en los que se presentan los enfoques de evaluación mediante métodos cualitativos. Aunque este uso no es innovador en el campo de la salud global (Belaid et al. 2016), hemos querido mostrar algunos enfoques originales que no se utilizan ampliamente en francés. Por lo tanto, estos tres capítulos proponen procesos interesantes en los que lo cualitativo se moviliza de manera rigurosa. El primero (Bujold y Fortin, capítulo 3) trata de la evaluación participativa de un programa destinado a la igualdad de género en Bolivia. Utiliza un método sistemático con tres componentes (cualitativo, participativo e interorganizativo). El segundo capítulo de esta sección (Rey y colaboradores, Capítulo 4) describe una evaluación realizada en México que utiliza otro enfoque cualitativo participativo que utiliza la fotografía como medio de expresión: *photo voice*. Se centra en las percepciones de los jóvenes indígenas sobre su participación en un programa basado en las artes para promover el cambio de comportamiento. Por último, el tercer capítulo de esta sección (Aka y colaboradores, capítulo 5) trata de la evaluación de un enfoque de investigación de la acción en Burkina Faso. El enfoque de evaluación adoptado combina un análisis externo con un análisis reflexivo para producir resultados que sirvan de guía para futuras estudios de investigación-acción.

La tercera sección ofrece la oportunidad de cumplir nuestro deseo de difundir más ampliamente el uso de métodos mixtos para la evaluación (Bujold y otros, 2018). En tres capítulos se muestra cómo realizar revisiones mixtas, así como la integración de datos que es fundamental para la pertinencia de estos métodos. Algunos ejemplos útiles en salud global facilitan esta demostración. El primero de estos capítulos (Hong y colaboradores, Capítulo 6) se centra en revisiones mixtas que combinan estudios cuantitativos, cualitativos y mixtos utilizando un enfoque sistemático. Se presenta un ejemplo de una revisión mixta de un proyecto de financiación basado en los resultados en Burkina Faso; se presentan cuatro diseños de síntesis. En el segundo capítulo de esta sección (Pluye, Capítulo 7) se propone un marco conceptual que incluye todos los tipos de

integración de los métodos cualitativos y cuantitativos, así como las diversas combinaciones posibles de estas estrategias. El siguiente capítulo (Pluye y otros, Capítulo 8) ilustra varios de estos tipos de combinaciones utilizando estudios de salud global.

La cuarta sección está dedicada a la evaluación de la efectividad y la eficiencia de las intervenciones. En los tres primeros capítulos de esta sección se explican los enfoques metodológicos para utilizar métodos estadísticos complejos con el fin de verificar si se han alcanzado los objetivos de las intervenciones. El primero (Benmarhnia y Fuller, Capítulo 9) presenta dos métodos cuasi-experimentales aplicables cuando los ensayos aleatorios controlados son imposibles de organizar: el método de la Diferencia en Diferencia (DD) y el enfoque de Regresión con Discontinuidad (RD). La aplicación de estos dos métodos se utiliza para medir los efectos de la edad mínima legal de consumo de alcohol en el consumo de los jóvenes. En el segundo capítulo de esta sección (Benmarhnia y McKinnon, capítulo 10) se presentan diferentes métodos para medir las desigualdades en materia de salud de una intervención de exención de tasas. Estas medidas incluyen el índice de concentración y el índice de Theil. En el tercer capítulo de la sección (Dumont, capítulo 11) se presentan los modelos estadísticos apropiados y las dificultades metodológicas relacionadas con los ensayos aleatorios por conglomerados utilizados para evaluar el impacto de una intervención para reducir la mortalidad materna y neonatal en los hospitales de Malí y el Senegal. En el cuarto capítulo de esta sección (Jouquet, capítulo 12) se presenta un análisis de la eficacia en función de los costos de una intervención en materia de VIH en Swazilandia, y además se presentan los análisis espaciales (Bonnet, capítulo 13). Finalmente, en esta versión en español, se adjuntan dos nuevos capítulos; la evaluación de la eficiencia de un programa para incrementar el número de partos vaginales mediante un estudio de coste efectividad (García Mochon y Bermúdez-Tamayo, capítulo 14) y nuevamente se describe un ensayo clínico aleatorizado por conglomerados, en este caso para evaluar la efectividad de una intervención para mejorar el autocuidado en diabetes adaptado a un contexto de restricción de recursos (Martin y Olry de Labry-Lima, capítulo 15)

Por último, en la quinta sección se presentan varios enfoques, todos ellos centrados en el análisis de los procesos de aplicación y la fidelidad en una variedad de contextos, desde Nunavik hasta Cuba y desde el África occidental hasta el África meridional. En lo que respecta a los métodos,

esta última sección es innovadora al proponer enfoques que rara vez se presentan o utilizan en la actualidad: un estudio de casos múltiples mediante un proceso de análisis piramidal (Ridde y Turcotte-Tremblay, capítulo 16), la evaluación de la fidelidad (Pérez y colaboradores, capítulo 17; Gauthier y colaboradores, capítulo 18) o el enfoque realista (Dossou y Marchal, capítulo 19).

No podríamos terminar esta introducción sin agradecer a todos los protagonistas de este libro. En primer lugar, los numerosos autores y autoras y sus colaboradores/as que cumplieron pacientemente con nuestros requisitos editoriales didácticos y aceptaron participar en una publicación cuyo tiempo de producción fue proporcional a su motivación. En segundo lugar, a Linda François, que nos ayudó con el formato de los manuscritos de los capítulos y que hizo un seguimiento de los autores y de las autoras para cumplir con los requisitos editoriales. Una de sus peticiones fue que se hicieran resúmenes de los capítulos en el mayor número posible de idiomas. Por lo tanto, debemos agradecer a aquellos que tuvieron la amabilidad de ayudarnos con estas traducciones. Finalmente, debemos agradecer a *Éditions science et bien commun* por su intención de publicar este libro en acceso abierto y todo su apoyo durante su producción, así como el apoyo financiero del *Institut de Recherche pour le Développement (IRD)* y el Equipo RENARD.

Referencias

- Beaudry, J. et Gauthier, B. (1992). L'évaluation de programme. Dans B. Gauthier (dir.), *Recherche sociale: de la problématique à la collecte de données* (2^e éd., p. 425-452). Québec : Presses de l'Université du Québec.
- Belaid, L., Bodson, O., Ridde, V., Samb, O. et Turcotte-Tremblay (2016). Innovations qualitatives dans l'évaluation des interventions en santé. Dans J. Kivits, F. Balard, C. Fournier et M. Winance (dir.), *Les recherches qualitatives en santé*. Malakoff : Armand Colin.
- Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117^e Cahiers scientifiques de l'ACFAS.

- Chen, H.-T. (2005). *Practical program evaluation: assessing and improving planning, implementation, and effectiveness*. Thousand Oaks, CA: Sage Publications.
- Craig, P., Di Ruggiero, E., Frohlich, K. L., Mykhalovskiy, E. et White, M., on behalf of the Canadian Institutes of Health Research (CIHR)–National Institute for Health Research (NIHR) Context Guidance Authors Group (listed alphabetically), R. Campbell et al. (2018). *Taking account of context in population health intervention research: guidance for producers, users and funders of research*.
<https://doi.org/10.3310/CIHR-NIHR-01>
- Hurteau, M., Houle, S. et Guillemette, F. (dir.). (2012). *L'évaluation de programme axée sur le jugement crédible*. Québec : Presses de l'Université du Québec.
- Larivière, N., et Corbière, M. (dir.). (2014). *Méthodes qualitatives, quantitatives et mixtes : dans la recherche en sciences humaines, sociales et de la santé*. Québec : Presses de l'Université du Québec.
- Patton, M. Q. (1997). *Utilization-focused evaluation* (3^e éd.). London: Sage Publications.
- Pluye, P., Potvin, L., Denis, J.-L., Pelletier, J. et Mannoni, C. (2005). Program sustainability begins with the first events. *Evaluation and program planning*, 28, 123-137.
- Ridde, V. et Dagenais, C. (dir.). (2012). *Approches et pratiques en évaluation de programme* (2^e éd.). Montréal : Presses de l'Université de Montréal.
- Ridde, V., Kouanda, S., Kobiané, J.-F. et Zongo, S. (dir.). (2016). *Pratiques et méthodes d'évaluation en Afrique*. Paris : L'Harmattan.
- Ridde, V. et Ouattara, F. (2015). *Des idées reçues en santé mondiale*. Montréal : Presses de l'Université de Montréal.
- Robert, E. et Ridde, V. (2014). L'approche réaliste pour l'évaluation de programmes et la revue systématique: de la théorie à la pratique. *Mesure et évaluation en éducation*, 36(3), 79-108.
- Smith, E., Haustein, S., Mongeon, P., Shu, F., Ridde, V. et Larivière, V. (2017). Knowledge sharing in global health research – the impact, uptake and

cost of open access to scholarly literature. *Health Research Policy and Systems*, 15: 73

<https://doi.org/10.1186/s12961-017-0235-3>

Taylor, S. (2018). "Global Health": meaning what? *BMJ Global Health*, 3(2), e000843.

Weiss, C. H. (1998). *Evaluation: methods for studying programs and policies* (2^e éd.). Upper Saddle River, NJ : Prentice Hall.

Cita

Ridde, Valéry y Christian Dagenais (2020). Introducción. In *Evaluación de las intervenciones sanitarias en salud global*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 1-9. Québec : Éditions science et bien commun y Marseille : IRD Éditions.

PARTE I

LA FASE PRE-EVALUATIVA Y LA SOSTENIBILIDAD

I. El estudio de la evaluabilidad

Una intervención de prevención de consumo de drogas en la escuela en Quebec

BIESSÉ DIAKARIDJA SOURA, JEAN-SÉBASTIEN FALLU, ROBERT BASTIEN Y FRÉDÉRIC N. BRIÈRE

Evaluación de la evaluabilidad

Definición del método

La Evaluación de la Evaluabilidad es un proceso riguroso de recogida y de análisis de datos que permite garantizar que la evaluación del programa se puede llevar a cabo y producir resultados satisfactorios y útiles para todas las partes implicadas.

Fortalezas del método

- Proceso participativo que permite la contribución de las partes implicadas a la mejora del programa y de la evaluación.
- Permite entender mejor el marco teórico del programa y desarrollar su marco lógico.
- Permite aprovechar tiempo y recursos, al recomendar formas de llevar a cabo una evaluación.
- Permite una mejor planificación de la propia evaluación.

Desafíos del método

- Puede llevar mucho tiempo si las partes implicadas no se entienden bien, o si las actividades no están bien organizadas, o si hace falta consultar a un gran número de personas.
- Puede incrementar los malos entendidos y provocar frustración en algunas partes implicadas.

Para Craig y Campbell (2015), la Evaluación de la Evaluabilidad (EE) es un enfoque sistemático destinado a planificar la evaluación de intervenciones o programas. Una de las contribuciones más significativas de la evaluación de la evaluabilidad se basa en el hecho de que puede aportar recomendaciones para mejorar el programa, sobre la mejor manera de desarrollarlo y sobre la evaluación posterior que debería hacerse. Si bien la evaluación de la evaluabilidad lleva desarrollándose desde la década de 1970, para responder a las dificultades y desafíos en la evaluación de programas gubernamentales en los Estados Unidos (Wholey 1976), solo recientemente su uso se ha incrementado entre profesionales de la evaluación como en el ámbito

académico (Trevisan y Yi 2003). Según Trevisan y Walser (2014), este nuevo interés ha coincidido con la creciente demanda de rendición de cuentas a los gestores de programas.

La evaluación de la evaluabilidad es una evaluación rápida de tipo exploratorio que puede llevarse a cabo en un período más o menos corto, entre algunas semanas a varios meses (Davies 2013; Wholey 2010). Incluye varias etapas que permiten recopilar información sobre las intenciones de quienes impulsan el programa, los objetivos que se quieren conseguir, tipo de datos que se recopilarán y forma de utilizar los resultados de la evaluación. La realización de una evaluación de la evaluabilidad también puede permitir desarrollar el modelo lógico del programa, una herramienta necesaria para realizar la evaluación en sí.

En resumen, la evaluación de la evaluabilidad se presenta como una práctica útil y pertinente para la evaluación de programas y para la toma de decisiones (Soura, Dagenais, Bastien, Fallu y Janosz 2016). Sin embargo, sigue siendo poco utilizado, especialmente en el entorno de habla francesa. De hecho, no hay una única terminología en francés para hacer referencia a este enfoque.

El objetivo de este capítulo es, primero, presentar brevemente la evaluación de la evaluabilidad y, luego, describir un ejemplo de su uso en el contexto de la evaluación de una intervención para prevenir el uso de sustancias psicoactivas en las escuelas, antes de presentar un análisis reflexivo de nuestra experiencia.

Presentación de la evaluación de la evaluabilidad

La evaluación de la evaluabilidad es un proceso riguroso de recopilación y análisis de datos que permite garantizar que la evaluación del programa sea factible y que pueda producir resultados satisfactorios y útiles para las partes implicadas (Kaufman-Levy y Poulin 2003). Por “parte implicada” nos referimos a todas las personas o entidades con un interés particular en la implementación de un programa y su evaluación. La evaluación de la evaluabilidad también puede verse como una evaluación exploratoria o una actividad de evaluación previa que debería permitir el establecimiento de las condiciones óptimas para una evaluación formativa o sumativa formal (Leviton, Khan, Rog, Dawkins y Cotton 2010a). Los primeros borradores de la

evaluación de la evaluabilidad se remontan a la década de 1970, cuando un grupo de investigadores del *Urban Institute* en Washington, Estados Unidos, implementó este enfoque en respuesta a las dificultades y desafíos encontrados en la evaluación de programas gubernamentales. Según estos investigadores, dichos programas se implementaron de manera insuficiente, eran demasiado incipientes para ser evaluados o tenían objetivos considerados poco realistas (Burrows, Bilodeau y Litvak 2012; Leviton, Khan, Rog, Dawkins y Cotton 2010 b; Nay y Kay 1982; Wholey 1976). Los críticos también se centraron en la discrepancia observada entre el programa evaluado y los resultados de esta evaluación (Leviton et al. 2010b). Posteriormente, se inició una reflexión sobre la capacidad de las evaluaciones, según fueron implementadas, para informar a quienes toman las decisiones sobre la formulación de políticas públicas y la necesidad de gastar muchos fondos en intervenciones con escaso impacto en la población diana (Trevisan y Yi 2003; Van Voorhis y Brown 1987). Es en este contexto en el que Horst y sus colegas se propusieron identificar las principales causas del fracaso de los programas y la incapacidad de la evaluación para satisfacer las necesidades de las partes implicadas (Horst, Nay, Scanlon y Wholey 1974). Este último recomendó que se llevara a cabo una evaluación previa de evaluabilidad, que aumentaría la capacidad de utilizar los resultados de la evaluación. Es esta actividad previa a la evaluación la que más tarde se llamó “evaluación de evaluabilidad” o “evaluación de la evaluabilidad”.

La realización de una evaluación de la evaluabilidad consiste en seguir los pasos descritos en los trabajos realizados tanto desde el ámbito académico como desde los organismos para el desarrollo (Trevisan y Walser 2014). A continuación se presenta la Tabla 1 con las cuatro etapas de una evaluación de la evaluabilidad según Dunn (2008), que se convierten en 10 según Smith (1989). La diferencia no está solo en el número de pasos a seguir, sino también en las primeras acciones a tomar. Así, Wholey (1987) propone comenzar involucrando a los posibles interesados en el proceso, mientras que para Thurston y Potvin (2003), primero es necesario seleccionar a la persona responsable de llevar a cabo el evaluación de la evaluabilidad. Independientemente del orden de las diferentes etapas, Smith (1989; 1990) sugiere que las acciones a tomar deben guiarse por el contexto y los objetivos a conseguir.

Tabla 1 : Las etapas de la evaluación de la evaluabilidad de los programas según algunos autores (Adaptado de Trevisan y Walser (2014) y Soura, Bastien y Fallu (2016). * etapas añadidas en Wholey (2010))

Etapas	Strosberg y Wholey (1983)	Wholey (1987, 2010)	Rog (1985)	Smith (1989)
1	Definir los recursos, actividades, objetivos y resultados esperados del programa.	Hacer participar a los eventuales beneficiarios y partes implicadas	Analizar el programa	Definir la finalidad, asegurar adhesiones y compromisos, poner en marcha el grupo de trabajo
2	Acordar la descripción del programa con las partes implicadas	Clarificar la formulación del programa	Analizar la puesta en marcha del programa	Definir los límites del programa a evaluar
3	Definir el tipo de datos que hay que recoger	Explorar la realidad de la puesta en marcha del programa	Analizar el sistema de recogida de la información	Identificar y analizar la documentación del programa
4	Verificar la puesta en marcha real del programa	Analizar si el programa es plausible*	Analizar si los objetivos del programa son plausibles	Desarrollar el marco teórico del programa
5	Comprobar si los recursos y actividades previstas permiten alcanzar los objetivos	Tener en cuenta posibles variaciones en el programa o en su implementación *	Preparar variaciones del programa	Identificar las partes implicadas y mantener entrevistas con ellas
6	Definir si el programa dispone de un sistema de gestión de la información	Tener en cuenta las posibles formas de evaluar y el uso que se dará a los resultados	Describir el uso de los resultados de la evaluación	Describir la percepción de las partes implicadas en el programa
7	Definir la parte del programa que será evaluada en función de los objetivos identificados			Identificar las necesidades, intereses y deferencia en las percepciones de las partes implicadas
8	Identificar las opciones de evaluación y de gestión por parte de los responsables del programa			Determinar si el modelo de programa es factible
9				Sacar conclusiones y hacer recomendaciones

Etapas	Strosberg y Wholey (1983)	Wholey (1987, 2010)	Rog (1985)	Smith (1989)
10				Planificar el uso de los resultados de la evaluación de la evaluabilidad

Tabla 1

Etapas	Thurston y Potvin (2003)	Dunn (2008)	Dunet et al. (2013)
1	Selección del evaluador responsable de realizar el estudio de evaluabilidad	Comprueba el modelo causal	Selecione el área de práctica
2	Identificar a los interesados en el programa	Acordar el propósito de la evaluación del impacto	Identificar los programas
3	Identificar y analizar la documentación	Evaluar la viabilidad de un modelo alternativo	Recopilación de información y evaluación del valor del estudio de evaluabilidad
4	Desarrollando el modelo lógico del programa y el plan de evaluación	Identificar un equipo de evaluación local	Selecione el/los programa(s) a ser evaluado(s)
5	Acordar el principio de la realización de una evaluación		Preparación de los equipos para las visitas sobre el terreno
6	Evaluar el tiempo y los recursos necesarios para la evaluación		Realizar visitas sobre el terreno
7			Preparar documentos de estudio de evaluabilidad
8			Determinar la posibilidad de una evaluación rigurosa

El primer modelo para realizar una evaluación de la evaluabilidad proviene de Wholey (1987), uno de los pioneros de este enfoque (Smith, 2005). Según él, una evaluación de la evaluabilidad debería permitir dar respuestas a cuatro cuestiones: 1) el marco teórico y el marco lógico son claros, es decir que los objetivos del programa y la información relevante están bien identificados y definidos; 2) los objetivos del programa son plausibles; 3) se pueden obtener datos relevantes del desarrollo del programa a un coste razonables; y 4) se aclara cómo los beneficiarios potenciales

utilizan los resultados de la evaluación. Al cumplir estas cuatro condiciones, la evaluación de la evaluabilidad debería permitir, como señala Smith (1989), tomar una de las siguientes decisiones: 1) dejar de desarrollar el programa, 2) realizar cambios en él, 3) realizar una evaluación formal, 4) dejar las cosas en su estado original, o 5) ignorar los resultados de esta evaluación.

El evaluación de la evaluabilidad puede llevarse a cabo en un corto período de tiempo, entre unas semanas y uno o dos meses, antes de realizar una evaluación formal adecuada que pueda centrarse en el proceso y/o el impacto del programa. También se puede extender a lo largo del tiempo, dependiendo de las actuaciones a realizar y la disponibilidad de las partes implicadas. En la mayoría de evaluaciones de evaluabilidad se utilizan sobretodo datos cualitativos recopilados a través de entrevistas, revisión bibliográfica, grupos focales, visitas de campo y observación (Esher et al.2011). El trabajo en evaluación de la evaluabilidad con datos cuantitativos aún es escaso (Trevisan 2007). Sin embargo, debe recordarse que al realizar la evaluación de evaluabilidad, el método seleccionado se guiará en gran medida por imperativos logísticos, temporales y económicos.

Si, a primera vista, puede parecer innecesario realizar una evaluación de la evaluabilidad antes de una evaluación formal, esta actividad previa a la evaluación se recomienda cuando el programa a evaluar se califica como complejo (Contandriopoulos, Rey, Brousselle y Champagne 2012; Kania et al.2013; Ogilvie et al.2011; Rogers 2008; Touati y Suárez-Herrera 2012), cuando el marco teórico y el lógico no son claros, o cuando el programa está mal documentado (Dunet et al.2013; Rutman 1997). La evaluación de la evaluabilidad puede ayudar a desentrañar la maraña de esta complejidad, al promover la descripción detallada del programa y el uso de las herramientas necesarias para una evaluación posterior.

El evaluación de la evaluabilidad puede ser útil tanto para el programa, para quienes lo evalúan y quienes toman las decisiones (Soura, Dagenais et al., 2016). Hoy en día hay consenso en que realizar una evaluación de la evaluabilidad puede ser beneficioso durante la fase de desarrollo del programa, durante su implementación y durante la fase posterior a la implantación. Con respecto a los evaluadores, pueden beneficiarse de la evaluación de la evaluabilidad en el sentido de que es una forma rápida de comprender el estado de un programa, e identificar las prioridades de las partes implicadas antes de dedicar muchos recursos a su evaluación. Además, la evaluación de la evaluabilidad puede contribuir a mejorar la

planificación de las actividades de evaluación. En cuanto a quienes toman decisiones, tienen la oportunidad de participar en el proceso de evaluación expresando su punto de vista y sus intereses. Además, la evaluación de la evaluabilidad permite a los gestores hacer una mejor distribución de los recursos económicos, materiales y humanos, al enfocar la evaluación solo en los aspectos del programa que puedan responder a los intereses de las partes implicadas.

Ejemplo de evaluación de la evaluabilidad: La Intervención en Red

El programa evaluado se llama Intervención de Red (IR). Es un programa de intervención destinado a evitar el uso de sustancias psicoactivas (SPA *en sus siglas en francés*) en las escuelas. Fue desarrollado por la *Commission scolaire des Navigateurs de Lévis* en Quebec, tras una amplia consulta con todos los actores en el entorno escolar. Esta consulta destacó la importancia de implementar prácticas más innovadoras y adaptadas para evitar el uso de sustancias psicoactivas por parte del alumnado. La intervención desarrollada ponía énfasis en tres elementos fundamentales: la creación de un vínculo fuerte entre el alumnado y la persona responsable del desarrollo de las actividades, iniciar la intervención en la escuela primaria, dando continuidad hasta la secundaria (3er año) y desarrollar actividades inspiradas en el Enfoque de escuelas saludables (AES *en sus siglas en francés*) (Désy 2009).

Una primera versión del proyecto se lanzó al comienzo del año escolar 2007-2008 con financiación del Fondo, en colaboración con el Centro de Salud y Servicios Sociales (CSSS) de Grand-Littoral. Luego se realizó un estudio de viabilidad para validar y medir el potencial impacto de esta intervención. Se hicieron ajustes para conseguir el formato actual de esta intervención. Para la fase piloto, la intervención se llevó a cabo en cinco escuelas primarias y una escuela secundaria de la zona.

La intervención de red (IR) funciona según el modelo de gestora de casos (*intervenante/e pivot*, IP *en sus siglas en francés*). Este modelo permite a una persona especialista coordinar varias acciones de diferentes entidades para encontrar una solución a una situación problemática (Farber, Deschamps y Cameron 2002; Canadian Breast Cancer Initiative 2002; Paradis, Racine y Ganó 2006). La intervención consiste en ubicar un/a IP en un centro de secundaria adscrito a un grupo de escuelas primarias ubicadas en el mismo

área geográfica. Esta agrupación de escuelas secundarias y primarias se denomina red. El IP puede ser personal de la escuela en quien se delegan las tareas vinculadas a la intervención, o alguien a quien se contrata específicamente para desempeñar este papel. Se ubica normalmente en el centro de secundaria. Visitarán de forma regular las escuelas primarias que se les asignaron. Por lo tanto, el/la IP puede hacer un seguimiento del alumnado durante un período más o menos largo, en función de su continuidad en el centro de secundaria. Este procedimiento le permite desarrollar actividades en ambas etapas educativas (primaria y secundaria) y permite un mejor conocimiento del alumnado. Quienes están en riesgo, se identifican más fácilmente para un adecuado seguimiento. Más allá de esta presencia física, el/la IP también debe desarrollar actividades preventivas del consumo de SPA, con la colaboración de docentes y otros actores dentro de la escuela y la comunidad.

Para el alumnado del último ciclo de primaria, las actividades se centran en los talleres del *Système C et D*, dos programas que ya se utilizan en las escuelas de Quebec. En particular, se dirigen a los factores de riesgo y protección comunes en la prevención del uso de SPA entre la juventud, al promover la adquisición de hábitos de vida saludables, conocimiento y asertividad, manejo del estrés, reacción frente a la presión del grupo de iguales y los medios de comunicación, y la resolución de conflictos (Departamentos de salud pública de las agencias de servicios sociales y de salud en Montérégie de Laval y Outaouais, 2007). También hay actividades sobre la transición entre primaria y secundaria, por ejemplo, visitas y reuniones de alumnado de primaria con alumnado de secundaria y docentes, conocimiento del funcionamiento del centro de secundaria, etc. En el centro de secundaria (del primer al tercer año), el alumnado participa en actividades centradas en los talleres del programa APTE (Garand-Butcher 2014; Vaugeois, Germain y Cunha Rêgo 2008). Se trata de actividades relacionadas con las creencias sobre el consumo de SPA, el conocimiento de los niveles de riesgo, la presión del grupo y las pandillas callejeras. Se considera que estas actividades refuerzan los hábitos adquiridos en la escuela primaria y están dirigidas por el/la IP y otros actores implicados (docentes, educadores sociales, enfermeras, servicios comunitarios, policía, etc.). Las actividades ofrecidas al alumnado de primaria y secundaria se presentan en la Tabla 2.

Tabla 2 : Lista de actividades de la intervención en red

Primaria (5º y 6º)	Secundaria (1º, 2º y 3º)
1. Hábitos de vida saludables	Consumo de SPA y sus consecuencias
2. Autoconocimiento y Asertividad	Hábitos de vida saludables
3. Relaciones sociales y comunicación	Resistencia a la presión de grupo negativa
4. Presión social, consumo y medios de comunicación	Relación docente-esudiante
5. Ciberdelito	Violencia y ciberdelito
6. Resolución de conflictos y gestión del estrés	Comprender la secundaria
7. Preparación para la secundaria	Entrevistas individuales
8. Actividades educativas y ocio	Actividades educativas y ocio

Se espera conseguir que, a corto plazo, la intervención mejore las habilidades sociales del alumnado, su manejo del estrés, su percepción de los riesgos del consumo de sustancias psicoactivas, promueva buenas relaciones entre docentes y alumnado, la habilidad para la resolución de conflictos y reduzca el nivel de exposición del alumnado al SPA. A medio plazo, se trata de retrasar la edad de inicio del consumo de SPA y prevenir el consumo excesivo y problemático de estas sustancias. El objetivo final de esta intervención es promover el éxito educativo del alumnado.

Datos para la evaluación de la evaluabilidad

Para llevar a cabo esta evaluación de la evaluabilidad, los datos cualitativos se obtuvieron principalmente de informantes clave seleccionados por su participación en el diseño y/o desarrollo de la intervención. La entrevista semiestructurada y el análisis documental fueron las principales estrategias de recogida de datos. El objetivo general era conocer el objeto de evaluación al inicio, antes de comenzar su evaluación en una fase posterior. En concreto, se trataba de comprender el marco teórico de la intervención, dado que la escasa documentación disponible en ese momento era ambigua al respecto. A partir de ahí, podríamos desarrollar

el modelo lógico y determinar el enfoque de evaluación que incluyera el funcionamiento de la intervención y su posible impacto en la población diana.

Sobre la base de las cuatro condiciones descritas por Wholey (1987) y presentadas anteriormente, esta evaluación de la evaluabilidad tuvo como objetivo responder las siguientes preguntas:

- ¿Se presentan claramente los fines de sus promotores y los objetivos de la intervención?
- ¿Los objetivos de la intervención son plausibles a la luz del conocimiento existente y los recursos disponibles?
- ¿Se cumplen las condiciones para la recogida de datos para la evaluación?
- ¿Qué utilidad tendrán los resultados de la evaluación para las partes implicadas?

Para responder a estas preguntas, se llevó a cabo la planificación de actividades. Se tuvieron en cuenta los objetivos, el tipo de datos recogidos, las fuentes de datos, los participantes y el período de recogida. La recogida de datos se llevó a cabo en un mes, un año después del inicio de la puesta en marcha del programa, con los participantes como informantes clave. Esta planificación se presenta en la Tabla 3

Tabla 3. Planificación de las actividades de evaluación

Criterios de evaluación	Sujetos/items tenidos en cuenta	Fuentes de datos
Fines y marco lógico de intervención	<ul style="list-style-type: none"> • Fines y objetivos de la intervención • Actividades y recursos previstos y disponibles • Población diana de la intervención • Resultados esperados 	<ul style="list-style-type: none"> • Entrevistas • Revisión bibliográfica
Factibilidad de los objetivos	<ul style="list-style-type: none"> • Condiciones de puesta en marcha de la intervención • Actividades previas realizadas • Adecuación de los recursos a los resultados esperados 	<ul style="list-style-type: none"> • Entrevistas • Revisión bibliográfica • Cuaderno diario
Disponibilidad de datos	<ul style="list-style-type: none"> • Identificación de un enfoque de evaluación de impacto y de proceso • Tipo de datos disponibles o para recoger • Actividades de seguimiento puestas en marcha • Disponibilidad de información sobre las actividades 	<ul style="list-style-type: none"> • Entrevistas • Diario de campo • Informes de reuniones
Utilidad de la intervención y de los resultados de la intervención	<ul style="list-style-type: none"> • Utilidad de la intervención para sus promotores • Uso de los resultados de la evaluación 	<ul style="list-style-type: none"> • Entrevistas • Informes de reuniones

Los datos recopilados se analizaron temáticamente en cuatro etapas. La primera etapa deductiva permitió indagar sobre cuestiones relacionadas con las condiciones del enfoque de la evaluación de la evaluabilidad según Wholey (1987). Un segundo paso, esta vez inductivo, estaba abierto a la aparición de nuevos temas (Saberí, Yuan, John, Sheon y Johnson 2013). Esta etapa permitió identificar temas relacionados con la satisfacción de los participantes con el desarrollo del programa y la buena colaboración con el IP. La tercera etapa consistió en agrupar todos los temas identificados en una matriz que permitía identificar subcategorías y citas textuales relevantes para el marco de análisis. Se realizó un reajuste para formar grupos grandes y más homogéneos. Finalmente, la cuarta etapa permitió extraer citas para ilustrar los resultados obtenidos. El procesamiento de datos se realizó

utilizando el software de procesamiento de datos cualitativo QDA Miner. Antes de presentar los resultados obtenidos, se describen a continuación las diferentes etapas de la implementación de la evaluación de la evaluabilidad.

Puesta en marcha de la evaluación de la evaluabilidad

La realización de una evaluación de la evaluabilidad no es un fin en sí mismo. Esta actividad debe satisfacer una necesidad específica. En general, sirve para generar una evaluación susceptible de producir resultados útiles para las partes implicadas. En este sentido, puede permitir, como ya se ha mencionado, una buena comprensión del marco teórico de programa y el desarrollo del modelo lógico. Antes de realizar una evaluación de la evaluabilidad, sus méritos deben explicarse a quienes promueven el programa y a todas las entidades que podrían beneficiarse de los resultados obtenidos. Los límites de dicha actividad también deben presentarse, para no generar expectativas innecesarias. Una reunión informativa con los diferentes actores puede ser una excelente oportunidad para lograr este objetivo.

Además, dado que la evaluación de la evaluabilidad se puede comparar con un análisis del proceso de un programa, para saber si cumple con ciertos criterios antes de evaluarlo formalmente, solo se debe contar con quienes disponen de información del programa. Por lo tanto, debe garantizarse su disponibilidad, y que la documentación, si existe, esté a disposición del equipo de evaluación. Finalmente, el conocimiento exhaustivo del contexto organizativo y sociopolítico en el que se desarrolla el programa representa un activo importante para comprender la dinámica de la toma de decisiones; ello puede contribuir a adecuar las recomendaciones al contexto y que sean adoptadas por las partes implicadas.

Procedimiento y descripción de las etapas

La realización de una evaluación de la evaluabilidad implica varias etapas que se han descrito en la literatura. Varios autores presentan los diferentes pasos a seguir para realizar una evaluación de la evaluabilidad, como se ha mostrado en la Tabla 1 anteriormente. Por motivos pedagógicos y para cumplir con los objetivos de este libro, la siguiente sección describe los

10 pasos propuestos por Smith (1989) en el enfoque más exhaustivo. Estos mismos pasos se han descrito de manera simplificada en otros lugares (Trevisan y Yi 2003). Sin embargo, debe tenerse en cuenta que este no es un proceso lineal, por lo que no está de más volver a reformular los objetivos de la evaluación de la evaluabilidad, por ejemplo, mientras se realizan las entrevistas con las partes implicadas.

Definir la finalidad, asegurar adhesiones y compromisos, poner en marcha el grupo de trabajo

Es muy importante identificar la finalidad antes de iniciar una evaluación de la evaluabilidad. Esto permite aclarar las expectativas desde el principio y alentar la participación de las partes implicadas. También en esta etapa, la creación de un grupo de trabajo puede fomentar la participación de los implicados y, por lo tanto, facilitar el desarrollo de las actividades. El grupo de trabajo establecido debe ser representativo de las diferentes entidades y pequeño, para permitir que funcione correctamente. En el caso presentado aquí, el grupo motor consistió principalmente en el diseñador de la intervención, el profesional clave y el evaluador principal. Se contactó a las otras entidades involucradas cuando surgió la necesidad, en particular para la validación del modelo lógico.

Definir los límites del programa a evaluar

La evaluación de la evaluabilidad se plantea como análisis de situación rápido de un programa antes de llevar a cabo una evaluación real. Como tal, no puede proporcionar respuestas para todo. Es importante establecer límites sobre lo que se puede y no se puede lograr en el tiempo disponible. Establecer límites puede ayudar a comprender mejor los objetivos de la evaluación de la evaluabilidad y evitar expectativas que no se puedan cumplir. En el caso presentado aquí, se trataba de comprender el programa en su totalidad, desarrollar el modelo lógico y proponer vías de evaluación del programa.

Identificar y analizar documentos relacionados con el programa

La consulta de la documentación sobre el programa es un paso importante en la realización de una evaluación de la evaluabilidad. Al inicio del proceso, es necesario identificar documentación que pueda proporcionar información útil sobre el programa y garantizar que está disponible. Se pueden incluir informes de reuniones, investigaciones o encuestas, solicitudes de financiación, normativa e incluso mensajes

intercambiados entre las partes implicadas. Estos documentos permiten comprender la finalidad de quienes ponen en marcha el programa y la teoría del cambio promovida. No es raro, como en este caso, que se disponga de poca documentación sobre el programa. En ese caso, será necesario consultar a las personas responsables del desarrollo del programa para recoger información.

Desarrollar / aclarar el marco teórico del programa

Uno de los beneficios de realizar una evaluación de la evaluabilidad es la capacidad de clarificar el marco teórico del programa y desarrollar el modelo lógico. El marco teórico del programa define la forma en que los promotores quieren lograr sus objetivos. Explica cómo van a conseguirse los objetivos con los recursos disponibles, las estrategias adoptadas y las actividades realizadas. En muchas situaciones, este marco teórico no es explícito o se expresa de manera vaga o fragmentada. Sólo desentrañando y aclarándolo se puede comprender el programa en su totalidad. Una vez aclarado el marco teórico, es posible desarrollar el modelo lógico, que es una representación esquemática de la interacción causa-efecto prevista entre los recursos utilizados, las actividades realizadas y los resultados esperados a mediano, corto o largo plazo. También deben tenerse en cuenta los factores favorables o desfavorables que podrían influir en el desarrollo del programa, tanto en una dirección como en la otra. El desarrollo del modelo lógico debe realizarse consultando a las partes implicadas. En cualquier caso, la versión final rara vez se parece al modelo inicial, debido a las modificaciones sucesivas que se pueden hacer a medida que se clarifica el marco teórico.

Identificar y entrevistar a las partes implicadas

Identificar y entrevistar a las partes implicadas es otro paso en la realización de una evaluación de la evaluabilidad, ya que estas personas pueden contribuir a una mejor comprensión del programa. Además, es gracias a su apoyo que el programa puede desarrollarse. Las entrevistas deben centrarse en lo que saben y perciben sobre el programa y su puesta en marcha. En el caso de IR, los entrevistados fueron identificados con la colaboración del IP que estaba en contacto directo con ellos. Una vez que estas personas han sido identificadas, es importante contactarlas con varios días de antelación para planificar la organización de las entrevistas.

Describir la percepción de las partes implicadas en el programa

Las entrevistas con las partes implicadas deben describir la percepción de estos informantes clave sobre el programa. Más allá de esta descripción, será necesario comparar los diferentes puntos de vista expresados para obtener una mejor comprensión de los desafíos del programa a desarrollar o ya desarrollado.

Identificar las necesidades, intereses y diferencias en las percepciones de las partes implicadas

En esta etapa, corresponde al asesor y a su equipo identificar las diferencias y convergencias en las percepciones, necesidades e intereses de los diferentes actores. Se debe prestar especial atención a las discrepancias debido a sus posibles efectos negativos en el programa. Comprender las diferencias e intentar resolverlas, ayuda a evitar malentendidos y controversias que pueden surgir cuando se introducen innovaciones en las organizaciones.

Determinar si el modelo de programa es factible

Cuando se realiza correctamente, la revisión de la literatura y las entrevistas deberían ayudar a determinar si el programa y sus objetivos son factibles. Esta factibilidad está basada en los recursos disponibles, pero también en la revisión de la literatura. Se trata de indicar el grado en que el programa se desarrolla de manera correcta y suficiente, y si es probable que las actividades realizadas produzcan los resultados esperados.

Sacar conclusiones y hacer recomendaciones

La realización de una evaluación de la evaluabilidad debe conducir a la formulación de recomendaciones que puedan orientar las medidas que haya que adoptar. Estas conclusiones y recomendaciones deben ser específicas para el programa que se evalúa y extraerse solo de los datos disponibles.

Planificar el uso de los resultados de la evaluación de la evaluabilidad

El análisis de los datos recogidos debería permitir hacerse una idea de cómo se va a desarrollar el programa. Las recomendaciones hechas deben ir en la dirección de proponer modificaciones al programa y/o seguir adelante con una evaluación formal o no formal. También es posible que las conclusiones de una evaluación de la evaluabilidad recomienden interrumpir el programa, porque su marco teórico no está adaptado al contexto o a la

población diana, pudiendo producir efectos lesivos. El programa también puede interrumpirse porque la evaluación de la evaluabilidad ha demostrado que no se está desarrollando según lo previsto.

Resultados obtenidos

Los resultados obtenidos en el marco de la evaluación de la evaluabilidad de la Red de Intervención se pueden agrupar en tres aspectos. Comencemos con el nivel de clarificación del marco teórico del programa. Al comienzo de las actividades de evaluación, se disponía de escasa documentación sobre el enfoque adoptado por sus promotores. Aunque los objetivos a alcanzar se hubieran presentado de forma clara, y hubo implicación de los actores para desarrollar el programa, pero la estrategia para lograr los objetivos no estaba tan clara. Además, algunos actores sobre el terreno no sabían a ciencia cierta lo que debía hacerse. La evaluación de la evaluabilidad también mostró que el desarrollo del programa no fue igual en todas las escuelas. Después de estas observaciones, se organizaron reuniones con las partes implicadas, durante las cuales se realizó una presentación más profunda del programa. A continuación, se le presentaron al IP los resultados de las consultas con otros actores, y se le trasladó la necesidad de documentar las actividades organizadas. Después de esto, se tomaron medidas para armonizar el grado de desarrollo en las escuelas participantes. Al mismo tiempo, se mejoró significativamente el seguimiento de las actividades.

Continuemos con el desarrollo del modelo lógico. Una vez que se aclaró el marco teórico del programa, fue más fácil desarrollar el modelo lógico, que no existía al comienzo de las actividades de evaluación. Se desarrolló después de un análisis de la documentación disponible y los debates con los promotores del programa. La versión inicial del modelo lógico sufrió varias modificaciones, antes de que se adoptara la versión final para respaldar la evaluación que debería llevarse a cabo (Soura, Bastien, et al.2016)

Finalmente, con respecto a la planificación de la evaluación, dado que el equipo responsable de la evaluación de la evaluabilidad era el mismo que debía llevar a cabo las actividades de evaluación propiamente dichas, se hicieron propuestas sobre los métodos para llevar a cabo esta evaluación. El debate con las partes implicadas y el análisis de la documentación, facilitaron la identificación del enfoque de evaluación a adoptar y el marco de análisis

más apropiado para informar sobre el proceso de desarrollo del programa. De hecho, parecía que el desarrollo del programa en las escuelas había causado controversia entre ciertos actores. Al mismo tiempo, se trataba de un programa novedoso para todas las escuelas involucradas. Todo esto llevó al equipo de evaluación a prestar más atención a la forma en que los actores intervenían para desarrollar el programa, que al análisis de si se estaba desarrollando fielmente. Por lo tanto, se realizó una revisión de la literatura para identificar un marco analítico que tuviera en cuenta esta situación. Después de esto, se eligió la teoría del *actor-red* para servir como punto de referencia para la evaluación del proceso. Desarrollado inicialmente para el estudio de la introducción de innovaciones tecnológicas en las organizaciones, este marco analítico ha demostrado ser interesante, ya que también presta atención a las controversias que surgen durante el desarrollo de un programa.

Referente a la planificación de la evaluación, la realización de la evaluación de la evaluabilidad también permitió obtener una estimación cuasi-experimental de los efectos del programa en los participantes. Se prefirió esta estimación, obtenida antes de realizar la evaluación de la evaluabilidad, a un estudio aleatorizado, debido a los recursos disponibles y las limitaciones vinculadas al desarrollo del programa. Permite comparar el grupo de estudiantes que participaron en el programa (grupo de intervención) con otro grupo de estudiantes que no estuvieron expuestos a él (grupo control). La viabilidad de este componente cuantitativo de la evaluación también se abordó durante la evaluación de la evaluabilidad. Sin duda, llevar a cabo esta evaluación de la evaluabilidad contribuyó en gran medida a comprender mejor el programa y a desarrollar las herramientas necesarias para la evaluación que siguió. Esta experiencia nos permitió realizar un análisis reflexivo sobre la realización de una evaluación de la evaluabilidad.

Análisis reflexivo

Podemos extraer tres lecciones principales de la realización de esta evaluación de la evaluabilidad. La primera es que la evaluación de la evaluabilidad es un proceso participativo, que requiere una interacción constante entre el equipo de evaluación y las partes implicadas. Incluso si no

es posible contar con su plena cooperación, es esencial y muy importante, al menos, que se establezca un proceso de validación de resultados, con el fin de recoger sus opiniones. Esto tiene la ventaja de maximizar las posibilidades de que se sigan las recomendaciones, especialmente cuando recomiendan cambios importantes en el desarrollo del programa, o una reorientación de sus objetivos. Además, la participación de las partes implicadas en la realización de la evaluación de la evaluabilidad, permite documentar mejor el programa y hacer recomendaciones contextualizadas.

En segundo lugar, realizar una evaluación de la evaluabilidad es beneficioso no solo para el equipo de evaluación, sino también para los propios profesionales, que tienen la oportunidad de reflexionar, contribuyendo a mejorar sus prácticas. Las etapas para llevar a cabo una evaluación de la evaluabilidad y las listas de verificación desarrolladas por los diferentes autores, son una buena forma de empezar, si no sabe por dónde empezar. Sin embargo, estas herramientas deben usarse de manera flexible y, sobre todo, adaptarse a la realidad y el contexto del programa que se va a evaluar, así como a la audiencia.

Tercero, incluso si los recursos son limitados, realizar una evaluación de la evaluabilidad, aunque solo sea a pequeña escala, puede ser muy beneficioso, tanto para el programa y sus promotores, como para el equipo de evaluación. Esta evaluación de la evaluabilidad fue una excelente manera de conocerse mejor todos los actores, antes de iniciarse la evaluación. Además, al alentar la participación de las partes implicadas, el evaluación de la evaluabilidad promueve su apropiación del proceso y, por lo tanto, representa un excelente ejercicio de empoderamiento de quienes no siempre tienen habilidades para evaluar el programa. Si el tiempo lo hubiera permitido, y las condiciones de trabajo hubieran sido diferentes, un estudio más profundo hubiera permitido planificar mejor la evaluación para hacerla en menos tiempo.

Como se ha señalado ya, la evaluación de la evaluabilidad tiene fortalezas, pero también algunas debilidades. Como puntos fuertes, hay que tener en cuenta que la evaluación de la evaluabilidad facilita un proceso participativo, durante el cual interactúan todas las entidades involucradas en el programa y la evaluación. Al tener en cuenta los puntos de vista de las diferentes entidades, es posible mejorar el programa y proponer resultados de evaluación adaptados a las preocupaciones de sus beneficiarios. La evaluación de la evaluabilidad también ayuda a aclarar el marco teórico del

programa y a desarrollar el modelo lógico. De hecho, gracias a la evaluación de la evaluabilidad, podemos lograr una mejor comprensión del marco teórico de un programa, y la finalidad que buscan sus promotores, lo que puede favorecer el desarrollo del modelo lógico que sigue siendo una herramienta esencial para la evaluación del programa. Otra ventaja de realizar una evaluación de la evaluabilidad es que puede contribuir a realizar mejoras sustanciales en el programa, ya sea en la fase de diseño o durante su desarrollo. Es un ejercicio que ahorra tiempo y recursos, ya que puede recomendar no realizar una evaluación, si no es útil para los implicados. Finalmente, la evaluación de la evaluabilidad puede facilitar la planificación de la evaluación futura, proponiendo estrategias e identificando enfoques y herramientas relevantes, para el contexto en el que se desarrolla el programa y para generar conocimiento útil para todas las partes implicadas.

La principal debilidad de la evaluación de la evaluabilidad se refiere al hecho de que, dado que requiere un proceso participativo, también puede llevar mucho tiempo detectar que el grupo de trabajo creado no funciona de manera óptima, o si no hay acuerdo entre sus miembros, o si el número de personas a consultar es muy alto. El otro punto débil es que la evaluación de la evaluabilidad también puede exacerbar los desacuerdos y causar frustraciones en ciertos actores, especialmente cuando las expectativas no se cumplen. Por lo tanto, es muy importante consensuar los objetivos a alcanzar desde el inicio de este proceso. Además, es obvio que la realización de una evaluación de la evaluabilidad presenta desafíos dependiendo del contexto en el que se lleva a cabo.

El primer desafío al realizar esta evaluación de la evaluabilidad fue el de poder movilizar a todas las partes implicadas, para que participaran en el proceso, dadas las limitaciones de tiempo vinculadas a las actividades lectivas. Como el programa se lleva a cabo en las escuelas, no siempre fue fácil para docentes y otros profesionales liberarse para participar en las entrevistas. Otro desafío en este proceso fue encontrar los recursos humanos y económicos necesarios para realizar el estudio de evaluabilidad, cuando no se incluyó en los objetivos iniciales de la evaluación. Es probable que esta situación, combinada con la larga distancia entre el lugar donde se ubicaba el programa y el lugar de trabajo de los evaluadores, haya alargado demasiado la realización del estudio de evaluabilidad. Cuando el estudio de evaluabilidad no es una de las preocupaciones de los actores implicados, puede ser difícil pedirles que esperen a los resultados estén disponibles

antes de comenzar las actividades de evaluación propiamente dichas. Finalmente, sin haber sido necesariamente un desafío en la realización de esta evaluación de la evaluabilidad, pero que merece ser mencionado, el juego de poder de cómo conciliar los puntos de vista de los diferentes actores, especialmente cuando son divergentes. Esta es la razón por la cual, para terminar de manera exitosa una evaluación de la evaluabilidad, es fundamental que haya un buen facilitador de grupo que sepa negociar.

Conclusión

En resumen, hay que tener en cuenta que, en los últimos años, la evaluación de la evaluabilidad está siendo cada vez más utilizada. De hecho, desde que se inició, numerosos trabajos han contribuido a facilitar su aplicación, al proponer los pasos a seguir y las listas de verificación como guías. Sin embargo, su uso en el entorno de habla francesa sigue siendo limitado. No obstante, realizar una evaluación de la evaluabilidad es una forma interesante de mejorar la comprensión del programa a evaluar y el desarrollo de las herramientas necesarias para planificar futuras actividades de evaluación. Fue muy útil en varios aspectos en el contexto de la evaluación del programa utilizado aquí como ejemplo, primero, promoviendo el conocimiento del programa entre las partes implicadas, luego, permitiendo que se desarrolle el modelo lógico y, finalmente, ayudando a planificar actividades de evaluación. Los resultados obtenidos sugieren que hubiera sido difícil no hacerlo en un programa cuyo desarrollo fue un desafío para sus promotores.

Para los novatos en este enfoque, es posible utilizar los pasos y las listas de verificación propuestas por los autores. Sin embargo, debe tenerse en cuenta que en la realización de una evaluación de la evaluabilidad influyen las características del programa, de las partes implicadas en el proceso, así como el contexto en el que se lleva a cabo. Por tanto, puede ser necesario adaptar el proceso a estas realidades. Es posible que la evaluación de la evaluabilidad no pueda responder todas las preguntas, pero puede ser una alternativa útil para evitar el *status quo* o la realización de una evaluación cuyos resultados no puedan ser utilizados por ninguna de las partes.

Referencias clave

Lessard, S., Stich, C. et Montreuil, A. (2012). *Évaluation de l'évaluabilité du Plan québécois de prévention du tabagisme chez les jeunes 2010-2015*. Québec : Institut National de Santé Publique du Québec.

Este informe de investigación describe la realización de la evaluación de la evaluabilidad del Plan de prevención de tabaco juvenil de Québec 2010-2015. Primero presenta el mandato dado al equipo de evaluación, así como el método adoptado y los resultados obtenidos. Este es un documento excelente que muestra el proceso de validación del modelo lógico con actores clave.

Burrows, S., Bilodeau, A. et Litvak, E. (2012). *Étude de la faisabilité de l'évaluation et des possibilités de recherche du Plan d'action montréalais 2012-2014 de la table régionale sur les saines habitudes de vie et la prévention des problèmes liés au poids – Phase 1*. Montréal : Direction de santé publique, Agence de la Santé et des Services Sociaux de Montréal.

Este informe se refiere a una evaluación de la evaluabilidad realizado para determinar el potencial de investigación y/o evaluación de varios proyectos en el marco del Plan de Acción de Montreal de la Mesa regional sobre estilos de vida saludables y prevención de problemas relacionados con el peso. El documento proporciona una visión general de cómo se puede llevar a cabo una evaluación de la evaluabilidad para varios proyectos al mismo tiempo.

Trevisan, M. S. et Walser, T. M. (2014). *Evaluability assessment : Improving evaluation quality and use*. Thousand Oaks: SAGE Publications, Inc.

Este libro, en inglés, es uno de los más recientes y, por lo tanto, uno de los más actuales en la evaluación de la evaluabilidad. Los presentan dos de los autores más prolíficos en este enfoque. Lo que hace que sea interesante para consultar es la referencia a estudios de casos, que nos permiten comprender cómo realizar una evaluación de la evaluabilidad. Además, los autores abogan por un modelo de la realización de una evaluación de la evaluabilidad en cuatro etapas, e introducen en este trabajo otros conceptos fundamentales para la evaluación, como el desarrollo del programa, la evaluación formativa y de impacto.

Sitio Monitoring and Evaluation NEWS

<http://mande.co.uk/category/lists/evaluability-assessments-bibliography/>

https://www.zotero.org/groups/211251/evaluability_assessments/items

Este sitio web es el primero en enumerar trabajos que datan de los primeros momentos de desarrollo de la evaluación de la evaluabilidad, desde la década de 1970 hasta 2012. Este directorio está siendo actualizado por su autor y está disponible para consulta en la plataforma de gestión bibliográfica Zotero. Aunque la mayoría de los documentos están en inglés, es un gran recurso para cualquiera que desee localizar referencias de la evaluación de la evaluabilidad.

<http://aea365.org/blog/tag/evaluability-assessment/>

Este sitio es un blog de la American Evaluation Association específicamente dedicado a la evaluación de la evaluabilidad. Se trata de publicaciones muy cortas de profesionales o investigadores sobre sus experiencias en la realización de una evaluación de la evaluabilidad. Los autores se refieren a las lecciones aprendidas de estas experiencias, que pueden ser muy útiles para aquellos que desean realizar una evaluación de la evaluabilidad por primera vez.

Referencias

- Burrows, S., Bilodeau, A. et Litvak, E. (2012). *Étude de la faisabilité de l'évaluation et des possibilités de recherche du Plan d'action montréalais 2012-2014 de la table régionale sur les saines habitudes de vie et la prévention des problèmes liés au poids – Phase 1*. Montréal : Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal.
- Contandriopoulos, A. P., Rey, L., Brousselle, A. et Champagne, F. (2012). Évaluer une intervention complexe : enjeux conceptuels, méthodologiques et opérationnels. *La Revue Canadienne d'Évaluation de Programmes*, 26(3), 1-16.
- Craig, P. et Campbell, M. (2015). *Evaluability assessment: a systematic approach to deciding whether and how to evaluate programmes and policies*. Scotland : What Works Scotland.
<http://whatworksscotland.ac.uk/wp-content/uploads/2015/07/WWS-Evaluability-Assessment-Working-paper-final-June-2015.pdf>
- Davies, R. (2013). *Planning evaluability assessments: A synthesis of the literature with recommendations*. UK : Department of International Development.

- Désy, M. (2009). *L'école en santé : recension des écrits*. Montréal: Direction de santé publique de l'Agence de la santé et des services sociaux de Montréal.
- Directions de santé publique des agences de la santé et des services sociaux de la Montérégie de Laval et de l'Outaouais. (2007). *Système D : activités de développement des compétences en milieu scolaire pour la prévention des toxicomanies et de l'usage des jeux de hasard et d'argent dans le cadre de la transition du primaire au secondaire*. Longueuil, QC.
- Dunet, D. O., Losby, J. L. et Tucker-Brown, A. (2013). Using evaluability assessment to support the development of practice-based evidence in public health. *Journal of Public Health Management and Practice*, 19(5), 479-482.
doi: 10.1097/Phh.0b013e318280014f
- Dunn, E. (2008). *Planning for cost effective evaluation with evaluability assessment*. Washington, DC: USAID.
http://pdf.usaid.gov/pdf_docs/PNADN200.pdf
- Esher, A., dos Santos, E. M., Azeredo, T. B., Luiza, V. L., Osorio-de-Castro, C. G. S. et Oliveira, M. A. (2011). Logic models from an evaluability assessment of pharmaceutical services for people living with HIV/AIDS. *Ciencia et Saude Coletiva*, 16(12), 4833-4844.
- Farber, J. M., Deschamps, M. et Cameron, R. (2002). *Analyse et évaluation de la fonction d'intervenant-pivot en réponse aux besoins qu'éprouvent les Canadiennes atteintes d'un cancer du sein sur les plans de l'information, de l'éducation et de l'aide à la prise de décisions*. Gouvernement du Canada.
<http://publications.gc.ca/collections/Collection/H39-663-2002F.pdf>
- Garand-Butcher, C. (2014). *Évaluation de l'implantation du programme Système D en Montérégie (Mémoire de maîtrise)*. Université de Montréal, Montréal, QC.
- Horst, P., Nay, J. N., Scanlon, J. W. et Wholey, J. S. (1974). Program management and the federal evaluator. *Public Administration Review*, 34(4), 300-308.
doi:10.2307/975239
- Initiative Canadienne sur le Cancer du Sein. (2002). *Analyse et évaluation de la fonction d'intervenant-pivot en réponse aux besoins qu'éprouvent les Canadiennes atteintes d'un cancer du sein sur les plans de l'information, de l'éducation et de l'aide à la prise de décisions*. Canada.

- Kania, A., Patel, A. B., Roy, A., Yelland, G. S., Nguyen, D. T. K. et Verhoef, M. J. (2013). Capturing the complexity of evaluations of health promotion interventions: A scoping review. *The Canadian Journal of Program Evaluation*, 27(1), 65-91.
- Kaufman-Levy, D. et Poulin, M. (2003). *Evaluability assessment: Examining the readiness of a program for evaluation*. Justice Research and Statistics Association, United States of America.
<https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=202398>
- Leviton, L. C., Khan, L. K., Rog, D., Dawkins, N. et Cotton, D. (2010a). Evaluability assessment to improve public health policies, programs, and practices. *Annual Review of Public Health*, 31, 213-233.
doi: 10.1146/annurev.publhealth.012809.103625
- Nay, J. N. et Kay, P. (1982). *Government oversight and evaluability assessment: It is always more expensive when the carpenter types*. Lexington, MA: Lexington.
- Ogilvie, D., Cummins, S., Petticrew, M., White, M., Jones, A. et Wheeler, K. (2011). Assessing the evaluability of complex public health interventions: Five questions for researchers, funders, and policymakers. *The Milbank Quarterly*, 89(2), 206-225.
doi: 10.1111/j.1468-0009.2011.00626.x
- Paradis, M., Racine, P. et Gagné, D. (2006). *Évaluation des interventions et des activités liées à la problématique de l'abus, la négligence et la violence à l'endroit des aînés*. Québec: Direction régionale de santé publique de l'Agence de la santé et des services sociaux de la Capitale-Nationale.
- Rog, D. J. (1985). *A methodological analysis of evaluability assessment* (Thèse de doctorat). Université Vanderbilt, Nashville, TN.
- Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, 14(1), 29-48.
- Rutman, L. (1997). Formative research and program evaluability. Dans L. Rutman (dir.), *Evaluation research methods: A basic guide* (p. 59-71). Beverly Hills, CA : Sage Publications.
- Saberi, P., Yuan, P., John, M., Sheon, N. et Johnson, M. O. (2013). A pilot study to engage and counsel HIV-positive African American youth via telehealth technology. *AIDS patient care and STDs*, 27(9), 529-532.

- Smith, M. F. (1989). *Evaluability assessment: A practical approach*. Boston Norwell, Mass. : Kluwer Academic; Distributors for North America Kluwer Academic Publishers.
- Smith, M. F. (1990). Evaluability assessment: Reflections on the process. *Evaluation and Program Planning*, 13(4), 359-364.
doi: 10.1016/0149-7189(90)90020-W
- Smith, M. F. (2005). Evaluability assessment. Dans S. Mathison (dir.), *Encyclopedia of evaluation* (p.137-140). Thousand Oaks, CA: Sage Publications.
- Soura, B. D., Bastien, R. et Fallu, J.-S. (2016). Étude d'évaluabilité d'une intervention visant à prévenir l'usage de substances psychoactives lors de la transition primaire-secondaire. *Revue Canadienne d'Évaluation de Programme*, 31(2), 211-231.
- Soura, B. D., Dagenais, C., Bastien, R., Fallu, J.-S. et Janosz, M. (2016). L'étude d'évaluabilité : pertinence pour l'évaluation de programme. *Revue Canadienne d'Évaluation de Programme*, 31(1), 18-33.
- Thurston, W. E. et Potvin, L. (2003). Evaluability assessment: A tool for incorporating evaluation in social change programmes. *Evaluation*, 9(4), 453-469.
- Touati, N. et Suárez-Herrera, J. C. (2012). L'évaluation des interventions complexes : quelle peut être la contribution des approches configurationnelles?. *La Revue Canadienne d'Évaluation de Programmes*, 26(3), 17-35.
- Trevisan, M.S. (2007). Evaluability assessment from 1986 to 2006. *American Journal of Evaluation*, 28(3), 290-303.
<http://doi:10.1177/1098214007304589>
- Trevisan, M. S. et Walser, T. M. (2014). *Evaluability assessment : Improving evaluation quality and use*. Thousand Oaks, CA: Sage Publications.
- Trevisan, M. S. et Yi, M. H. (2003). Evaluability assessment: A primer. *Practical Assessment Research et Evaluation*, 8(20), 2-9.
- Van Voorhis, P. V. et Brown, K. (1987). *Evaluability assessment: A tool for program development in corrections*. National Institute of Corrections.
- Vaugeois, P., Germain, M. et Cunha Rêgo, L. (2008). *Programme APTE (Rapport synthèse)*. Montréal : Centre québécois de lutte aux dépendances.

- Wholey, J. S. (1976). Role of evaluation and evaluator in improving public programs: Bad news, good news, and a bicentennial challenge. *Public Administration Review*, 36(6), 679-683.
doi: 10.2307/975063
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. *New Directions for Program Evaluation*, (33), 77-92.
doi: 10.1002/ev.1447
- Wholey, J. S. (2010). Exploratory evaluation. Dans J. S. Wholey, H. P. Hatry et K. E. Newcomer (dir.), *Handbook of practical program evaluation* (p. 81-99). San Francisco, CA: Jossey-Bass.

Resumen / Résumé / Abstract

La realización de una Evaluación de la Evaluabilidad (EE) es un paso importante en la planificación de la evaluación real de una intervención o programa. Es un ejercicio que permite recopilar datos rápidamente sobre el progreso de un programa con el fin de tomar decisiones sobre posibles cambios o realizar una evaluación formal. En el caso presente, la EE se aplicó a un programa de prevención de abuso de sustancias basado en la escuela. Se describen las diferentes etapas de la conducción de una EE, los resultados obtenidos y las lecciones aprendidas. Además, se presentan las fortalezas y debilidades de la EE, como una fase previa a la evaluación de un programa, y los desafíos de su implementación.

Réaliser une étude d'évaluabilité (ÉE) est une étape importante permettant de planifier l'évaluation proprement dite d'une intervention ou d'un programme. C'est un exercice qui permet de collecter rapidement des données sur le déroulement d'un programme afin de prendre des décisions sur d'éventuelles modifications à apporter ou la conduite d'une évaluation formelle. Dans le cas actuel, l'étude d'évaluabilité a été appliquée à un programme de prévention de l'usage de substances psychoactives en milieu scolaire. Les différentes étapes de la conduite d'une étude d'évaluabilité y sont décrites ainsi que les résultats obtenus et les leçons apprises. Par

ailleurs, les forces et faiblesses de l'étude d'évaluabilité, en tant que phase pré-évaluative d'un programme, et les défis de sa mise en œuvre sont présentés.

Conducting an evaluability Assessment (EA) is an important step in planning the actual evaluation of an intervention or program. It is an exercise that allows to quickly collect data on the implementation of a program in order to make decisions about possible changes or conduct a formal evaluation. In this actual case, EA was applied to a school-based substance abuse prevention program. The different steps to conduct an EA are described as well as the results and the lessons learned. In addition, the strengths and weaknesses of EA, as a pre-evaluative phase of a program, and the challenges of its implementation are presented.

Biessé D. Soura actualmente realiza una pasantía postdoctoral en los *Centros para el Control y la Prevención de Enfermedades* (CDC) en Atlanta en el servicio de monitoreo, evaluación y análisis de datos. Como parte de sus actividades, realiza misiones a países que reciben fondos de emergencia para la lucha contra el VIH del gobierno estadounidense, como Ruanda, Camerún, Costa de Marfil y Haití. Es miembro del programa de capacitación en Prevención, Promoción y Políticas Públicas (4P) de la *Red de Investigación de Salud de la Población de Quebec* (RRSPQ). Tiene un doctorado en psicoeducación (Universidad de Montreal), una maestría en psicología organizacional y del trabajo (Universidad Paris Descartes) y un certificado en salud pública (Universidad Emory, Atlanta). Su tesis se centra en la evaluación de los procesos y los efectos de una intervención dirigida a prevenir el uso de sustancias psicoactivas (SPA) en las escuelas y recibió uno de los premios a las mejores tesis de la RRSPQ para 2016. Investiga sobre el uso de enfoques innovadores y participativos para el monitoreo y la evaluación de programas, la prevención del consumo de VIH y SPA, métodos de investigación mixtos y transferencia de conocimiento.

Jean-Sébastien Fallu es profesor asociado en la Escuela de Psicoeducación de la Universidad de Montreal. Tiene un doctorado en psicología de la misma universidad y completó una beca posdoctoral en el Centro de Adicción y Salud Mental, afiliado a la Universidad de Toronto. Es investigador habitual en el Instituto Universitario de Adicciones del Centro Integrado de

Salud y Servicios Sociales del Centre-Sud-de-l'Île-de-Montréal, el Grupo de Investigación sobre Sustancias Psicoactivas y el Instituto de investigación en salud pública en la Universidad de Montreal. También trabaja como Director de la revista Drogas, salud y sociedad. Sus intereses de investigación incluyen la etiología, prevención y reducción de los daños del consumo problemático de sustancias, así como las políticas relacionadas. Fundador del Grupo de Investigación e Intervención Psicosocial de Montreal, este logro le valió el premio Forces Avenir 2002 en la categoría Sociedad, Comunicación y Educación.

Robert Bastien es investigador en el *Departamento de Salud Pública* (DSP) de Montreal y profesor clínico asistente en el Departamento de Medicina Social y Preventiva de la Universidad de Montreal. Tiene una maestría y un doctorado en educación, así como una licenciatura en diseño. Está interesado en políticas de prevención, discursos y prácticas. Su trabajo intenta identificar los vectores de influencia de la investigación social en las políticas e identificar nuevas palancas para que la investigación contribuya a renovar políticas, prácticas y servicios. Al colaborar con personas que trabajan en los campos de las artes, la sociedad y la salud, provoca el encuentro entre varias disciplinas, desarrolla nuevas preguntas y experimenta con nuevas metodologías de investigación y creación.

Frédéric N. Brière es profesor asistente en la Escuela de Psicoeducación de la Universidad de Montreal e investigador habitual del Grupo de Investigación sobre Entornos Escolares (GRES), el equipo RENARD y el Instituto de Investigación en Salud Pública de Universidad de Montreal (IRSPUM). Ha estado interesado en el desarrollo, las consecuencias y la prevención de la depresión en adolescentes durante años. En términos más generales, su trabajo se centra en la evaluación de programas para promover un buen desarrollo psicosocial para jóvenes y las mejores formas de transferir estas intervenciones basadas en la evidencia de la investigación a la práctica.

Cita

Biessé Diakaridja Soura, Jean-Sébastien Fallu, Robert Bastien y Frédéric N. Brière (2020). El estudio de evaluabilidad. Una intervención de prevención del uso de drogas en la Escuela en Quebec. En *Evaluación de las*

intervenciones sanitarias en salud global. Métodos avanzados. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 13-41. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

2. Evaluación de la sostenibilidad

Una intervención de financiación basada en los resultados en Malí

MATHIEU SEPPEY Y VALÉRY RIDDE

Evaluación de la sostenibilidad

Definición del método

La sostenibilidad es el proceso que conduce a la continuidad de una intervención. La evaluación de la sostenibilidad es el estudio de la presencia de rutinas para asignar un grado de sostenibilidad (nulo, precario, bajo, medio, alto). La evaluación de la sostenibilidad consiste en analizar los procesos que han permitido que la intervención perdure en el tiempo y explicar el grado de continuidad.

Fortalezas del método

- Comprender mejor el proceso de sostenibilidad, un fenómeno complejo en un contexto específico, que debe estudiarse en múltiples dimensiones.
- Desarrollar un análisis basado en un conjunto amplio de puntos de vista
- Triangular los métodos de recogida de datos (entrevistas, observaciones, investigación documental)

Desafíos del método

- Necesidades de recursos (tiempo, presupuesto) para analizar los datos
 - Tener en cuenta la validez interna de la información (posibles sesgos)
 - Dificultad para comprender el concepto y moverse hacia atrás y adelante, entre la teoría y los datos de campo
-

En este capítulo se presenta la evaluación de la sostenibilidad mediante el ejemplo de una evaluación de una intervención sanitaria en varios centros de salud de Malí, concretamente un proyecto piloto de financiación basada en los resultados (FBR).

Entre las diferentes metodologías posibles para evaluar la sostenibilidad de una intervención, la que se utiliza es el estudio de casos (Yin 2013), que puede ser un caso único, un caso múltiple, casos con niveles de análisis anidados u holísticos, etc. El estudio que se presenta, es el de un solo caso con diferentes niveles de análisis anidado, es decir, diferentes centros de salud que ofrecen diversos servicios.

Hay muchas evaluaciones de sostenibilidad que se limitan a estudiar el aspecto financiero de la intervención o sus efectos a largo y mediano plazo.

La utilización del estudio de casos permite investigar más dimensiones de la intervención que se está evaluando: sus recursos, su funcionamiento, sus determinantes (facilitadores o barreras), sus efectos a corto y largo plazo, etc. Por consiguiente, estas dimensiones se investigarán mediante diferentes instrumentos de recogida de datos (entrevistas, observaciones, investigación documental, grupos de discusión, etc.), lo que permitirá triangular la información reunida y comprender así mejor el fenómeno general objeto de estudio. Cabe señalar que, además del estudio de casos, en la evaluación que aquí se presenta, se utiliza también un marco conceptual de sostenibilidad de la literatura científica. Este marco permite adoptar una perspectiva sistémica del fenómeno estudiado y estructurar mejor la evaluación centrándose en las dimensiones esenciales y subyacentes del fenómeno.

La metodología del estudio de casos es muy útil en la evaluación de fenómenos complejos como la sostenibilidad, ya que permite obtener el máximo de información sobre diversas dimensiones relacionadas con el objeto de estudio (Contandriopoulos y otros 1990; Lamoureux 2000; Yin 2013). Según Robert Yin (2013), este tipo de enfoque metodológico debería, sin embargo, limitarse a tres situaciones: 1) cuando sea necesario responder a preguntas de investigación del “¿Cómo?” “2) cuando el investigador pueda ejercer muy poco o ningún control sobre la intervención, y 3) cuando el objeto del estudio sea actual, lo que permite la exploración directa del sujeto estudiado. Esta metodología también es necesaria cuando los agentes que participan en la evaluación son numerosos, diversos y tienen puntos de vista diferentes. La utilización de un marco conceptual en un estudio de casos también facilita la comprensión del objeto del estudio al identificar los diversos elementos relacionados con él.

Sin embargo, hay varios problemas de aplicación de esta metodología. El primero está relacionado con la cantidad de información necesaria para comprender el caso. Cuanto más compleja sea la intervención a evaluar, más información necesitará el investigador para comprenderla. Además, para ciertos detalles de la intervención, habrá que reunir información de determinados especialistas o expertos. El segundo desafío está relacionado con la complejidad del análisis de estos datos, que provienen de una gran variedad de fuentes (documentos, entrevistas, observaciones, etc.) y actores. Incluir toda esta información en una sola línea explicativa puede ser difícil.

Sin embargo, hoy en día se utilizan varias herramientas informáticas y programas de software para ayudar a los/as investigadores/as a realizar este tipo de análisis.

Antecedentes

Malí

El estudio de casos que se presenta en este capítulo se centra en un proyecto piloto de financiación basada en los resultados que se está llevando a cabo en Malí. Malí es un país Saheliano sin litoral del África occidental que ocupa el puesto 179 en el Índice de Desarrollo Humano en 2015 (Programa de las Naciones Unidas para el Desarrollo 2015). En 2015, el país tenía más de 15,8 millones de habitantes, la mitad de los cuales (50,6%) vivía por debajo del umbral de pobreza con 1,25/Dólares día en paridad de poder adquisitivo (ibíd.). En 2015, la mortalidad neonatal y la de los menores de cinco años eran muy elevadas, con 77,6 y 122,7 muertes por cada 1.000 nacidos vivos, respectivamente. En 2014, el gasto total en salud fue sólo el 6,9% del PIB (División de Estadística de las Naciones Unidas 2016).

Malí es un país con una administración muy descentralizada, con varios niveles de representación: nacional, regional, de distrito, de zona y comunal; también podrían añadirse aldeas y barrios, pero no tienen competencias en los ámbitos de la salud y la educación (Syll 2005). La organización del sistema de salud basado en este modelo incluye hospitales nacionales, especializados y regionales, que están a cargo de la Dirección Nacional de Salud (DNS). Estas instituciones sanitarias representan el nivel terciario (de 2ª o 3ª nivel) de la prestación de atención de la salud y están distribuidas en las ocho regiones de Malí. Los primeros centros de referencia (nivel secundario) son los centros de salud de referencia (CSREF) que están presentes en los distritos sanitarios vinculados al nivel administrativo de las zonas (que no deben confundirse con los distritos de Bamako) (Observatorio de Salud de África 2014). En los servicios de primera línea se encuentran los centros de salud comunitarios (CSCOM), que están situados en las zonas sanitarias vinculadas a las comunas. La gestión cotidiana de los centros de salud comunitarios es privada, ya que de ella se ocupa una asociación de salud comunitaria (ASACO) integrada por representantes electos de las distintas

aldeas de la comuna y que vela por el buen funcionamiento de los centros de salud comunitarios y la prestación del conjunto mínimo de actividades (MPA) (Observatorio de Salud de África 2014; Audibert y de Roodenbeke 2005). La MPA representa todos los servicios que un CSCOM debe ofrecer, es decir, servicios de promoción, prevención y atención. Las actividades van desde el simple diagnóstico y tratamiento de una enfermedad hasta las campañas de vacunación, el seguimiento de niños sanos (SES), las sesiones de información, educación y comunicación (IEC), las consultas pre y postnatales (CNA y CNP), los partos, la planificación familiar (PF) y otros. Los tres componentes del CSCOM ofrecen estos servicios: dispensario, la farmacia y la maternidad.

Financiación basada en los resultados

El Organismo Neerlandés de Desarrollo (SNV) y el Real Instituto Tropical (KIT) han ejecutado un proyecto piloto sobre financiación basada en los resultados en Malí, concretamente en la región de Koulikoro, una de las más pobres del país (Ministerio de Salud de Malí 2012; Secretaría General 2009; Toonen et al. 2014). El proyecto se ejecutó entre febrero de 2012 y diciembre de 2013 en los distritos sanitarios de Dioïla, Fana (ambos en el cerco de Dioïla) y Banamba (zona de Banamba) y contó con la participación de un total de 26 CSCOM (18 para el distrito de Dioïla y 4 para cada uno de los distritos de Fana y Banamba) y 3 CSREF (1 para cada distrito) (Toonen y otros, 2014).

Este proyecto piloto tenía por objeto aumentar la calidad y la cantidad de la atención ofrecida en esos distritos sanitarios mediante mecanismos de verificación, participación comunitaria, la rendición de cuentas en la gestión de los fondos, los vínculos entre la motivación y los resultados, y las adquisiciones (Toonen y otros 2014; Toonen, Kone y Dao s.f.). La adquisición consistió en el pago de una subvención para los servicios de salud a los que se dirigía el proyecto, que evaluó esos servicios cualitativa y cuantitativamente.

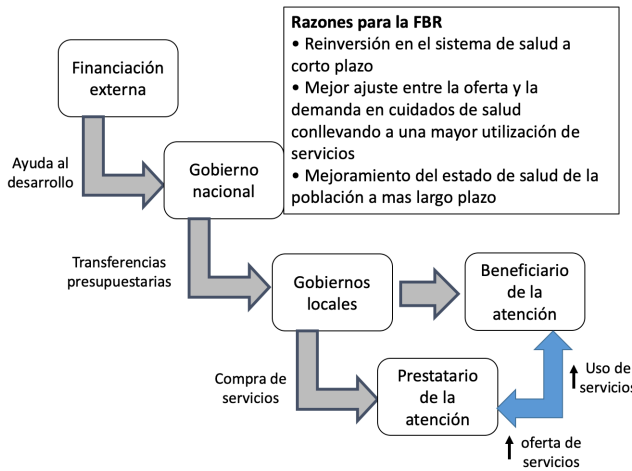


Figura 1.
Modelo lógico
de
financiación
basado en
resultados
(FBR)

Estos subsidios, comúnmente conocidos como “primas”, se distribuyeron a los centros de salud (60%), así como a los proveedores de servicios de salud (40%). La proporción se invirtió para los CSREF. Estas primas se repartieron entre el personal sanitario, que recibió cada uno un promedio de 140.500 francos CFA por trimestre (para un total de 65.419.921 francos CFA/\$138.100 dólares canadienses para el proyecto), y las organizaciones de salud, que recibieron un total de 98.129.882 francos CFA (207.150 dólares canadienses) (Toonen y otros, 2014). Las primas individuales constituían una parte sustancial de los ingresos de la mayoría de los empleados, que podían así hasta duplicar sus ingresos (Toonen y otros, 2014).

En la aplicación del FBR participaron diversos agentes, con diferentes funciones: los equipos técnicos de las CSCOM y los CSREF (servicios de atención), los ayuntamientos y las ASACO (contratación y compras), el equipo de gestión de distrito y la Dirección Regional de Salud (DRS) (regulación/verificación a nivel de las CSCOM y los CSREF, respectivamente) y diversas organizaciones no gubernamentales independientes (ONG) (verificación cruzada de resultados).

Al final del proyecto (2014), se realizó una evaluación final sobre el traspaso de responsabilidades al gobierno de Malí y la ampliación de la escala

(Toonen et al. 2014), pero sin abordar el elemento de la sostenibilidad del proyecto. Por lo tanto, era pertinente evaluar la sostenibilidad del proyecto de la FBR en Malí, así como el proceso de sostenibilidad.

Plan de evaluación

En este estudio, la evaluación de la sostenibilidad se centró en dos objetos: el proceso de sostenibilidad y el mantenimiento o estado de situación de la intervención. Por consiguiente, la evaluación de la sostenibilidad se centra en los diferentes mecanismos de sostenibilidad futura, mientras que la evaluación del estado de situación, se centra en los logros de una intervención en un momento determinado (Tabla 1).

Tabla 1 : Plan de evaluación de un proyecto de FBR en Malí

	Sostenibilidad	Continuidad
Objetivos de la evaluación	Comprender el proceso de sostenibilidad	Evaluar el nivel de sostenibilidad del proyecto
Elementos de evaluación	Determinantes (Sostenibilidad) Fases (sostenibilidad) Eventos	Actividades “rutinizadas” (a través de determinantes y fases)
Recogida de datos	Entrevistas semiestructuradas Investigación documental	
Tipo de análisis	Correspondencia con un marco conceptual Elaboración de una explicación inductiva	Correspondencia con un marco conceptual

A pesar de esta diferencia, la evaluación de procesos y resultados está íntimamente relacionada, ya que ambas se basan en un fundamento común; los determinantes de la sostenibilidad son la memoria organizativa, los códigos/valores, las normas/procedimientos, la capacidad de adaptación y las relaciones con las partes interesadas (véase Seppey y otros 2017 para más información sobre estos determinantes). Una característica especial de la evaluación de la sostenibilidad es que es el resultado de un juicio del investigador que asigna un nivel de sostenibilidad a la intervención en relación con los datos recogidos sobre el terreno.

Metodología

Revisión de la literatura y marco conceptual

La evaluación de un fenómeno complejo requiere una buena comprensión previa de sus diversos mecanismos, determinantes u otros componentes, de ahí la necesidad de iniciar la evaluación con una evaluación de la bibliografía sobre sostenibilidad. Para este estudio de caso, varios artículos científicos, informes organizativos, presentaciones y otros documentos contribuyeron a una comprensión sistemática y a la elaboración de un marco conceptual que reúne los diversos elementos de la sostenibilidad. Estos elementos se presentan en la figura 2: i) sus determinantes, ii) el proceso de sostenibilidad, iii) el nivel de sostenibilidad resultante de los elementos anteriores, iv) los diferentes contextos en los que se aplica la intervención, y v) los acontecimientos críticos y el tiempo (Johnson y otros, 2004; Chambers, Glasgow y Stange, 2013; Pluye, Potvin y Denis, 2004; Moullin y otros, 2015). Por consiguiente, el trazado del marco conceptual permite vincular los diferentes subconceptos de la sostenibilidad (adaptación, contexto organizativo, planificación de la sostenibilidad o sostenibilidad débil) a los mecanismos.

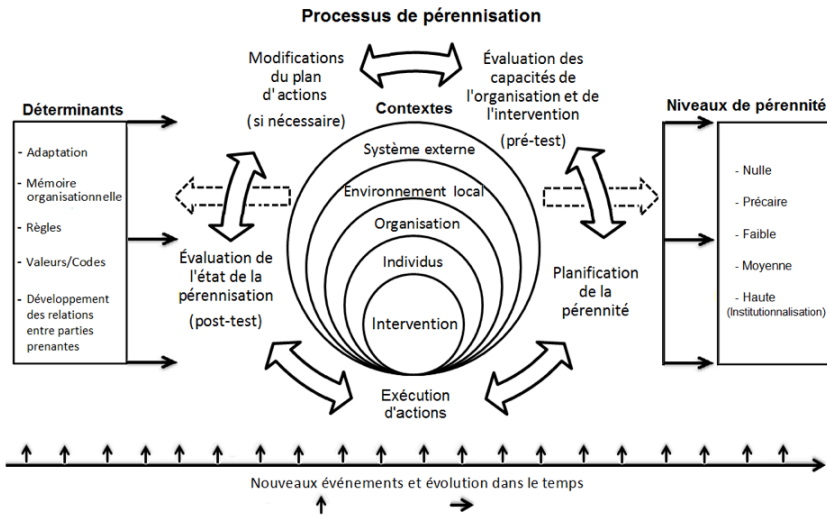


Figura 2. Marco conceptual de la sostenibilidad. Adaptado de Johnson y otros (2004), Chambers y otros (2013), Pluye, Potvin y Denis (2004) y Moullin y otros (2015).

Proceso de sostenibilidad		
	<ul style="list-style-type: none"> · Modificaciones del plan de acción (si es necesario) · Evaluación de capacidades de la organización y de la intervención (pre-test) · Evaluación del estado de la sostenibilidad (post-test) · Planificación de la sostenibilidad · Ejecución de acciones 	
Determinantes		Niveles de sostenibilidad
<ul style="list-style-type: none"> · Adaptación · Memoria organizacional · Reglas · Valores/códigos · Desarrollo de relaciones entre las partes involucradas 	<ul style="list-style-type: none"> Contexto Sistema externo Ambiente local Organización Individuos Intervención Ejecución de acciones Nuevos eventos y evaluación en el tiempo 	<ul style="list-style-type: none"> · Nulo · Precario · Débil · Medio · Alto (institucionalización)

Construcción y validación de la herramienta

El instrumento de recogida de datos es la guía de entrevista semiestructurada que surgió del examen de la literatura y del marco conceptual. Para este estudio de caso, se utilizó como base el modelo de una guía previamente utilizada durante la investigación sobre la sostenibilidad de las intervenciones en Haití (Ridde, Pluye y Queuille 2006). Sin embargo, para corresponder mejor al marco conceptual, este modelo tuvo que ser adaptado con la incorporación, eliminación y modificación de ciertas cuestiones.

Esta guía de entrevistas tiene cinco secciones relacionadas con los diferentes determinantes del marco y permite recopilar información relacionada con diferentes acontecimientos críticos también vinculados al proceso de sostenibilidad. Las preguntas eran abiertas y permitían reunir un máximo de información relacionada con los temas tratados. Dado que la guía que sirvió de modelo se había utilizado en Haití, hubo que adaptarla al contexto maliense (Ridde, Pluye y Queuille 2006). Esta adaptación se hizo en dos etapas, primero a través de un colega maliense y luego durante una visita sobre el terreno en la que se probó la guía con posibles participantes. Posteriormente, se hicieron modificaciones en el vocabulario, el orden de las preguntas, la complejidad de ciertos conceptos, etc. La guía se adaptó, por tanto, a las necesidades de los participantes. Este paso de adaptar la guía de entrevistas es esencial para una recolección efectiva de datos.

Muestreo

En aras de la exhaustividad, se seleccionó una amplia variedad de centros de salud como lugares de recolección, con el apoyo de personas que trabajan en los CSREF. Para garantizar esta exhaustividad se utilizaron varios criterios de selección: sitios con alta o baja asistencia, en zonas urbanas o rurales, y con particularidades en cuanto a su ubicación. Para garantizar la obtención de datos ricos y pertinentes, también se dio prioridad a los sitios con baja rotación de personal, acceso físico seguro y un mayor número de compras por FBR. En total, la muestra de sitios incluye tres centros de salud en cada uno de los distritos de Dioila y Fana de un total de 26 posibles, así como los dos CSREF respectivos.

En cuanto a los participantes, se utilizaron dos tipos de muestreo. El primero fue razonado y estratificado y subrayó las diferencias entre los participantes y la exhaustividad de los puntos de vista (Palinkas y otros, 2015). El criterio de inclusión en el estudio fue la participación en el proyecto, y los estratos fueron los diferentes papeles de los participantes: directivos (9), cuidadores (22), miembros de la COHA o consejo de la zona (9), y representantes locales (4). Los directores de los centros actuaron como especialistas en la selección de los participantes como autoridades de los diversos centros, facilitando el acceso de los participantes (Savoie-Zajc 2007). El otro tipo de muestreo se realizó mediante el método de bola de nieve, que consiste en obtener la ayuda de un participante para encontrar otro participante que nos ayude a encontrar otro participante, y así sucesivamente (Pires 1997). Este tipo de muestreo tuvo muchos participantes similares: personas que participaron en el diseño del proyecto, así como en su ejecución. Los participantes procedían de diversas organizaciones (Direcciones Nacionales y Regionales de Salud, SNV y KIT; cinco en total) y fueron seleccionados por recomendación de los agentes relacionados con el FBR que se habían reunido previamente. El muestreo final alcanzó un nivel de saturación de datos con un total de 49 entrevistas. La calidad del muestreo puede afirmarse por tres elementos: 1) la representatividad en número de la muestra en relación con la población objeto de estudio (por ejemplo, algunas centros de salud tenían menos de 10 empleados, mientras que el número objetivo de participantes en el estudio era de cinco por cada centro), 2) la participación de diferentes actores en la selección de los participantes (la cooperación entre los investigadores y los diversos actores sobre el terreno redujo el sesgo de selección), y 3) la exhaustividad de los grupos de actores (por ejemplo, el médico, la matrona, el gerente, el diseñador).

Recogida de datos

La fase de recogida de datos se llevó a cabo en dos etapas: del 22 al 29 de diciembre de 2015 y del 6 al 10 de febrero de 2016. Estas dos fases consistían en viajar (en autobús, motocicleta, carro, autostop, etc.) a los diferentes lugares que podían estar muy alejados (varios kilómetros de pistas de la carretera asfaltada) y reunir a un número máximo de participantes durante un período de uno a tres días. La segunda fase de la recolección fue más selectiva, ya que consistió en llegar a las diversas personas que no pudieron

participar en la investigación en la primera fase. Los diversos contactos de estas personas que estaban ausentes o se sumaron a la toma de muestras de la bola de nieve se hicieron durante la primera fase de la recogida, lo que facilitó la concertación de citas y redujo el tiempo asignado a la recogida de datos.



Ejemplo de una clínica (Fuente: M. Seppey)

Todas las entrevistas se grabaron con el consentimiento de los participantes y duraron desde menos de 30 minutos hasta más de 2 horas (\pm 1 hora y 15 minutos en promedio). La mayoría de las entrevistas se realizaron en francés, pero varios participantes optaron por realizarlas en bambara, el idioma local. Un intérprete acompañó al investigador y sirvió de

punto entre el bambara y el francés. Es importante señalar que varios participantes pudieron realizar sus entrevistas en francés, pero con frases o segmentos completos en bambara. Por lo tanto, la presencia de un intérprete era necesaria en todo momento.

Estas entrevistas normalmente tenían lugar en el centro de salud, que era el lugar de trabajo de los participantes. Sin embargo, las entrevistas a menudo podían verse perturbadas por personas que saludaban al equipo de investigación, personas curiosas, ruidos de motocicletas, cacareos, etc., lo que era típico del contexto local y del lugar de trabajo, donde generalmente era difícil encontrar un lugar tranquilo para la entrevista.



Ejemplo de dificultad de acceso a un centro de salud (Fuente: M. Seppey)

También se realizó una búsqueda bibliográfica, ya que es importante que los datos recopilados mediante las entrevistas sean corroborados por documentos oficiales y viceversa, ya que las intervenciones suelen ser diferentes sobre el papel que en la realidad (Chambers, Glasgow y Stange

2013). Por lo tanto, se realizó un análisis documental a lo largo de todo el estudio. Se solicitó el acceso a cada uno de los documentos mencionados en las entrevistas, pero en realidad se recogieron muy pocos documentos debido al bajo nivel de archivo en los centros de salud y al temor a ser penalizado por ello, entre otras cosas.

Gestión, procesamiento y análisis de datos

Tras la recogida de datos, se transcribieron los audios de las entrevistas. Las entrevistas en Bambara fueron transcritas por un técnico de investigación, mientras que las de francés fueron transcritas por el investigador principal. Para verificar la integridad de cada transcripción, el investigador principal tuvo que releer todas las entrevistas francesas (30 de las 49 entrevistas) con la grabación. Las grabaciones



Mahamadou Diabaté (Asistente de investigación en MISELI) y Mathieu Sepey (Investigador en IRSPUM) acampando en centros de salud (Fuente: M. Sepey)

de audio, así como las transcripciones literales, se grabaron en un sistema de almacenamiento e intercambio de información en línea en formato electrónico. Por lo tanto, los equipos de investigación de Malí y Quebec tuvieron acceso a esta información.

Todas las palabras fueron codificadas usando el software ©QDA Miner. La codificación con este programa informático consistía en asignar temas (por ejemplo, la adaptación, los procedimientos, etc.) a segmentos diferentes de las palabras, las frases y los párrafos. Este proceso de abstracción de datos (que pasa de lo empírico a lo teórico) está estructurado por un árbol de códigos que se basa en el marco conceptual. El árbol de códigos también fue revisado por pares para asegurar la presencia de cada elemento del marco conceptual.

En cuanto al análisis, se utilizaron dos tipos de análisis: el análisis de concordancia (la asociación del marco conceptual con la literalidad) y la elaboración de una explicación mediante el análisis inductivo e iterativo de

los datos que permite el surgimiento de diferentes temas (Contandriopoulos y otros, 1990; Yin 2013). El análisis de coincidencia se realizó sobre la base de los diferentes temas que se habían codificado. Para cada componente del marco conceptual se propusieron elementos empíricos que ilustraban la situación en Malí (por ejemplo, la incompreensión de ciertos términos médicos ilustra el uso de un lenguaje inapropiado). También se señalaron las lagunas entre el marco conceptual y los elementos empíricos, que fueron las principales conclusiones de la investigación. A continuación, se elaboró una explicación del fenómeno de la sostenibilidad mediante un análisis iterativo e inductivo de los datos. Varios elementos hicieron posible este tipo de análisis, entre ellos la división de la recogida de datos en dos fases: en primer lugar, se efectuó una escucha y transcripción de las grabaciones previamente a la vuelta al campo. Esto hizo posible el poder refutar o confirmar las hipótesis durante una segunda fase de la recogida de datos. Por tanto, la escucha de las grabaciones, la relectura de las notas tomadas durante las entrevistas, la codificación de las notas literales, la creación de tablas de resumen por centro de salud y por tipo, y la creación de una lista de los datos recogidos, se llevó a cabo en dos fases de recogida.

La escucha de los audios, la relectura de las notas de las entrevistas y la presentación los resultados preliminares son actividades que han permitido apropiarse mejor de los datos y analizar mejor los diversos aspectos en estudio. También se pudo hacer una explicación inductiva a través de los diversos temas (fuera del marco conceptual) que se añadieron al árbol de códigos durante la etapa de codificación.

Resumen de los resultados

Los resultados del estudio se derivan de las diferentes etapas de la recogida de datos. Para mayor claridad, los resultados se compararon con el marco conceptual y los temas añadidos (Seppey et al. 2017). Llegamos a la conclusión de que el proyecto del FRB en Malí tenía un bajo nivel de sostenibilidad ya que, según nuestra evaluación, pocas actividades se “rutinizaron” después de dos años de la finalización de la intervención, que era uno de los factores determinantes de la sostenibilidad. Además, la continuidad de las actividades de la FBR que habían perdurado no estaban aseguradas a corto plazo debido a la inestabilidad de los recursos, la falta de participación de diversos agentes o la ausencia de beneficios claros para

ellos. En cuanto a la sostenibilidad, varias fases fueron insuficientes o inexistentes, como la planificación y la evaluación. Para varias partes interesadas, la sostenibilidad de la intervención se logró mediante la ampliación de la escala, que es, sin embargo, un proceso totalmente diferente. Los rumores de que el Banco Mundial asumiría la responsabilidad de ampliar el proyecto parecieron haber eclipsado la planificación de la sostenibilidad del proyecto de marcos reguladores que se estaban llevando a cabo. El proceso de sostenibilidad también se vio fuertemente influenciado por la falta de recursos estables para sostener la intervención. Al final de la financiación del proyecto, varias actividades no pudieron continuar. Por ejemplo, no se contó con recursos para la supervisión externa, se abandonaron varias normas y procedimientos del FBR por considerarse demasiado exigentes o innecesarios.

Análisis reflexivo

Fortalezas y debilidades

El punto fuerte de este estudio sobre la sostenibilidad es la utilización de un marco conceptual que ha permitido una comprensión más sistémica y amplia del fenómeno. En el caso de la evaluación de un fenómeno complejo, nos parece indispensable utilizar ese marco para no ocultar ciertas dimensiones que podrían afectar a los resultados del estudio. Este marco también permitió estructurar mejor toda la investigación, desde la construcción de la guía de entrevistas hasta la redacción del informe final. El uso de un marco durante el análisis puede parecer muy deductivo, pero no impide el uso de un análisis inductivo que permita que surjan diferentes temas (Blais y Martineau 2006).

Una limitación de esta evaluación podría ser la ausencia de otros métodos de recogida de datos, como la observación, para validar cierta información recopilada durante las entrevistas. Por ejemplo, habría sido interesante observar las diferentes atenciones ofrecidas así como las diferentes prácticas de los centros para compararlas con los relatos de los servicios mencionados durante las entrevistas. Podrían haberse planteado varias discrepancias entre lo que se comunicó y lo que se observó. Esto puede explicarse por el hecho de que la mayoría de los participantes se

encontraban en una situación de conflicto de intereses, ya que se habían beneficiado directa o indirectamente del proyecto (a través de las primas del proyecto o de un empleo remunerado relacionado con el proyecto FBR). Esta situación se reflejó en los datos, que pueden sobrestimar los efectos y la sostenibilidad del proyecto u ocultar los aspectos negativos o insostenibles del mismo, aunque los resultados muestran en general lo contrario. Si se adoptara un método de recogida de datos más orientado al discurso (entrevistas), los participantes podrían tener una mayor propensión a responder favorablemente a las preguntas relacionadas con el proyecto. Las observaciones realizadas durante y después del proyecto FBR habrían permitido ver efectivamente los diferentes efectos del proyecto FBR y su sostenibilidad.

Cuestiones de aplicación de la evaluación

Algunos otros elementos pueden constituir tanto limitaciones como apoyos para la evaluación. Dado que el investigador era un *toubabou* (nombre común de un hombre blanco) de fuera del contexto del estudio que quería discutir el tema del FBR, una proporción significativa de los participantes lo consideraron un individuo vinculado al FBR y capaz de reintroducir el proyecto. Esta preconcepción del papel del investigador puede haber tenido un impacto negativo en los datos, ya que casi todos los participantes pidieron al final de la entrevista volver al proyecto. A pesar de las estrategias puestas en marcha para reducir este sesgo, los participantes seguían siendo objeto de deseabilidad social (Savoie-Zajc 2016). Antes de cada entrevista, se aclaró el papel del investigador y los objetivos del estudio y se hizo hincapié en la independencia del estudio del proyecto FBR. Si los participantes parecían estar hablando específicamente a favor del proyecto FBR, se les recordó que debían aclarar los objetivos del estudio.

Por otra parte, la edad y el enfoque del investigador pueden haber facilitado la entrada en el campo. Al ser más joven que casi todos los participantes entrevistados y utilizar medios de transporte (autostop, motocicletas, autobuses o a pie) y alojamiento básico (tiendas, salas de hospital, salas de visita), el equipo de investigación (el intérprete y el investigador principal) a menudo se ganaron la simpatía de las personas de los centros de salud. Esto ciertamente permitió una mejor disponibilidad y disposición de los participantes, así como un cierto nivel de comodidad con

el equipo de investigación. El uso de Bambara por el investigador principal durante el contacto y las conversaciones informales también permitió potencialmente una mejor relación de confianza con los participantes.

Lecciones aprendidas

En el presente estudio sobre la sostenibilidad se pueden señalar dos lecciones importantes sobre la forma de realizar la labor de evaluación. En primer lugar, este estudio destaca la importancia de la adaptación al contexto, que es tan importante para la sostenibilidad de una intervención como para un proyecto de investigación. Por ejemplo, en el caso de esta evaluación de la sostenibilidad, el lenguaje utilizado para las entrevistas podría ser a veces inapropiado, a pesar de la validación previa de la guía sobre el terreno. Las respuestas muy cortas y siempre positivas de algunos participantes podrían explicarse por una mezcla de deseo de complacer y una posible malinterpretación en el que sin duda influyó el bajo nivel de educación de algunos participantes. Las formulaciones también podrían estar mal adaptadas. Por ejemplo, en el contexto de Malí se considera que la “capacitación” es una actividad formal que generalmente termina con un certificado o diploma, lo que no era el caso del investigador. Por lo tanto, se hace esencial aclarar los posibles malentendidos y equívocos. Esto puede facilitarse mediante la intervención de un intérprete bien capacitado en el tema que se estudia o mediante la elección de un vocabulario adaptado a cada participante. Tan pronto como surjan dudas sobre una posible confusión, el investigador deberá aclarar la situación y tomar nota de esta nueva información en las entrevistas posteriores. Por lo tanto, el instrumento de recogida debe adaptarse constantemente, tanto la guía de la entrevista como el propio equipo de investigación en términos de vocabulario y comportamiento.

Desde una perspectiva más general, la logística de una evaluación de salud global suele requerir mucha flexibilidad y adaptabilidad debido a circunstancias imprevisibles (averías en el transporte, enfermedades, pérdida de redes de comunicación, etc.). Sin embargo, esto puede mitigarse mediante una planificación cuidadosa de la labor que debe realizarse y el tiempo que debe asignarse a ella. Planificar un día para hacer 90 km de transporte puede parecer muy fácil, pero cuando se suman las averías de los autobuses, las paradas frecuentes, los controles de carretera, etc., el

día se acaba rápidamente. La programación de demasiadas reuniones en un día también puede ser problemática, ya que pueden producirse ausencias, aplazamientos o invitaciones a diversos eventos no planificados. Por lo tanto, es esencial tomar y planificar el tiempo “libre” para aprovechar al máximo las diversas oportunidades o inconvenientes. Dividir su estudio en diferentes momentos también puede ayudarle a reunir más información (por ejemplo, planear un segundo intervalo de tiempo para reunirse con los participantes ausentes o añadidos) e iniciar el proceso de análisis antes. Así pues, una primera fase de análisis puede permitir al investigador volver a los distintos actores y obtener respuestas a las primeras preguntas. El proceso iterativo de análisis permite reunir una información más rica y validar las numerosas hipótesis preliminares.

Referencias claves

Sitio web <http://www.etudiantsetdeveloppement.org/article/comment-assurer-la-perennite-de-votre-projet-de-si>

Este sitio fue creado por una red francesa de Estudiantes y Desarrollo que promueve la transferencia de conocimientos en relación con el desarrollo entre pares. Además de la sostenibilidad, se presentan varios temas relacionados con el desarrollo.

Sitio web <https://evalorix.com/boutique/innovations-en-sante/evaluation-perennite-programmes-sante-publique/>

Evalorix.com es un sitio donde encontrará una multitud de herramientas para mejorar las prácticas en la salud y otros campos. Aquí puede encontrar herramientas de evaluación relacionadas con la sostenibilidad.

Ridde, V., Pluye, P. et Queuille, L. (2006). Évaluer la pérennité des programmes de santé publique : un outil et son application en Haïti. *Revue d'Épidémiologie et de Santé Publique*, 54, 421-431.

Este artículo es un ejemplo de un estudio de caso sobre la evaluación de la sostenibilidad de una intervención sanitaria. La guía de entrevistas que se proporciona al final de este artículo se utilizó para el estudio de casos en Malí.

Blais, M. et Martineau, S. (2006). L'analyse inductive générale : description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives*, 26(2), 1-18.

Este artículo ofrece una introducción muy concreta a lo que es el análisis cualitativo (y más específicamente el análisis inductivo). El artículo aclara las diferentes etapas del análisis inductivo y presenta un resumen de otros tipos de análisis.

Brouselle, A., Champagne, F., Contandriopoulos, A.-P. et Hartz, Z. (2011). *L'évaluation : concepts et méthodes*. Montréal : Les Presses de l'Université de Montréal.

Este libro presenta una diversidad de evaluaciones que pueden ser utilizadas en diferentes contextos. Por lo tanto, cada tipo de evaluación se explica con sus objetivos, los elementos evaluados, el proceso de evaluación, etc.

Referencias

- Audibert, M. et de Roodenbeke, E. (2005). *Utilisation des services de santé de premier niveau au Mali : analyse de la situation et perspectives*. Banque Mondiale, Région Afrique, Département du Développement Humain.
- Blais, M. et Martineau, S. (2006). L'analyse inductive générale : description d'une démarche visant à donner un sens à des données brutes. *Recherches Qualitatives*, 26(2), 1-18.
- Chambers, D., Glasgow, R. et Stange, K. (2013). The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. *Implementation Science*, 8 (1), 117.
- Contandriopoulos, A.-P., Champagne, F., Potvin, L., Denis, J.-L. et Boyle, P. (1990). *Savoir préparer une recherche : la définir, la structurer, la financer*. Montréal: Presses de l'Université de Montréal.
- Johnson, K., Hays, C., Center, H. et Daley, C. (2004). Building capacity and sustainable prevention innovations: a sustainability planning model. *Evaluation and Program Planning*, 27(2), 135-149. doi:<http://dx.doi.org/10.1016/j.evalprogplan.2004.01.002>
- Lamoureux, A. (2003). *Recherche et Méthodologie en Sciences Humaines* (2^e éd.). Laval: Éditions Études Vivantes.
- Ministère de la Santé du Mali. (2012). *Carte sanitaire du Mali mise à jour – 2011 : Rapport de synthèse*. edited by Secrétariat Général. s.l.

- Moullin, J., Sabater-Hernandez, D., Fernandez-Llimos, F. et Benrimoj, S. (2015). A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. *Health Research Policy and Systems*, 13 (1), 16.
- Observatoire africain de la santé. (2014). État de santé et tendances. http://www.aho.afro.who.int/profiles_information/index.php/Mali:Health_Status_and_Trends/fr
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. et Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544.
doi:10.1007/s10488-013-0528-y
- Pires, A. (1997). Échantillonnage et recherche qualitative : essai théorique et méthodologique. *La recherche qualitative. Enjeux épistémologiques et méthodologiques*, 113-169. Montréal: Gaëtan Morin.
- Pluye, P., Potvin, L. et Denis, J.-L. (2004). Making public health programs last: conceptualizing sustainability. *Evaluation and Program Planning*, 27(2), 121-133.
<http://dx.doi.org/10.1016/j.evalprogplan.2004.01.001>
- Ridde, V., Pluye, P. et Queuille, L. (2006). Évaluer la pérennité des programmes de santé publique : un outil et son application en Haïti. *Revue d'Épidémiologie et de Santé Publique*, 54(5), 421-431.
[http://dx.doi.org/10.1016/S0398-7620\(06\)76740-2](http://dx.doi.org/10.1016/S0398-7620(06)76740-2)
- Savoie-Zajc, L. (2007). Comment peut-on construire un échantillonnage scientifiquement valide?. *Recherches qualitatives*, 99-111(5).
- Savoie-Zajc, L. (2016). L'entrevue semi-dirigée. Dans B. Gauthier et I. Bourgeois (dir.), *Recherche sociale : de la problématique à la collecte des données* (6^e éd., p. 337-364). Québec : Presses de l'Université du Québec.
- Secrétariat général du Ministère de la Santé du Mali. (2009). *PRODESS II prolonge 2009 2011 : Composante santé*.
- Seppey, M., Ridde, V., Touré, L. et Coulibaly, A. (2017). Donor-funded project's sustainability assessment: a qualitative case study of a results-based financing pilot in Koulikoro region, Mali. *Globalization and health*, 13(1), 86.

- Syll, O. (2005). *La décentralisation en Afrique subsaharienne : Mali* (Mémoire de maîtrise). Besançon: Université de Franche-Comté.
- Toonen, J., Dao, D., Matthijssen, J. et Koné, B. (2014). *Évaluation finale : accélérer l'atteinte de l'ODD 5 dans la région de Koulikoro – Projet pilote financement basé sur les résultats dans les cercles de Dioïla et Banamba*. Amsterdam: Institut Royal Tropical.
- Toonen, J., Koné, B. et Dao, D. s.d. *Le Financement Basé sur les Résultats (FBR) au Mali*. KIT.
- United Nations Development Programme. (2015). *Human Development Reports*.
<http://hdr.undp.org/en/countries/profiles/MLI>
- United Nations Statistics Division. (2016). *Country Profile: Mali*.
<http://data.un.org/CountryProfile.aspx?crName=Mali>.
- Yin, R.K. (2013). *Case Study Research: Design and Methods*. Thousand Oaks: Sage Publications.

Resumen / Résumé / Abstract / Resumen en bambara

La financiación basada en los resultados es un tipo de reforma sobre la financiación de los sistemas de salud muy popular en África. Se están llevando a cabo varios proyectos piloto con el objetivo de proporcionar más atención sanitaria y mejorar su calidad. Sin embargo, la sostenibilidad de este tipo de intervención en África todavía no se ha estudiado suficientemente. Por consiguiente, en este capítulo se explica la forma en que se ha evaluado la sostenibilidad de esta intervención mediante una metodología cualitativa. La lectura de este capítulo permite comprender mejor el proceso de investigación que se llevó a cabo gracias a su exhaustiva presentación, desde el primer paso (por ejemplo, la realización de una revisión bibliográfica) hasta el último (por ejemplo, la asignación del nivel de continuidad de la intervención). Una sección más reflexiva permite a los/as lectores/as la implicación en la experiencia de los investigadores, al comprender mejor los retos que se plantean (tanto metodológicos como profesionales) y las lecciones aprendidas.

Le financement basé sur les résultats (FBR) est un type de réforme du financement du système de santé très populaire en Afrique. Plusieurs projets pilotes sont mis en œuvre dans le but d'offrir plus de soins de santé ainsi que d'en améliorer la qualité. Cependant, la pérennisation de ce type d'intervention en Afrique a encore été trop peu étudiée. Ce chapitre explique donc comment la pérennité de cette intervention a été évaluée au moyen d'une méthodologie qualitative. En lisant ce chapitre, on pourra mieux comprendre le processus de recherche qui a été mis en œuvre grâce à sa présentation exhaustive, de la première étape (ex. : la réalisation d'une revue de la littérature) jusqu'à la dernière (ex. : l'attribution du niveau de pérennité de l'intervention). Une section plus réflexive permet aussi aux lecteurs et lectrices de s'immiscer davantage dans l'expérience des chercheurs en ayant une meilleure compréhension des défis rencontrés (tant méthodologiques que professionnels) et des leçons apprises.

Results-Based Financing (RBF) is a new type of health financing reform that is very popular in Africa. Several pilot projects are being implemented on this continent, and elsewhere, to provide better and more health care services. However, too little research has been done on the sustainability of this type of intervention in Africa. This chapter explains how this intervention was evaluated, through the concept of sustainability, and following a qualitative methodology. By reading this chapter, we will be able to better understand the research process that has been carried out thanks to its exhaustive presentation, from the first step (e.g. the completion of a review of the literature) to the last (e.g. the attribution of the level of sustainability of the intervention). A more reflexive section also allows readers to become more involved in the researcher's experience by having a better understanding of the challenges encountered (both methodological and professional) and the lessons that they have learned.

Resumen en Bambara, traducido por Ahamadou Diakité y Mahamadou Diabaté, a quienes agradecemos.

Nafolo ko mi siysin né do diabi sɔɔɔlila (FBR) yé nafolo taabolo kura yé mi maakadi keneya kan farafina kɔnɔ. Poroje kuntiléna ciaman be sékan walassa ka sé ko furakeli lagini boŋan ka yele mali do a la ani ka keɕioŋoya numa ke o hokumu kɔnɔ taabolo bilali sé kan kuolaye farafina kɔnɔ kalan ciamaba ma a

siɲi sé kan o la. Ni Bolofara ɲinfoli ba yira ciokodi taabolo kudayi diatemilɛli kɛra ni baara mine, ɲinini taabolo bolé yuman na. Ni Bolofara kalan ni na, ɲinini taabolo fouamu bibali la sé kan sannan yelin yirali la ni kalan biolo folonɲi (Misali la : o ɲafé sebeni waleyali la) fo lavan (Misali la : taabolokudayi diaté dero dili la). Hakilina baara bolodali boyan k ata a fɛ ki sé do a ko la kɔsɔbɛ ɲinini kela yé dɔniya mi sɔɔ ɲɛleya minounou yéra faamu cioko yuman na (kaata o kalan taabolo ni o baara taabolo) ani fouamu mi sɔɔ la kɔɔ.

Mathieu Seppey descubrió el mundo de la investigación después de pasar un tiempo efectuando ayuda humanitaria. Fue cuando vio diferentes problemas en el “campo” cuando sintió la necesidad de efectuar investigación y la importancia de comprender mejor los diferentes mecanismos que se activan por los diversos proyectos o programas de intervención. Por lo tanto, Mathieu volvió a la escuela para comprender mejor las intervenciones realizadas en todo el mundo, y más concretamente en el África occidental. Mientras desarrollaba una maestría en administración de la salud, viajó a Malí para comprender mejor estos mecanismos, especialmente los de la sostenibilidad de un proyecto de financiación basado en los resultados (tema de su tesis). Hoy en día, continúa su investigación doctoral, su tesis es sobre la definición de la sostenibilidad de las intervenciones sanitarias.

Valéry Ridde es Director de Investigación en el CEPED, una Unidad Conjunta de Investigación asociada a la Universidad de París Descartes y al Instituto de Investigación para el Desarrollo (IRD). Es investigador habitual del *Institut de recherche en santé publique de l'Université de Montréal* (IRSPUM) e investigador asociado del *Centre interdisciplinaire de recherche en développement international et société* (CIRDIS) de Montreal. Es editor asociado de las revistas *BMJ Global Health* y *Global Health Promotion*, miembro del consejo editorial del *Canadian Journal of Program Evaluation* y *Qualitatif! Revue africaine de sciences sociales*. Su investigación se centra en la cobertura universal de la salud, la financiación de los servicios de salud, la evaluación de programas, las políticas de salud pública y la promoción de la salud.

Cita

Mathieu Seppey y Valéry Ridde (2020). La evaluación de la sostenibilidad. Una intervención de financiación basada en los resultados en Malí. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 43-65. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

PARTE II

LAS APROXIMACIONES CUALITATIVAS Y PARTICIPATIVAS

3. Evaluación cualitativa, informatizada, participativa e inter-organizacional (EQUIPO)

Ejemplo de un programa para mujeres víctimas de la violencia en Bolivia

MATHIEU BUJOLD Y JEAN-ALEXANDRE FORTIN

El método EQUIPO

Definición del método

Tiene tres componentes (cualitativo, participativo e inter-organizacional) y sigue seis etapas principales:

1. Introducción al campo y constitución de los equipos de evaluación
2. Adaptación del material didáctico al contexto sociocultural
3. Capacitación introductoria en metodologías cualitativas
4. Reclutamiento de participantes y recogida de datos cualitativos
5. Análisis cualitativo efectuado por ordenador
6. Foro Inter-organizacional: un espacio para compartir

Fortalezas del método

- Varios niveles de impacto potencial: institucional, personal, técnico, profesional, de colaborativo
- El componente cualitativo promueve el aprendizaje de los puntos fuertes y débiles de los programas y las ONG desde la perspectiva de los beneficiarios y añade una dimensión humana a los indicadores cuantitativos
- El componente participativo motiva a los/as empleados/as y beneficiarios al empoderarlos, al tiempo que promueve su participación democrática en el proceso de evaluación de un programa en el que participan
- El componente inter-programático e inter-organizacional proporciona una perspectiva global sobre las formas de mejoramiento de los programas

Desafíos del método

- Movilización espontánea de los/as empleados/as (no impuesta por la dirección)
- Planificación de los recursos humanos para la evaluación y las experiencias de las organizaciones
- Participación de los beneficiarios en todas las etapas de la evaluación
- Apoyo de una organización que aplica el método EQUIPO
- Convencer a los financiadores del valor añadido de los métodos cualitativos

Un programa es un conjunto organizado y coherente de actividades cuyo propósito es producir un cambio para los/as posibles participantes

(Plante 1994; Ridde y Dagenais 2012b). Los programas de desarrollo tienen como objetivo específico producir cambios en la vida de las personas afectadas por cuestiones a menudo relacionadas con factores socioculturales complejos (por ejemplo, las relaciones entre géneros). Aunque los modelos lógicos son herramientas esenciales para la planificación y evaluación de programas (Porteous 2012), las realidades sobre el terreno son a menudo mucho más complejas de lo esperado. Los procesos de cambio no pueden resumirse en una lógica lineal de causa y efecto y su evaluación requiere la participación de una multitud de partes interesadas de diferentes organizaciones. Los tres componentes del método EQUIPO tienen en cuenta estas realidades complejas.

El componente cualitativo de EQUIPO

El componente cualitativo de este método tiene por objeto comprender los procesos de cambio o inercia desde la perspectiva de la persona que experimenta una situación determinada (por ejemplo, la participación, como empleado/a o beneficiario/a, en un programa para reducir la violencia). Según las personas interesadas (beneficiarios/as, técnicos/as, directivos/as), ¿cómo y por qué un programa ha producido o no cambios en la vida de las personas beneficiarias? ¿Cuáles son los puntos fuertes y débiles del programa? ¿Cómo se puede mejorar ahora? ¿Cuáles son los desafíos que hay que superar? Para explorar estas cuestiones se pueden utilizar métodos cualitativos.

Los datos cualitativos son utilizados cada vez más en la evaluación de programas, por ejemplo, mediante entrevistas semiestructuradas o grupos de discusión. Sin embargo, el análisis de este tipo de datos representa un desafío importante (Patton 2015). El análisis temático es un método accesible para derivar el significado de un conjunto de datos cualitativos (Miles, Huberman y Saldaña 2014; Paillé 2016). Para describir este método de manera sencilla, establezcamos un paralelismo con una lectura sistemática de los textos en la que se resaltan los pasajes importantes y se marcan en el margen con códigos temáticos para recordar su contenido. Intuitivamente, el/la evaluador/a realiza análisis temáticos al leer los informes o entrevistar textualmente, destaca los pasajes pertinentes y anota su contenido. A continuación, reúne esta información en grupos en un documento que presenta el resultado de su análisis. Sin embargo, la tarea se complica cuando

es necesario hacer un análisis sistemático de varios documentos y seguir una multitud de temas (por ejemplo, logros, desafíos, facilitadores, recomendaciones). El análisis temático mediante el ordenador es un proceso riguroso de organización de estos extractos de textos destacados en categorías temáticas (por ejemplo, desafíos) que pueden contener subcategorías (por ejemplo, desafíos tecnológicos). En un proceso deductivo, estas categorías están predeterminadas, mientras que, en un proceso inductivo, surgen en el transcurso del análisis.

Los criterios de calidad para la investigación o evaluación cualitativa difieren de los de los enfoques cuantitativos (Laperrière 1997; Patton 1999). Aunque no garantiza un análisis de calidad, el uso de CAQDAS (Software de Análisis de Datos Cualitativos Asistido por Ordenador), como NVivo, puede aumentar el rigor del análisis temático (Bazeley y Jackson 2013; Bujold 2016). El aprendizaje de este tipo de software es parte de los programas de educación continua que se ofrecen a los evaluadores de programas.

El componente participativo de EQUIPO

Este componente nos recuerda que la evaluación del programa no es un fin en sí mismo, sino un medio para mejorar las condiciones sociales y comunitarias. La evaluación participativa procura empoderar a las personas interesadas haciéndolas participar activamente en el proceso de evaluación entre ellos y no sobre ellos (Cargo y Mercer 2008; Jagosh y otros, 2012). Mientras que el componente cualitativo explora los puntos de vista y las experiencias vividas por los beneficiarios y los técnicos del programa, el componente participativo de EQUIPO une a las personas más motivadas para formar equipos de co-evaluadores/as para llevar a cabo las diversas etapas del proyecto de evaluación (planificación, reunión y análisis de datos, difusión de resultados). Este enfoque también tiene por objeto cambiar la percepción negativa de la evaluación de programas, que a menudo se reduce a la rendición de cuentas y se percibe como un control del rendimiento por parte de los encargados de un programa (Ridde y Dagenais 2012a). El nivel de participación de los/as co-evaluadores/as puede variar desde una simple consulta hasta su participación en todas las etapas de la evaluación (Daigneault y Jacob 2009).

El componente interorganizacional de EQUIPO

Este componente amplía el aspecto participativo al estimular el intercambio de métodos y resultados de evaluación de programas entre las ONG que comparten misiones similares. La EQUIPO tiene por objeto el trabajo en equipo entre las personas interesadas de las organizaciones para abordar mejor las complejas cuestiones socioculturales, como la justicia de género.

Recuadro 1. Requisitos previos para la aplicación de EQUIPO

- Responder a una necesidad de evaluación cualitativa expresada por las ONG locales.
- Obtener el apoyo de la dirección de las ONG para fomentar la participación de los/as empleados y los/as beneficiarios/as.
- Poder ofrecer capacitación y apoyo de un asesor de análisis cualitativo y participativo.
- Facilitar a las personas participantes el uso de herramientas informáticas.
- Confianza en la disponibilidad de recursos de las ONG (humanos, materiales y de tiempo).
- Obtener un compromiso de colaboración interinstitucional de las ONG participantes.

Antecedentes: Justicia entre las mujeres y los hombres en Bolivia

El proyecto piloto EQUIPO se llevó a cabo como parte del programa de cooperación voluntaria “Actuar mejor e influir mejor”, ejecutado por Oxfam-Québec entre 2009 y 2015 en 11 países, entre ellos Bolivia. Una de las áreas de intervención del programa fue “Justicia entre mujeres y hombres”. La violencia, el feminicidio y el acoso a las mujeres en el trabajo siguen siendo frecuentes en Bolivia (Ministerio de justicia 2015). Debido a la importancia de esta cuestión, Oxfam-Quebec se ha asociado con el Centro Juana Azurduy (CJA) y el Centro para las mujeres Gregoria Apaza (CPMGA), dos organizaciones no gubernamentales locales con programas para promover la igualdad entre mujeres y hombres.

Tabla 1 : Descripción de las ONG asociadas

	Centro de Promoción de la Mujer Gregoria Apaza	Centro Juana Azurduy
Misión	Contribuir al reconocimiento y al ejercicio de los derechos de la mujer	Contribuir al empoderamiento de las mujeres, individual y colectivamente, fortaleciendo su capacidad de representarse a sí mismas, de reclamar sus derechos y de redistribuir equitativamente los recursos disponibles
Ubicación (datos de 2012)	El Alto (Bolivia). Altitud 4150m. Población: 848.452	Sucre, Bolivia. Altitud 2780m. Población: 261.201
Fundación	8 de junio de 1984	10 de septiembre de 1989
Comienzo de la asociación con Oxfam-Qc	2005	2003
Número de empleados (2012)	40 (29M/11H)	50 (30M/20H)
Número de beneficiarios (estimación para 2012)	2 000	1500
Poblaciones destinatarias	Urbanas. Mujeres y jóvenes. Mestizo y aymara.	Urbano y rural. Mujeres y jóvenes. Mestizos y quechuas.

A petición de la CJA y la CPMGA, Oxfam-Québec contrató en febrero de 2012 a un asesor en análisis cualitativo (Mathieu Bujold) para cubrir una necesidad de capacitación en métodos de evaluación cualitativa y participativa. Otra necesidad expresada por los asociados era la de estimular la colaboración con otras organizaciones bolivianas con misiones similares. Con este objetivo en mente, los miembros del personal de la CJA y la CPMGA se reunieron en marzo de 2012. Este último expresó la necesidad de estar equipado para llevar a cabo conjuntamente un proyecto de evaluación cualitativa y participativa de los programas de atención a las mujeres víctimas de la violencia. Para responder a esta necesidad se propuso a los asociados el proyecto piloto de cualitativa, informatizada, participativa e inter-organizacional (EQUIPO).

Tabla 2 : Objetivos del EQUIPO

Objetivo general	Fortalecer la capacidad de los asociados (en este caso, CJA y CPMGA) en los métodos de evaluación cualitativa y participativa.
Subobjetivos cualitativos	<p>Capacitar a los/as interesados/as (beneficiarios/as, técnicos/as y directivos/as) de las ONG asociadas en la recogida y el análisis de datos cualitativos mediante ordenador (NVivo).</p> <p>Generar una base de datos cualitativa para explorar las experiencias de los/as beneficiarios/as de las ONG asociadas, recoger sus opiniones sobre los programas y recoger sus recomendaciones.</p> <p>Promover el aprendizaje en la acción mediante la evaluación de los procesos de los programas de las ONG asociadas.</p>
Subobjetivos participativos	<p>Desarrollar las aptitudes de las personas interesadas en la evaluación participativa de los programas.</p> <p>Promover la participación activa y democrática de los interesados como co-evaluadores/as de los programas de las organizaciones asociadas.</p>
Objetivos inter-organizacionales	<p>Promover los intercambios entre las ONG con misiones similares o complementarias.</p> <p>Compartir las experiencias y los desafíos encontrados en la aplicación del método EQUIPO.</p> <p>Fortalecer una cultura participativa de evaluación y supervisión de las ONG asociadas.</p>

Presupuesto y compromiso de los socios

- Apoyo de Oxfam-Québec
- Un asesor, Mathieu Bujold (MB), en análisis de datos cualitativos.
- Un asesor, Jean-Alexandre Fortin (JAF), en desarrollo organizacional.
- Dos licencias completas del software NVivo.
- Dos manuales de entrenamiento traducidos al español.
- Apoyo financiero que cubre el 75% de los costos de la recogida de datos cualitativos, la transcripción y la traducción al español de los grupos de discusión y las entrevistas semidirigidas realizadas en quechua o aymara.
- Participación de las ONG locales
- Reclutamiento de co-evaluadores/as (empleados/as, aprendices y beneficiarios/as).
- Contribución financiera que cubre el 25% de los gastos de

transcripción y traducción.

- Grabadoras de audio y portátiles para la instalación del software NVivo.
- Gastos de viaje y un refrigerio ofrecido a los/as participantes durante los grupos de discusión y las entrevistas.

EQUIPO: un método de aprender haciendo

Introducción al campo y constitución de equipos de evaluación

A su llegada, Mathieu Bujold (MB), experto en análisis de datos cualitativos, empezó a colaborar con Jean-Alexandre Fortin (JAF), experto de campo y asesor de desarrollo organizacional de Oxfam. JAF había estado trabajando con las dos ONG locales durante más de cuatro años. Como primer paso, se planificaron reuniones con la dirección de la ONG y los/as trabajadoras/as interesados/as para comprender mejor sus necesidades y expectativas en relación con los enfoques cualitativos y participativos. Cada ONG seleccionó tres programas para ser evaluados (CJA: Defensoría de las mujeres, Acción Política, Trabajo Digno; CPMGA: Servicio psicoLegal, Brigadistas, Agentes Comunitarios). Se crearon entonces equipos de personas evaluadoras para participar en el proyecto piloto EQUIPO. La instrucción general era que al menos un/a empleado/a de los programas en cuestión debía participar en la evaluación de un programa distinto del suyo. Cada equipo de evaluadores/as consideró entonces a los beneficiarios potencialmente interesados en participar en el proceso de evaluación.

Adaptación del material educativo

El consultor en análisis cualitativo compartió manuales de capacitación (Bujold 2014 y 2016) y varias guías metodológicas que se han traducido al español (realización de grupos focales y entrevistas; guía de transcripción; protocolo de validación entre analistas). Dado que estas herramientas se habían desarrollado para la investigación académica, tuvieron que adaptarse al contexto de la evaluación de los programas de desarrollo. Por ejemplo, se simplificaron los aspectos teóricos y se promovió el aprendizaje en la acción. A pesar de estas atenuaciones metodológicas, se han mantenido los criterios de calidad fundamentales de la evaluación cualitativa, por ejemplo,

la optimización de la adecuación empírica entre las realidades evaluadas, los datos recogidos y los productos de la evaluación (Olivier de Sardan 2008). También se ha aplicado un proceso de validación inter-analítica que moviliza a un grupo interdisciplinario de analistas para interpretar un conjunto de datos (Miles, Huberman y Saldaña 2014).

Capacitación introductoria en metodologías cualitativas

Dado que el uso del CAQDAS requiere una comprensión de las metodologías cualitativas, se ofrecieron por primera vez dos días de capacitación sobre el tema a los/as empleados/as y pasantes de las ONG. El software NVivo se enseñó durante las etapas de recogida y análisis de datos. Se ofreció capacitación específica a los facilitadores y observadores masculinos y femeninos de los grupos de discusión (empleados/as de los co-evaluadores y de las co-evaluadoras). La función de los/as facilitadores/as es estimular el debate a fin de explorar a fondo los temas de discusión y promover la participación equitativa de mujeres y hombres. Las personas observadoras se encargan de señalar los aspectos que no son audibles en el audio (por ejemplo, las emociones) y especialmente la cronología de los intercambios (quién habló y cuándo). Las notas del/la observador/a son esenciales para la persona responsable de la transcripción. También se ofreció capacitación específica a las personas encargadas de las entrevistas (co-evaluadores/as beneficiarios/as) y las transcripciones (co-evaluadores/as en formación).

Reclutamiento de participantes y recogida de datos cualitativos

Para cada programa a evaluar, se planificaron dos grupos de enfoque: el primero con una cohorte de nuevos beneficiarios y el segundo con mujeres que habían completado recientemente el programa (< 6 meses). Dada la carga emocional del tema, los grupos de discusión fueron dirigidos por el personal de las ONG y no por los beneficiarios. Un psicólogo estaba presente en cada grupo. Se capacitó y apoyó a cuatro beneficiarios en la planificación y realización de entrevistas individuales semiestructuradas con otros beneficiarios. Las guías de debate y de entrevistas semiestructuradas fueron elaboradas conjuntamente por los/as empleados/as y los/as

beneficiarios/as. Si bien algunos grupos de discusión tenían temas de debate específicos, como la percepción de la violencia, todos los grupos tenían que abordar temas genéricos (Recuadro 2).

Recuadro 2. Temas genéricos para la discusión en los grupos de discusión de EQUIPO

1. ¿Por qué vino al centro, en primera instancia?
2. ¿Su experiencia con el centro ha cambiado algo en su vida?
3. ¿Está satisfecho/a con los resultados del entrenamiento que recibió?
4. En su opinión, ¿cuáles son los puntos fuertes y débiles del programa?
5. ¿Cómo podría el centro mejorar sus servicios?

Los grupos de discusión duraron unas dos horas, mientras que las entrevistas duraron un promedio de 30 minutos. Cada actividad de recogida de datos se registró y luego se transcribió electrónicamente para facilitar el análisis temático con NVivo. Dada la sensibilidad del tema, se prefirió la grabación de audio al formato de vídeo, lo que hace que el papel del observador sea aún más importante a la hora de señalar quién habla y cuándo. Siendo la confidencialidad particularmente importante en el contexto (violencia), se explicó a los/as participantes que sus nombres serían cambiados y que sería imposible reconocer sus palabras en los informes finales. Los aspectos éticos de la investigación cualitativa fueron objeto de capacitación para los/as co-evaluadores/as. Los grupos de discusión en comunidades remotas se realizaban en quechua o aymara y requerían el empleo de personal de traducción y transcripción. Un total de 11 grupos de discusión y 14 entrevistas con más de 80 beneficiarios de seis programas fueron realizadas, grabadas, transcritas y analizadas por equipos de analistas formados por personal de AJC y CPMGA.

Análisis cualitativo mediante el ordenador

Siguiendo un enfoque de análisis temático híbrido, tanto deductivo como inductivo (Fereday y Muir-Cochrane 2006), la codificación asistida por NVivo de las verbatim ha permitido que surjan categorías temáticas del discurso de los beneficiarios, al tiempo que se presta atención a los indicadores previstos por las organizaciones. Se aplicó un riguroso proceso de validación interanalítica de la codificación. Cada grupo elaboró primero un diccionario de codificación en el que se definían las categorías y subcategorías temáticas previstas y se daba un ejemplo de un extracto

ficticio o real que se codificaría en él. Los analistas aplicaron esta herramienta analizando individualmente el mismo grupo de enfoque con NVivo. Los miembros del equipo se reunieron entonces con el consultor de análisis para comparar su codificación y optimizar el diccionario de codificación añadiendo, fusionando o eliminando categorías y/o subcategorías. Una vez alcanzada cierta homología de codificación, es decir, cuando los analistas codificaban en general los mismos extractos de texto en las mismas categorías, el resto del corpus de datos se distribuía entre los analistas. Los analistas se reunieron entonces semanalmente con el consultor para presentar sus análisis y actualizar el diccionario de codificación. Una vez que se codificó completamente el corpus de datos, se inició una etapa de interpretación. Los analistas se repartieron la tarea de releer los extractos codificados en cada tema y registrar sus interpretaciones en memorandos, ilustrándolos con citas de los beneficiarios para optimizar la correspondencia empírica entre las realidades exploradas, los datos generados y los productos de la evaluación. Para cada categoría, se añadieron a los memorandos gráficos que representaban la distribución de la codificación por número de participantes. Estos memorandos se utilizaron para elaborar presentaciones para el foro interinstitucional.

Foro Interinstitucional: Un espacio para compartir

En octubre de 2012, los equipos de analistas se reunieron en El Alto, en las oficinas de la CPMGA, para presentar sus resultados y los métodos desarrollados. En el marco de esta reunión, se celebró un grupo de debate para examinar las experiencias de los co-evaluadores/as del proyecto piloto de la EQUIPO, su percepción de los puntos fuertes y débiles de este método y las modificaciones que deben introducirse en él.

Recuadro 3. Temas examinados en el grupo de debate durante el foro interinstitucional

1. ¿Puede describir su experiencia con este proyecto piloto?
2. ¿Cuáles son los puntos fuertes y débiles del método EQUIPO?
3. ¿Qué desafíos encontró?
4. ¿Cómo podría mejorarse este método en el futuro?
5. ¿Qué podría promover su sostenibilidad?

En las siguientes secciones se presenta un análisis de este grupo de discusión, que también siguió un proceso de análisis temático asistido por NVivo.

Impacto del proyecto piloto de EQUIPO

El análisis temático del grupo de discusión interinstitucional reveló diferentes niveles de impacto de este proyecto piloto (institucional, personal, técnico, profesional, de colaboración).

Impactos institucionales

En el plano institucional, los/as participantes observaron que esa experiencia permitía a las organizaciones aprender de sus puntos fuertes y débiles. Isis, psicóloga de la AJC, dice que este proyecto “nos ayudó a ajustar nuestras estrategias... para reconocer lo que hacemos bien y lo que necesitamos para arreglar, pensar e innovar” (Isis, AJC). (Isis, CJA). Este aprendizaje en acción se hizo desde la perspectiva de los beneficiarios de las ONG. Sonia, técnica de recursos humanos de la CPMGA, dice que ha “sido capaz de profundizar y acercarse a la realidad de la vida de las personas para descubrir nuestras fortalezas y debilidades” (Sonia, CPMGA). Según los participantes, este ejercicio de reflexión permitió ir más allá de la contemplación de las cosas buenas que se han hecho, para analizar también las lagunas que hay que llenar. Christina, Directora de la CPMGA, nos recuerda que “es muy fácil mirarte el ombligo y decirte que eres bueno (risas) [...] pero este proceso va más allá [...] también se trata de ver tus debilidades [...] para mí, la pregunta fundamental de este ejercicio es “¿Cómo podemos mejorar?”(Christina, CPMGA).

Este proyecto piloto habría proporcionado a las personas interesadas una nueva forma de evaluar los programas. Según Pilar, planificadora de programas de la CPMGA, “esta experiencia nos permitió explorar espacios que no habíamos explorado antes, ... conocer de una manera diferente cómo llegamos a la gente” (Pilar, CPMGA). La evaluación cualitativa, mediante el enfoque inductivo que propugna, habría añadido una perspectiva central sobre las realidades que se evalúan. Calvino, un psicólogo de la AJC, discute la importancia de “conocer la realidad de la situación” (Pilar, CPMGA).

Este proyecto piloto habría proporcionado a las personas interesadas una nueva forma de evaluar los programas. Según Pilar, planificadora de programas de la CPMGA, “esta experiencia nos permitió explorar espacios que no habíamos explorado antes, ... conocer de una manera diferente cómo llegamos a la gente” (Pilar, CPMGA). La evaluación cualitativa, mediante el enfoque inductivo que defiende, habría añadido una perspectiva central sobre las realidades que se evalúan. Calvino, un psicólogo de la AJC, discute la importancia de “conocer la realidad desde el fondo [de la población] hasta la cima [de la institución]” (Calvino, AJC). Esta perspectiva inductiva es un complemento esencial de los enfoques deductivos que generalmente se defienden en la planificación y evaluación de programas. La evaluación cualitativa aporta profundidad a las evaluaciones cuantitativas que generalmente se adoptan. Pilar argumenta que uno de los errores de las ONG es “planificar los objetivos del programa sólo cuantitativamente... Por ejemplo, con el indicador “Número de mujeres que han participado en este programa” [...], nos decimos a nosotros mismos: ¡que bien! Alcanzamos nuestro objetivo de cobertura (risas)!” Pilar, CPMGA”. Continúa diciendo que el proyecto ha hecho posible completar las cifras explorando..:

Cómo se siente la gente. Cómo afectamos a sus vidas. Cuando hacemos este tipo de evaluación, podemos acercarnos a sus experiencias [...]. Esto nos lleva más lejos. Escuchándolos, te das cuenta, más allá de los números, de lo que haces bien o mal (Pilar, CPMGA).

Los métodos cualitativos añaden una dimensión humana a los indicadores cuantitativos de vigilancia y evaluación preexistentes. Según el director de la CPMGA, este proyecto ha enriquecido “los indicadores en términos cualitativos y esto es muy positivo” (Christina, CPMGA). Jova hace comentarios similares: “Habíamos desarrollado muchas herramientas para medir nuestros indicadores cuantitativos, pero el ejercicio que hemos hecho va más allá de lo que hemos hecho hasta ahora. Analizar lo que la gente dice es una nueva experiencia” (Jova, CJA).

La experiencia también habría aclarado la comprensión de las metodologías cualitativas. Jova dice que se dio cuenta de que la organización tenía un concepto erróneo de lo que es cualitativo:

la formulación de nuestros indicadores, que pensábamos que eran cualitativos, apoyaba más el análisis cuantitativo... Por ejemplo, el “porcentaje de mujeres que experimentaron un aumento de la autoestima”. No hicimos ningún análisis del discurso de las mujeres para averiguar cómo habían mejorado o no su autoestima (Jova, CJA).

Impactos personales

En el plano personal, este proyecto habrá sensibilizado a los/as empleados/as y a la administración sobre la importancia de tener en cuenta las opiniones de las personas beneficiarias interactuando directamente con ellas. Jova dice que este proyecto “ha sido muy gratificante para los co-investigadores/as, hemos sido muy sensibilizados... es mucho más emocionante interactuar con la gente de esta manera” (Jova, AJC). Según Jova, esta experiencia también fue motivadora para las beneficiarias: “las niñas y las mujeres estaban muy contentas de formar parte de este proyecto. Lorena de la CJA especifica que “es importante que las personas beneficiarias sean objeto de evaluación dinámica... no sólo encuestadas. ¡No son sólo papeles!” (Lorena, CJA). El aspecto participativo de EQUIPO era coherente con los valores compartidos por la comunidad y algunos movimientos sociales de la Bolivia contemporánea. Según René, sociólogo del CPMGA, “involucrar a las personas en el proceso de evaluación es coherente con un punto de vista postcolonial” (René, CPMGA).

Implicaciones técnicas

En el plano técnico, la puesta en marcha de EQUIPO habrá sentado las bases de un sistema de organización de la información. Según Lorena, “NVivo nos permite organizar la información... tenemos que continuar porque tenemos otras ideas que han surgido” (Lorena, CJA). Más allá del análisis de datos, NVivo puede utilizarse para gestionar, organizar y centralizar toda la información relacionada con un proyecto o programa ofrecido por una organización. Lorena dice que espera “introducir a la organización en esta nueva forma de ver el trabajo”. Este tipo de herramienta puede ser utilizada para rastrear proyectos a largo plazo. René dice que con “NVivo, más allá del análisis, puedes documentar todo lo que ha sido hecho por la organización y

crear una memoria más rica” (René, CPMGA). Según el director de la CPMGA, “fue enriquecedor conocer una nueva herramienta de gestión de datos” (Christina, CPMGA). Al final del proyecto piloto, Christina inició un proyecto para centralizar y analizar con NVivo todos los informes y la literatura producidos por CPMGA entre 2007 y 2012.

Impacto profesional

En el plano profesional, a pesar de algunos problemas técnicos, el aprendizaje de los métodos y programas informáticos para el análisis cualitativo fue una ventaja para los/as co-evaluadores/as, dado que pocas personas de la comunidad de desarrollo boliviana están familiarizadas con este tipo de instrumento. Según el director de la CPMGA, “este proyecto permitió a los técnicos enriquecer sus currículos” (Christina, CPMGA). Lorena bromea: “Me atrevo a decir que en Sucre somos las únicas cuatro personas de la empresa que saben usar NVivo (risas)” (Lorena, CJA). El aprendizaje de este tipo de software sigue siendo un desafío, especialmente para aquellos que no tienen muchos conocimientos de informática: “La dificultad para mí fue la brecha tecnológica, me sentí completamente obsoleto [...] tengo que mantenerme al día con la computadora” (Jova, CJA). A continuación, subraya la importancia del apoyo de un consultor de análisis: “sin él estaríamos perdidos”; así como la reunión con sus colegas de la CPMGA [la organización asociada], que permitió un intercambio sobre los desafíos encontrados.

Impactos de la colaboración

A nivel de colaboración, los intercambios interprogramáticos e interinstitucionales habrán sido enriquecedores para las organizaciones al proporcionarles una perspectiva global sobre las formas de mejorar los programas. Según Christina, “la riqueza fundamental ha sido la interacción entre los miembros del equipo técnico de los diferentes programas para que puedan ver por sí mismos cómo podemos mejorar” (Christina, CPMGA). Lorena continúa presentando la importancia del aspecto interorganizativo del EQUIPO, que habría sentado las bases para la colaboración:

una de las cosas más importantes fue escuchar los resultados de nuestros colegas del CPMGA... una institución que conocía de lejos y me di cuenta de que teníamos mucho en común y estoy muy emocionado de seguir compartiendo nuestros métodos.

A través de este intercambio, los miembros de las organizaciones habrán adquirido una perspectiva complementaria sobre el tema de la violencia contra la mujer en otra población [Quechua vs. Aymara]. Esta experiencia también les permitió “conocer mejor a las mujeres de El Alto [de mayoría aymara] [...] para comparar lo que hemos visto en la población de Sucre [de mayoría quechua], para poder enriquecer nuestros proyectos” (Lorena, CJA).

Fortalezas, debilidades y desafíos

La importancia de la movilización espontánea de los/as empleados/as

En el análisis temático del grupo de discusión se destaca que este proyecto piloto ha permitido alcanzar el principal objetivo del método EQUIPO, que es “fortalecer las capacidades de los asociados en la utilización de métodos cualitativos informatizados y enfoques participativos de evaluación de programas”. Más concretamente, los sub-objetivos participativos de este método parecen haberse alcanzado. De hecho, se ha capacitado a las personas interesadas en el CJA y el CPMGA en la reunión y el análisis de datos cualitativos mediante ordenador; se han generado bases de datos cualitativos para explorar las experiencias de las personas beneficiarias de las ONG; y se ha estimulado un proceso de aprendizaje práctico. Los sub-objetivos participativos del proyecto promovieron la participación activa y democrática de las personas interesadas como co-evaluadores/as de los programas de las organizaciones asociadas. En consonancia con los sub-objetivos interinstitucionales, la realización de un foro creó un espacio para promover un intercambio entre la CJA y la CPMGA, dos ONG con misiones similares. Este evento también alentó el intercambio de experiencias en la aplicación del método EQUIPO y fortaleció la cultura de la evaluación cualitativa y participativa de las ONG asociadas.

Uno de los puntos fuertes de este proyecto es el hecho de que surgió de una solicitud de las ONG locales. Los proyectos de recogida de datos

cualitativos ya estaban en marcha antes de que llegara el asesor. Por ejemplo, Isis había realizado una serie de entrevistas para evaluar el impacto de un programa de empoderamiento de los jóvenes de AYC. Sin embargo, dice que “no sabía qué hacer con toda esta información, cómo transcribirla y analizarla... me frustró mucho” (Isis, AYC). (Isis, CJA). Continúa diciendo, “era muy importante para mí aprender una herramienta como NVivo porque esta investigación estaba pendiente. El acompañamiento ofrecido por la consejera le permitió analizar los datos y publicar un libro (Barahona 2012).

La movilización espontánea de los empleados es un elemento que ha favorecido la realización del proyecto en la CJA. Se observó una menor motivación entre los empleados cuya participación fue “sugerida” por la dirección de la CPMGA. Por otro lado, los empleados de esta ONG tuvieron más tiempo libre de su tiempo de trabajo, en comparación con la AJC, donde los empleados a menudo tenían que hacer el análisis en su tiempo libre. Uno de los retos compartidos por los co-evaluadores/as del AJC era “poner la evaluación cualitativa en el plan operacional anual... para darle tiempo” (Isis, AJC).

(Isis, CJA). Christina cree que es importante “crear una cultura institucional de evaluación cualitativa e incluir este método en la planificación anual, de lo contrario será relegado a un segundo plano... Nos hemos dado cuenta de la importancia de este tipo de análisis, por lo que es necesario incluirlo en nuestros objetivos” (Christina, CPMGA). Sonia, técnica de recursos humanos de la CPMGA, dijo que “reservará tiempo en los horarios de los empleados, períodos predeterminados en los que se verán unas cuantas veces al mes [...] No tengo ninguna duda de que se comprometerán, porque es una herramienta muy enriquecedora para su trabajo diario” (Sonia, CPMGA). Jova señala, sin embargo, que uno de los retos es “hacer un seguimiento de la aplicación de las recomendaciones que surjan de esta evaluación”. ¿Cómo pueden aplicarse en la medida de lo posible? ...] ¿Cómo podemos incorporarlos a nuestro trabajo para tratar de mejorar nuestras debilidades?” Jova, CJA”.

Experiencias y recursos de organización

La CJA y la CPMGA son dos ONG bien establecidas que emplean a más de 40 personas y prestan servicios a miles de beneficiarios (Cuadro 1).

Sus programas están bien establecidos y cuentan con el apoyo de diversas instituciones de cooperación internacional. Si bien los recursos de estas organizaciones siguen siendo limitados, les permiten contratar personal calificado, acoger a pasantes universitarios y adquirir equipo para la evaluación de la calidad (computadora y grabadora). Sin afirmar que el EQUIPO no puede ser replicado con ONG más pequeñas, la experiencia y los recursos de las organizaciones fueron dos factores que favorecieron este proyecto piloto.

Aumentar el nivel de participación y compromiso de los/as beneficiarios/as

Si bien el personal de las ONG participó en todo el proyecto, los beneficiarios “sólo participaron en la realización y transcripción de las entrevistas” (Jova, CJA). Según Jova, uno de los desafíos será involucrarlos en el análisis de datos. Pilar también presenta el tema de la participación a largo plazo de las personas beneficiarias: “¿Cómo conseguimos que la gente participe desde el principio y se quede? ...] Al final, no nos quedaban muchos... ¿Cómo los involucramos en el proceso? ...] Tendremos que pensar en esto en la planificación” (Pilar, CPMGA). Para Lorena, es importante fortalecer “la comunicación entre los/as co-evaluadores/as para ver otras perspectivas, abrir espacios para compartir e incluir momentos en nuestra planificación anual” (Lorena, CJA). Calvino añade que “será muy estimulante ser más participativo, no sólo para la institución, sino también para los financiadores” (Calvino, CJA).

Ser apoyado por una organización que aplique el método del EQUIPO

Si la evaluación cualitativa asistida por ordenador era una experiencia nueva para el AJC y el CPMGA, también lo era para Oxfam-Québec, que ha dado ejemplo aplicando este proceso a la evaluación de sus propios programas. De hecho, la misma serie de talleres sobre métodos cualitativos y participativos se ofreció a una docena de agentes y cooperantes de Oxfam-Québec en 2012. Se llevó a cabo un vasto proyecto utilizando el método EQUIPO en Perú y Bolivia. Se llevaron a cabo 17 grupos de discusión, que llegaron a 144 beneficiarios y empleados de 12 ONG locales [incluidas CJA y CPMGA], y se analizaron en equipo para explorar los puntos fuertes y débiles

de los programas de Oxfam-Québec con los socios. Los/as coevaluadores/as de CJA y CPMGA señalaron que apreciaban este esfuerzo de Oxfam-Québec, que da ejemplo aplicando lo que sugiere a los propios asociados. Según Jova, este es un aspecto que distingue a Oxfam-Québec: “En materia de cooperación internacional, ustedes [Oxfam] han incorporado este tipo de análisis cualitativo, y esto me parece excelente, pero las otras instituciones de cooperación no están todavía allí” (Jova, CJA).

Convencer a los donantes del valor añadido de los métodos cualitativos

Uno de los principales retos de la aplicación de la evaluación cualitativa en el contexto del desarrollo internacional es convencer a los donantes que favorecen la medición de los indicadores cuantitativos. Según Jova, la mayoría de ellos:

piden que se vigilen sólo los indicadores cuantitativos... Cuando presentamos los cualitativos, nos dicen que “hagamos un esfuerzo para cuantificar, para ver la cobertura”. Nos estamos moviendo en una dirección, pero la cooperación está en otra dirección, tenemos que seguir el modelo lógico, los indicadores cuantitativos... Realmente, conseguir una concesión para lo cualitativo ha sido difícil (Jova, CJA).

Pilar hace comentarios similares: “Quieren ver los efectos en sus indicadores cuantitativos” (Pilar, CPMGA).

Más allá de la evaluación, parecería que los métodos cualitativos no son favorecidos por los donantes al planificar los programas. Según Calvin, esto explicaría la ineficacia de muchos proyectos que no se adaptan a las realidades del terreno y a las necesidades de las poblaciones destinatarias:

varios proyectos no funcionan porque el contexto no ha sido debidamente evaluado [...] la cooperación internacional nos llega con proyectos ya desarrollados y no adaptados a la realidad. Creo que con una herramienta como NVivo podemos analizar lo que quiere la población objetivo, es una buena manera de contextualizar (Calvin, CJA).

Una vez que han visto el valor añadido de lo cualitativo en la acción, los miembros de las ONG locales deben centrarse en convencer al mundo exterior. Según René, “sabemos que lo cualitativo nos permite ver cosas que no se pueden ver con lo cuantitativo, pero el problema está en el exterior, donde lo cualitativo no se reconoce [...] es importante convencer de lo contrario nuestros esfuerzos serán en vano” (René, CPMGA). Según Jova, una de las estrategias para promover los métodos cualitativos en un contexto dominado por los números sería presentarlos como complemento de los cuantitativos: “los aspectos cualitativos se presentarán primero como complemento de los cuantitativos” (Jova, CJA).

Análisis reflexivo y conclusión: hacia los métodos mixtos

Este proyecto experimental de la EQUIPO ha contribuido a fomentar la capacidad y el interés de los asociados en la utilización de métodos cualitativos y participativos de evaluación de programas. El aprendizaje del software de análisis cualitativo ha permitido construir una base de datos cualitativos que puede ser alimentada para supervisar la aplicación de las lecciones aprendidas. Sin embargo, como lo expresa Jova al final de la sección anterior, debe considerarse que los/as evaluadores/as son más dirigidos a los enfoques cuantitativos que de los cualitativos. Presentar la primera cualitativa para complementar la cuantitativa, como propone Jova, es una estrategia introductoria que se ha utilizado en otros campos, en particular en las ciencias de la salud, que históricamente han estado muy dominadas por los enfoques cuantitativos (Creswell y Plano Clark 2011). Para continuar la labor sobre el aspecto interinstitucional, sería interesante que las dos organizaciones no gubernamentales elaboraran una base de datos complementaria que integrara sus datos cuantitativos y cualitativos.

Pilar subraya la complementariedad entre la evaluación de “indicadores cuantitativos que nos muestran, por ejemplo, cuántas personas han completado el programa, e indicadores cualitativos que nos permiten explorar su realidad, por ejemplo, si han recaído en la violencia y por qué. Entonces, ¿cómo podríamos seguir ayudándoles?” Pilar, CPMGA”. René hace hincapié en la posible contribución de los métodos mixtos en la evaluación:

“Una mirada cuantitativa nos permite explorar los impactos, mientras que los procesos se pueden entender más desde un punto de vista cualitativo. ... Una perspectiva mixta sería lo mejor” (Rene, CPMGA).

Los métodos mixtos son cada vez más valorados en la evaluación de programas (Bamberger 2012; Patton 2015; Bujold, El Sherif et al. 2018). El análisis temático asistido también se utiliza para realizar meta-análisis de los informes de evaluación de los programas de desarrollo (Bujold 2017). Una de las deficiencias de este proyecto ha sido su falta de integración del análisis de datos cualitativos con las evaluaciones cuantitativas del programa. Un siguiente paso para el EQUIPO podría ser la Evaluación Cualitativa-Cuantitativa Integrativa y Participativa Organizacional, que integra el análisis de datos cuantitativos y cualitativos en el mismo proyecto de evaluación. Sin embargo, los métodos mixtos de enseñanza y aprendizaje presentan otros desafíos que deberán tenerse en cuenta (Bujold, Hong y otros, 2018).

El establecimiento de una comunidad de práctica interdisciplinaria e intercultural es una estrategia para superar los desafíos de los métodos mixtos en la evaluación. Por ello, ofrecemos a todas las personas lectoras interesadas en los métodos mixtos la oportunidad de hacerse miembros de la *l'association Méthodes mixtes Francophonie*, que tiene por objeto promover la colaboración entre investigadores/as y profesionales francófonos de diferentes zonas geográficas y culturales.

Agradecimientos

Los autores desean dar las gracias a Oxfam-Québec y sus representantes, al CPMGA y al CJA, así como a los/as co-evaluadores/as de este proyecto piloto, a los voluntarios y a todos los beneficiarios que participaron. Este proyecto fue posible gracias al apoyo del Gobierno del Canadá (World Affairs Canada), que financió el programa “Actuar mejor e influir mejor” ejecutado por Oxfam-Québec.

Referencias clave

Patton, M. Q. (2015). *Qualitative research et evaluation methods: integrating theory and practice*. Thousand Oaks, California : SAGE Publications.

Un valioso recurso para explorar los diferentes aspectos y aplicaciones de la investigación cualitativa y los métodos de evaluación.

Miles, M. B., Huberman A. M. et Saldaña J. (2014). *Qualitative data analysis : a methods sourcebook*. Lieu d'édition : Maison d'édition.

Un instrumento esencial para profundizar en los métodos de análisis de datos cualitativos.

Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117^e Cahiers scientifiques de l'ACFAS.

Primer libro en francés y de libre acceso sobre métodos mixtos. Disponible gratuitamente en la siguiente dirección :

<http://methodesmixtesfrancophonie.pbworks.com>.

Para obtener más información sobre las ONG y asociaciones mencionadas en este capítulo y sus programas de desarrollo, los lectores pueden consultar los siguientes enlaces :

- <http://centrojuanaazurduy.org>
- <http://gregorias.org.bo/>
- <http://methodesmixtesfrancophonie.pbworks.com>
- <http://oxfam.qc.ca/>

Referencias

Bamberger, M. (2012). *Introduction aux méthodes mixtes dans l'évaluation d'impact*. Repéré sur le site d'InterAction : <https://www.interaction.org/document/introduction-aux-méthodes-mixtes-dans-l'évaluation-d'impact>

Barahona, I. (2012). *Young labor right: A study of young migrants from rural 1st and 2nd generation*. Edited by Juanas Azurduy Center. Sucre, Bolivia.

Bazeley, P. et Jackson, K. (2013). *Qualitative Data Analysis with NVivo* (éd.). Colorado: Sage Publications.

Bujold, M. (2014). *Méthodologies de recherche et d'évaluation qualitatives : de la théorie à la pratique, guide de formation*. Montréal, Canada : CAQI.

- Bujold, M. (2016). *NVivo: un support à l'analyse qualitative, guide de formation*. Montréal, Canada : CAQI.
- Bujold, M. (2017). *Analyse thématique et critique de rapports initiaux d'évaluation des portefeuilles pays d'ONU Femmes : réalisations, défis et recommandations à considérer*. New York : ONU Femmes.
- Bujold, M., El Sherif, R., Bush, P. L., Johnson-Lafleur, J., Doray, G. et Pluye, P. (2018). Ecological content validation of the information assessment method for parents (IAM-parent): A mixed methods study. *Evaluation and Program Planning*, 66, 79-88.
<https://doi.org/10.1016/j.evalprogplan.2017.09.011>
- Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. et Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117^e Cahiers scientifiques de l'ACFAS.
<http://methodesmixtesfrancophonie.pbworks.com>
- Cargo, M. et Mercer, S. L. (2008). The value and challenges of participatory research: strengthening its practice. *Annual Review of Public Health*, 29, 325-50.
doi: 10.1146/annurev.publhealth.29.091307.083824
- Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (éd.). Los Angeles : Sage Publications.
- Daigneault, P.-M. et Jacob, S. (2009). Toward accurate measurement of participation: Rethinking the conceptualization and operationalization of participatory evaluation. *American Journal of Evaluation*, 30(3), 330-348.
- Fereday, J. et Muir-Cochrane, E. (2006). Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1), 80-92.
- Jagosh, J., Macaulay, A. C., Pluye, P., Salsberg, J., Bush, P. L., Henderson, ... Greenhalgh, T. (2012). Uncovering the benefits of participatory research: implications of a realist review for health research and practice. *The Milbank Quarterly*, 90(2), 311-46.
doi: 10.1111/j.1468-0009.2012.00665.x
- Laperrière, A. (1997). Les critères de scientificité des méthodes qualitatives. Dans J.-P. Deslauriers, J. Poupert, L.-H. Groulx, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (p. 365-390). Boucherville, QC : Gaëtan Morin.

- Miles, M. B., Huberman, A. M. et Saldaña, J. (2014). *Qualitative data analysis : a methods sourcebook*. Lieu d'édition : Maison d'édition
- Ministerio de justicia. (2015). *Vidas dignas, mujeres libres de violencia*. Bolívie : Éditeur.
- Olivier de Sardan, J.-P. (2008). *La rigueur du qualitatif: les contraintes empiriques de l'interprétation socio-anthropologique*. Louvain-la-Neuve, BE: Academia-Bruylant.
- Paillé, P et Mucchielli, A. (2016). *L'analyse qualitative en sciences humaines et sociales* (4^e éd.). Paris : Armand Colin.
- Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(5 Pt 2), 1189-1208.
- Patton, M. Q. (2015). *Qualitative research et evaluation methods : integrating theory and practice*. Thousand Oaks, CA : Sage Publications.
- Plante, J. (1994). *Évaluation de programmes*. Sainte-Foy, QC : Presses de l'Université Laval.
- Porteous, N. L. (2012). La construction du modèle logique d'un programme. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes* (p. 89-107). Montréal : Les Presses de l'Université de Montréal.
- Ridde, V. et Dagenais, C. (2012a). Introduction générale à l'évaluation de programmes. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programmes* (p. 13-32). Montréal : Les Presses de l'Université de Montréal.
- Ridde, V. et Dagenais, C. (2012b). *Approches et pratiques en évaluation de programmes : nouvelle édition revue et augmentée*. Montréal : Les Presses de l'Université de Montréal.

Resumen / Résumé / Abstract

Los programas de desarrollo tienen como objetivo, producir cambios en comunidades con problemas socioculturales complejos, por ejemplo la desigualdad de las mujeres, donde se entrecruzan una pluralidad de significaciones. La evaluación de estos programas, necesita de un enfoque sistemático que movilice la colaboración entre las partes interesadas de diferentes organizaciones. Los tres componentes de la Evaluación Cualitativa

Informatizada y Participativa inter-Organizacional (ÉQUIPO), apuntan a responder a estas realidades complejas. El objetivo del componente cualitativo es comprender los procesos de cambio desde puntos de vista diversos de aquellos o aquellas que experimentan una situación compleja. El componente participativo está orientado a promover la participación democrática de las partes interesadas y recuerda que la evaluación de programas no es un fin en sí, sino un medio para mejorar las condiciones sociales y colectivas. El componente inter-organizacional estimula los intercambios entre las ONG compartiendo misiones similares. Este método de aprendizaje en la acción, ha sido desarrollado en el 2012 durante un proyecto piloto de evaluación de programas de ONG bolivianas, promoviendo la igualdad entre hombres y mujeres. El objetivo de este capítulo no es presentar los resultados de la evaluación específica de programas, sino más bien describir el método ÉQUIPO, el contexto en el cual ha emergido, y la presentación de repercusiones y desafíos de su aplicación desde las perspectivas de los co-evaluadores que participaron en este proyecto piloto.

Les programmes de développement visent à produire des changements dans des communautés touchées par des problématiques socioculturelles complexes, par exemple l'égalité entre les femmes et les hommes, où s'entrecroise une pluralité de significations. L'évaluation de ces programmes nécessite une approche systématique mobilisant la collaboration entre les parties prenantes de différentes organisations. Les trois volets d'Évaluation Qualitative Informatisée et Participative inter-organisationnelle (ÉQUIPO) visent à répondre à ces réalités complexes. Le volet qualitatif vise à comprendre les processus de changement depuis les points de vue de ceux ou celles qui expérimentent une situation complexe. Le volet participatif en favorisant la participation démocratique des parties prenantes rappelle que l'évaluation de programmes n'est pas une fin en soi, mais un moyen d'améliorer les conditions sociales et collectives. Le volet inter-organisationnel stimule les échanges entre des ONG partageant des missions similaires. Cette méthode d'apprentissage dans l'action a été développée en 2012, lors d'un projet pilote d'évaluation de programmes d'ONG boliviennes promouvant l'égalité entre les femmes et les hommes. L'objectif de ce chapitre n'est pas de présenter les résultats de l'évaluation spécifique des programmes, mais plutôt de décrire la méthode ÉQUIPO, le contexte dans

laquelle elle a émergé, et de présenter ses répercussions et les défis de son application depuis les perspectives des co-évaluateurs et évaluatrices qui ont participé à ce projet pilote.

Development programmes aim to produce change in communities affected by complex socio-cultural issues, such as gender equality, where a plurality of meanings intersect. Evaluating these programmes requires a systematic approach that involves collaboration between stakeholders from different organizations. The three components of the Qualitative Informative and Inter-Organisational Participatory Assessment (EQUIPO) aim to respond to these complex realities. The qualitative component aims to understand change processes from the perspectives of those experiencing a complex situation. The participatory component, by promoting the democratic participation of stakeholders, reminds us that programme evaluation is not an end in itself, but a means of improving social and collective conditions. The inter-organizational component stimulates exchanges between NGOs sharing similar missions. This method of learning in action was developed in 2012, during a pilot project to evaluate programmes of Bolivian NGOs promoting gender equality. The purpose of this chapter is not to present the results of the specific program evaluation, but rather to describe the TEAM methodology, the context in which it emerged, and to present its implications and the challenges of its application from the perspectives of the co-evaluators who participated in this pilot project.

Mathieu Bujold, es Doctor en antropología, e investigador posdoctoral en el Departamento de Medicina Familiar de la Universidad McGill. Sus experiencias de investigación y sus antecedentes académicos y profesionales se encuentran en la intersección de las ciencias sociales, las ciencias de la salud, las ciencias de la educación y la evaluación de programas. Desde 2007, ha ofrecido capacitación en análisis cualitativo mediante ordenador y métodos mixtos a varios estudiantes, investigadores/as y profesionales y ha sido experto invitado en varios departamentos universitarios, asociaciones de investigación y organizaciones internacionales que trabajan en el desarrollo (por ejemplo, el Banco Mundial). Además del desarrollo metodológico, sus intereses en materia de investigación incluyen el

pluralismo médico, la atención integrada de la salud, la colaboración interprofesional en la atención de la salud y la atención centrada en la persona.

Jean-Alexandre Fortin, tiene una maestría en ayuda humanitaria internacional, es coordinador del Programa de Acceso e Innovación de Oxfam Quebec en Bolivia. Sus más de diez años de experiencia en Bolivia con Oxfam-Quebec le han dado un profundo conocimiento del contexto sociopolítico del país. Su educación universitaria y su formación profesional se especializan en la gestión de programas de desarrollo para mujeres y jóvenes, en particular en las áreas de planificación, supervisión y evaluación de programas. Desde 2003, ha trabajado con varias organizaciones no gubernamentales que trabajan en el desarrollo internacional, como el Centro de Cooperación Internacional para la Salud y el Desarrollo, Canada World Youth y Oxfam-Quebec.

Cita

Mathieu Bujold y Jean-Alexandre Fortin (2020). Evaluación cualitativa, informatizada, participativa e inter-organizacional (EQUIPO). Ejemplo de un programa para mujeres víctimas de la violencia en Bolivia. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 69-94. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

4. El método foto voz

Una intervención con poblaciones marginadas para el acceso al agua potable, la higiene y el saneamiento en México

LYNDA REY, WILFRIED AFFODÉGON, ISABELLE VIENS, HIND FATHALLAH Y MARIA JOSÉ ARAUZ

El método foto voz

Definición del método

“Foto voz es un proceso por el cual los individuos pueden identificar, representar y mejorar su comunidad mediante el uso de una técnica fotográfica específica, confiando las cámaras a los individuos para que actúen como registradores y potenciales catalizadores del cambio en sus propias comunidades” (Wang y Burris, 1997, pág. 369).

Fortalezas del método

- Metodología flexible y fácilmente adaptable a diversos contextos, objetivos y públicos
- Material visual (fotos) como plataforma para la reflexión personal y como un canal para la auto-expresión
- Permite una atmósfera de intercambio y confianza
- Apela a la responsabilidad individual y colectiva de los/as participantes frente a los problemas identificados en su comunidad.
- Estimula la creatividad de los/as participantes
- Promueve la libertad de expresión: los comentarios positivos y menos positivos se expresan libremente en el proyecto

Desafíos del método

- Requiere una preparación importante: disponibilidad de recursos técnicos (impresión de fotografías grandes de calidad y tamaño adecuados), espacio físico adecuado, equipo
- Disponibilidad de recursos humanos capacitados con experiencia en el trabajo con jóvenes y poblaciones vulnerables
- Habilidades de los/as facilitadores/as de la actividad: comprensión del método, ingenio, flexibilidad, alta receptividad y capacidad de escucha
- Riesgo de influencia a veces involuntaria de algunos líderes de los/as participantes durante la actividad, en detrimento de otros/as participantes
- Presencia de un traductor capaz de mantener la fluidez de la comunicación y la confianza de todos los/as participantes en un contexto en el que la lengua materna de los/as participantes no es la de la persona facilitadora del ejercicio

Foto voz es un método cualitativo de recogida de datos que se utiliza a menudo en la investigación de la acción, en el que los/as participantes toman fotografías para explorar temas complejos o delicados. Creada a principios de los años 90, se basa en la idea de que las fotos pueden ayudar a los/as participantes, a menudo marginados socialmente, a contar una

historia, identificar los problemas a los que se enfrentan y reflexionar sobre la formulación de sus propias soluciones. De esta manera, la fotografía no sólo se utiliza para responder a preguntas de investigación descriptivas: también es un medio para iniciar un verdadero cambio de paradigma (Fradet 2012).

Photovoice is a process by which people can identify, represent, and enhance their community through a specific photographic technique. It entrusts cameras to the hands of people to enable them to act as recorders, and potential catalysts for change, in their own communities (Wang et Burris 1997, p. 369).

Foto voz es un proceso por el cual los individuos pueden identificar, representar y mejorar su comunidad a través de una técnica fotográfica específica, confiando las cámaras a las personas para que actúen como registradores y potenciales catalizadores del cambio en sus propias comunidades (traducción libre).

¿Cuál es el propósito de *foto voz*?

Foto voz es un método visual y participativo, tiene generalmente tres objetivos principales: 1) permitir a los individuos destacar los puntos fuertes y las preocupaciones de su comunidad; 2) promover el conocimiento y el diálogo crítico discutiendo temas importantes en grupo; 3) influir en las decisiones políticas destacando los temas sociales (Wang y Burris 1997; Wang 1999; Wang y Redwood-Jones 2001; Fradet 2012).

De hecho, a través de la fotografía, se revela el poder de la imagen para definir lo que es real y normal. La imagen permite interpretar la realidad que nos rodea. También actúa como vehículo de las emociones que despiertan las condiciones económicas y socioculturales que afectan a los individuos y a sus comunidades. El proceso de *foto voz* es muy movilizador: es un espacio que promueve la conciencia de las realidades sociales y donde se produce la creación de representaciones y significados simbólicos. Permite a otros ver el mundo a través de los ojos del/la fotógrafo/a para que coexistan múltiples perspectivas e interpretaciones. A través del proceso y de los sujetos sociales que participan activamente en él, *foto voz* tiene el potencial de actuar como catalizador del cambio social.

De *foto voz* al foto-lenguaje: ¿hay aspectos comunes?

Existen varias variantes de *foto voz*, una de las cuales se llama “foto-lenguaje”. Las foto-voces difieren del foto-lenguaje en sus fundamentos. Mientras que *foto voz* forma parte de un enfoque de concienciación, emancipación y cambio social, el foto-lenguaje se basa originalmente en un enfoque más terapéutico. Creado en 1965 por un grupo de psicólogos franceses que trabajaban con adolescentes, el foto-lenguaje fue concebido de manera intuitiva. El objetivo era utilizar fotografías para apoyar el discurso de los jóvenes que tenían dificultades para expresarse y hablar en grupo sobre sus diversas y a veces dolorosas experiencias personales. También es un enfoque específico de trabajo en grupo basado en una colección de fotos elegidas para ilustrar un tema particular. El foto-lenguaje se centra más en la conciencia individual, centrada en los problemas personales o incluso psicológicos, y en hablar delante de todos los/as participantes (Vacheret, 2011).

Otra diferencia importante existe entre las dos técnicas a nivel de sus respectivos procesos. En efecto, en el caso del foto-lenguaje, las fotos son tomadas por un/a fotógrafo/a y la selección previa de las fotos la hace el/la facilitador/ de la actividad, a diferencia del enfoque foto-vocalización en el que los/as participantes toman las fotos por sí mismos (White, Sasser, Bogren y Morgan 2009). Sin embargo, las diferencias se desvanecen cuando el enfoque del foto-lenguaje se alinea con los fundamentos teóricos y los conceptos de la concientización como instrumento liberador en un proceso de educación y transformación social (Freire, 1970). La única diferencia que queda es entonces a nivel del autor de las fotos (el fotógrafo vs. los/as participantes).

En ambos enfoques se reconoce el potencial influyente de la imagen/foto utilizada como “voz familiar, poder, vehículo de emociones, indicador y revelador simbólico de las realidades sociales, catalizador del cambio e instrumento de promoción y acción social” (Fradet, 2012).

Como proceso, *foto voz* y el foto-lenguaje juegan un papel en el despertar de la “conciencia crítica”, es decir, una profunda comprensión de cómo funciona el mundo y cómo la sociedad, la política y las relaciones de poder afectan a la realidad de un individuo o grupo. La técnica de la fotografía, al “desvelar” y exponer las realidades sociales, puede influir en el surgimiento del cambio social (Danchin y Sax 2009; White et al. 2009).

Foto voz, un método de investigación y evaluación participativa

Foto voz se presenta generalmente como un método de investigación de acción participativa, pero también puede utilizarse como un método de investigación cualitativa o como un instrumento de recogida de datos para la planificación y la evaluación (White et al., 2009). Basado en la comunidad, este método es flexible y puede servir para diferentes propósitos (movilización comunitaria, evaluación de necesidades, evaluación de procesos y resultados desde una perspectiva participativa) (Catalani y Minkler 2010; Wang y Burris 1997). Aplicado como parte de la planificación de la intervención y la evaluación de las necesidades, el proceso de votación fotográfica involucra a la comunidad a fin de comprender mejor sus realidades, en particular a través de la perspectiva de las personas que no están en posiciones de poder o autoridad. Es una forma de legitimación del conocimiento comunitario (Mayfield-Johnson, Rachal y Butler 2014; Strack, Lovelace, Jordan y Holmes 2010).

Cuando se utiliza como método de evaluación participativa, *el foto voz* proporciona acceso a información concreta y útil sobre las comunidades destinatarias (Goo-Kuratani y Lai 2011). Puede complementarse con varios métodos, como las encuestas de hogares, las técnicas de grupos focales y de grupos nominales, equilibrando al mismo tiempo los objetivos de la investigación, la acción y la evaluación. Por lo general, sobre la base de las discusiones entre los/as participantes y los miembros del equipo de investigación, se identifica una pregunta. Luego, con sus cámaras, los/as participantes salen a buscar fotografías que ayuden a contestarla (Wang y Burris 1997). En un estudio realizado por Wang y otros (1996), se utilizó *foto voz* con mujeres rurales chinas que participaban en un proyecto de salud comunitaria para evaluar de manera participativa la eficacia, los puntos fuertes y débiles del proyecto. Se discute la forma en que *foto voz* puede contribuir a la capacidad de una comunidad para reflexionar sobre su propia cultura. *Foto voz* también fue utilizada por Duffy (2010) en una evaluación formativa en la que el investigador interrogó a mujeres solteras sobre los beneficios de participar en un proyecto de investigación. *Foto voz* es tanto más prometedora cuando se aplica en colaboración con poblaciones vulnerables o marginadas como las mujeres, los jóvenes, las personas que padecen una determinada enfermedad, las personas sin hogar (Valiquette-Tessier y otros, 2015) o los trabajadores del sexo (Desyllas, 2014).

El proceso de *foto voz*

Por regla general, la implementación de *foto voz* se lleva a cabo en tres fases. Requiere la presencia de un facilitador que esté familiarizado con las realidades históricas, políticas y socioculturales del entorno de intervención. La primera fase consiste en la capacitación y el fomento de la capacidad de los/as participantes que se familiarizan con los instrumentos (cámaras, baterías, películas, proyectores, álbumes, etc.). La segunda fase comprende la recogida de datos, que corresponde a los debates en grupos de personas (previamente identificadas) sobre cuestiones de desarrollo de la comunidad. Por último, la última fase es el análisis cualitativo de los comentarios/discursos de los/as participantes, a través de los cuales relacionan las fotografías con sus realidades (Catalani y Minkler 2010; Wang y Burris 1997).

Ventajas y desafíos del *foto voz*

Palibroda y otros (2009) identificaron varios beneficios y desafíos relacionados con el uso de *foto voz*, incluso a nivel de los/as participantes, el proceso de investigación y facilitación y el nivel de la comunidad. Cabe recordar que la aplicación de *foto voz* promueve: i) el fortalecimiento de la participación de la comunidad en la acción y la promoción, ii) una mejor comprensión de las necesidades de la comunidad, y iii) un mayor empoderamiento individual y colectivo (Catalani y Minkler 2010).

El enfoque también presenta varios desafíos. En primer lugar, los líderes de opinión influyentes de las comunidades desfavorecidas pueden utilizar su condición para imponer sus puntos de vista y mantener el statu quo frente al desarrollo o las cuestiones sociales, en detrimento de la mayoría marginada. Los/as participantes también pueden tener dificultades para presentar una realidad compleja o abstracta a través de sus fotografías.

En segundo lugar, dado que el foco de *foto voz* está en el individuo, también puede hacer hincapié en los puntos de vista individuales en lugar de en el consenso de la comunidad. Cuando las relaciones de poder favorecen a algunos individuos en detrimento de otros, *foto voz* podría perpetuar las desigualdades sociales. La discusión sobre temas a veces delicados, puede inducir sentimientos negativos.

Por último, las comunidades/individuos que viven en una situación de vulnerabilidad, pueden tener otras prioridades, además de contar con poco tiempo. Los resultados de *foto voz* también pueden no estar a la altura de las expectativas de los miembros de la comunidad, especialmente en lo que respecta al cambio social.

En el siguiente cuadro se presentan las ventajas y limitaciones del proceso de *foto voz*.

Tabla 1 : Resumen de las ventajas y desafíos del método *foto voz* según Wang y Burris (1997)

Ventajas	Desafíos
<ul style="list-style-type: none"> • Enfoque descriptivo a través de una forma de comunicación basada en imágenes • Promueve la interacción entre los participantes a través de ideas, historias, etc. • Valora el conocimiento de las comunidades • Fomenta la participación de la comunidad • Método adaptado a las comunidades vulnerables y marginadas • Puede utilizarse como instrumento de identificación de necesidades y método de evaluación • Un enfoque de creación de capacidad, aprendizaje y potenciación para priorizar las necesidades y encontrar soluciones 	<ul style="list-style-type: none"> • Aprender a tomar fotos puede ser difícil • Complejidad del análisis de los datos recogidos durante los debates • Sólo puede abarcar un número limitado de cuestiones sociales • Discrepancia entre el método y la práctica: necesidad de adaptarse al contexto de la aplicación y la audiencia en tiempo real • El riesgo de mantener el status quo por parte de los líderes de opinión • Hace más hincapié en el juicio personal a expensas del consenso de la comunidad. • Puede perpetuar las desigualdades sociales al dar más voz a los grupos de presión

En la segunda parte de este capítulo, presentamos un experimento de *foto voz* utilizado como método de evaluación formativa que se acerca más al *foto-lenguaje* descrito anteriormente. De hecho, en el caso que se presenta aquí, las fotos fueron tomadas por los miembros del equipo del proyecto y no por los/as participantes debido a ciertos parámetros (método no previsto al principio del proyecto, pero ajuste ágil durante la ejecución que permite una mejor adaptación al contexto y a los/as participantes, limitaciones de tiempo y la necesidad de obtener datos casi en tiempo real para adoptar medidas correctivas antes del despliegue de las actividades en la fase 2 del proyecto). Esta experiencia es tanto más interesante cuanto que *foto voz* se utiliza como método de evaluación basado en el arte (Desyllas 2014). Proporciona una forma “alternativa” de hacer un juicio de valor sobre una

intervención utilizando herramientas y enfoques creativos. El objetivo del enfoque era evaluar el componente artístico de un proyecto de agua, higiene y saneamiento ejecutado en una comunidad indígena de México.

Foto-elección para evaluar el arte social: entre la complejidad y la creatividad

Antecedentes

En un contexto en el que 700 millones de personas todavía no tienen acceso al agua potable y el 80% de ellas viven en zonas rurales (OMS/ UNICEF 2015), la Fundación One Drop, con sede en Montreal, en asociación con una fundación privada y una organización no gubernamental mexicana, puso en marcha en 2015 un proyecto piloto destinado a proporcionar acceso sostenible al agua potable, la higiene y el saneamiento en dos municipios (Mixtla de Altamirano y Tehuipango) del Estado de Veracruz (México).

Ambos municipios están poblados por comunidades indígenas que comparten la misma lengua (el 98% de los habitantes hablan náhuatl). Se caracterizan por una elevada tasa de analfabetismo y la falta de agua potable e infraestructura de saneamiento. Durante un largo período del año, la infraestructura hidráulica no funciona, principalmente debido a los altísimos costos de mantenimiento. Para satisfacer sus necesidades de agua, la gente depende del agua de los ríos y de otras fuentes no potables. Algunas familias tienen depósitos en los que pueden almacenar agua de lluvia, pero la cantidad recogida sigue siendo insuficiente.

Además, aunque la mayoría de los hogares tienen letrinas (alrededor del 96% en Mixtla y el 86% en Tehuipango, según los datos del proyecto), estas infraestructuras son inadecuadas debido a la falta de higiene y seguridad.

En cuanto a la higiene personal, se observó la ausencia, la baja frecuencia y la práctica inapropiada de lavarse las manos y el cuerpo con jabón y de cepillarse los dientes.

Todas estas prácticas, combinadas con el hecho de que la mayoría de los hogares no tratan el agua antes de beberla, promueven la proliferación de gérmenes patógenos y enfermedades transmitidas por el agua dentro de las comunidades y las familias.

Tabla 2 : Contexto de las comunidades objetivo (INEGI 2010; One Drop 2015)

Municipios objetivo (Comunidades)	Tehuipango (Apitpitzactitla)	Mixtlade Altamirano (Mixtlantlapak)
Tasa de analfabetismo	Más del 50%	Más del 50%
Lengua	Náhuatl (98%)	Náhuatl (98%)
Población destinataria	23.746	10.387
Tasa de cobertura de saneamiento	86%	96%
Tasa de cobertura de agua	41%	55%
Higiene	Comportamientos higiénicos inadecuados dan lugar a una alta incidencia de enfermedades relacionadas con el agua	

La lógica de la intervención

En respuesta a estos problemas estructurales y de comportamiento, un proyecto ejecutado por la ONG mexicana financió la construcción de la infraestructura de agua y saneamiento. Sin embargo, la infraestructura por sí sola no basta para fomentar la adopción de buenas prácticas y hábitos a nivel individual y comunitario. Por consiguiente, la Fundación One Drop y sus asociados decidieron integrar un componente de arte social en el proyecto de infraestructura: un enfoque sistemático que utiliza todas las formas de arte para abordar las cuestiones sociales mediante la sensibilización y la movilización social. One Drop apuesta por el potencial del arte social como motor de transformación social y cambio de comportamiento en los proyectos de agua, higiene y saneamiento (WASH) que apoya (www.onedrop.org).



El rincón de la higiene en el aula. Crédito de la foto: One Drop, 2016

El arte social, que se lleva a cabo en colaboración con artistas locales, permite a las personas participar activa y creativamente en la solución de los problemas de su comunidad. La intervención a través de las artes conduce al desarrollo y la realización de acciones individuales y/o colectivas que responden a las necesidades de los

individuos y las comunidades. Los/as participantes que se involucran en estos proyectos artísticos asumen responsabilidades y adquieren nuevas habilidades al participar en experiencias enriquecedoras, al tiempo que desarrollan su autoestima y su autovaloración. Muchos agentes de desarrollo las han utilizado durante decenios, en particular en la esfera de la salud, especialmente en lo que respecta al VIH/SIDA (Onuekwe 2015; Moyer-Gusé 2008; Singhal y Rogers 1999). El arte social se encuentra en la encrucijada de varias estrategias de comunicación para el desarrollo, la educación popular y especialmente el enfoque de educación-entretenimiento.

En México, los objetivos del componente del proyecto Arte Social para el Cambio de Conducta fueron: 1) integrar el arte social como instrumento de sensibilización, movilización y cambio de conducta individual y colectiva en un proyecto existente de EHA y 2) evaluar los efectos del arte social en el conocimiento, las actitudes y las prácticas en torno a dos conductas objetivo: el lavado de manos y el tratamiento del agua. Concretamente, ese enfoque tenía por objeto, por una parte, diseñar e integrar actividades de arte social en dos comunidades destinatarias y, por otra, elaborar un marco integrado de vigilancia, evaluación y aprendizaje con métodos innovadores para evaluar los efectos del arte social en las variables de interés mencionadas anteriormente.

A continuación se presenta el modelo lógico de la intervención social en el arte, es decir, las actividades planificadas y los resultados previstos a corto, medio y largo plazo, teniendo en cuenta el contexto del proyecto y el entorno de las comunidades destinatarias.

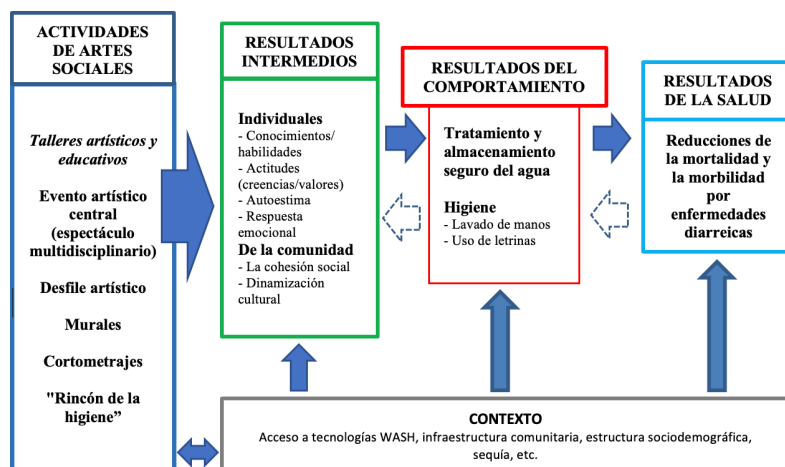


Figura 1 : Modelo Lógico del Proyecto (basado en Figueroa; Kincaid 2010)



Pinturas murales en los tanques de agua.
Crédito de la foto: One Drop, 2016

Como parte de esta innovadora iniciativa, se organizaron talleres de artes sociales que incluían teatro, danza (contemporánea y tradicional), música, circo, artes visuales y realización de cortometrajes para niños y adolescentes en las dos comunidades del proyecto. Los temas abordados por los artistas en esos talleres, así como los mensajes

transmitidos, se centraron en el lavado de manos y el tratamiento del agua. El proceso creativo y las formas de arte se inspiraron en gran medida en historias locales, cuentos y otros bienes socioculturales. Los talleres se basaron en procesos colectivos de cocreación para mejorar las actitudes, los conocimientos y las prácticas de los/as participantes en relación con el lavado de manos y el tratamiento del agua, al tiempo que se les enseñaban técnicas artísticas.

Según este proceso, se esperaba que, como resultado de su experiencia de despertar artístico, aprendizaje cognitivo y reflexión crítica, se

fortalecieran las capacidades/conocimientos de los jóvenes como agentes de cambio en sus comunidades, se les aumentara la autoestima y las actitudes hacia las cuestiones abordadas fueran positivas y proactivas a fin de fomentar un cambio de comportamiento sostenible a nivel individual y comunitario.



Creación del Rincón de la Higiene con los/as profesores/as. Crédito de la foto: One Drop, 2016

Además de los talleres artísticos mencionados, se ha puesto en marcha la actividad “Rincón de la higiene”. Consistía en la creación de un espacio dedicado a la higiene personal en las aulas. Este rincón se creó como parte de talleres de bricolaje, pintura y collage divertidos y educativos en los que participaron alumnos y profesores.

Esta actividad se llevó a cabo con el fin de fomentar la creatividad de los profesores en la elaboración de materiales educativos para concienciar a los niños de la escuela primaria sobre las cuestiones relacionadas con la higiene y adoptar prácticas apropiadas en un entorno divertido y atractivo.

En el marco del proyecto se organizaron otros eventos dirigidos a las comunidades en general: desfiles de baile, actuaciones multidisciplinares, difusión de cortometrajes producidos en los talleres y murales.

En el cuadro siguiente se presentan las actividades realizadas, los grupos destinatarios, los/as participantes y los temas abordados.

Tabla 3 : Actividades artísticas sociales en las zonas de destino

Actividades establecidas	Descripción y participantes	Grupos destinatarios/ públicos	Temas específicos
Talleres artísticos y pedagógicos	Teatro, danza contemporánea y tradicional, música, circo, artes visuales, cortometrajes, etc.	Jóvenes de 12 a 18 años	Aprendizaje técnico y cognitivo en varios campos artísticos integrando contenidos relacionados con el lavado de manos y las prácticas de tratamiento del agua a nivel doméstico
Eventos artísticos (Espectáculos)	Creación, producción y presentación de espectáculos multidisciplinares por jóvenes acompañados de artistas	Toda la comunidad	Prácticas de lavado de manos y tratamiento del agua en el hogar
Desfiles artísticos	Creado y organizado por los jóvenes acompañados por los/as artistas/ Participación activa de los jóvenes, artistas y miembros de la comunidad en el desfile final: desfile cultural basado en las experiencias tradicionales de las caravanas	Toda la comunidad	
Murales	Pinturas de los jóvenes apoyadas por los artistas/ Inauguración de los jóvenes durante el desfile final	Toda la comunidad	
Cortometrajes	Creación por parte de los jóvenes de los talleres formados por los artistas/ Presentación por parte de los jóvenes durante el desfile final	Toda la comunidad	
“Rincón de la higiene”	La participación de todo el alumnado en las clases correspondientes	Estudiantes de primaria (6-12 años)	Prácticas de lavado de manos y de higiene personal de los niños y niñas

El plan de evaluación de arte social



Emisión de cortometrajes. Crédito de la foto: One Drop, 2016

En el marco de este proyecto, se elaboraron y utilizaron varios instrumentos de supervisión y evaluación cuantitativos y cualitativos, en particular para medir los efectos de las intervenciones artísticas. El método de foto voz se adaptó y se aplicó como herramienta de evaluación con los jóvenes que participaban en los talleres de arte social para

obtener sus opiniones, percepciones y satisfacción con estos talleres.

Se elaboró un plan de evaluación en colaboración con el equipo del proyecto en las zonas seleccionadas. Se trata de un plan que aborda varias cuestiones, pero sólo se presentan aquí las que fueron objeto de foto voz.

Tabla 4. Plan de evaluación del componente de arte social del proyecto

Objetivo de la evaluación	Preguntas de evaluación (desarrolladas en colaboración con el equipo del proyecto y los artistas)	Procesos e instrumentos de recogida de datos
Evaluar las percepciones de los jóvenes sobre las actividades artísticas en las que han participado como parte del componente de Arte Social para el Cambio de Conducta relacionado con el agua potable y la higiene (tratamiento del agua y lavado de manos)	REACCIONES	Las fotos tomadas por el equipo del proyecto y los artistas según criterios bien establecidos e identificados por los jóvenes
	¿Cuáles fueron las reacciones de los participantes a las actividades de arte social?	
	APRENDIZAJE	Discusiones de grupo
	¿Cuáles fueron las lecciones aprendidas por los jóvenes a través de su participación en actividades de arte social?	Dibujos hechos por jóvenes para expresar sus reacciones y percepciones
	EFFECTOS SOBRE LAS PRÁCTICAS Y EL COMPORTAMIENTO	Entrevistas individuales con jóvenes líderes (no consideradas en este texto)
	¿Cuál ha sido la influencia del arte social en las prácticas de agua y saneamiento como el lavado de manos y el tratamiento adecuado del agua para reducir el riesgo de enfermedades?	Encuestas por hogares sobre conocimientos, actitudes y prácticas (CAP) relacionados con el agua, la higiene y el saneamiento (no se consideran en este texto)
	RESULTADOS	
	¿Cuáles son los efectos próximos del arte social en la comunidad?	

Aplicación de *foto voz* como un enfoque evaluativo en el contexto artístico y social

Objetivos

Se eligió *foto voz* como enfoque de evaluación formativa. El enfoque tenía la ventaja de adaptarse al contexto indígena en el que se ejecutaba el proyecto, al carácter artístico de las actividades y a los jóvenes participantes. La actividad de *foto voz* tenía dos objetivos:

- Compartir las representaciones de los/as participantes sobre cuestiones de agua y saneamiento, exponer ideas preconcebidas, identificar estereotipos e ir más allá de ellos dando voz a la experiencia personal de cada individuo.
- Permitir que los/as participantes se expresen libremente y liberen sus emociones, individualmente y en grupo, sobre el tema de interés, reflexionando críticamente sobre las soluciones a considerar.

Proceso

La fase de planificación

El proceso de votación fotográfica requiere preparativos que requieren tiempo y recursos humanos. Para garantizar la aplicación eficaz del proceso de evaluación sobre el terreno, se colocó en la comunidad a un coordinador de proyecto especializado en arte social y evaluación, apoyado por el equipo del proyecto (una persona encargada de la traducción, una persona encargada de tomar notas durante el ejercicio y un fotógrafo/camarógrafo). Estas personas fueron capacitadas y orientadas por el evaluador interno de One Drop con sede en Montreal, que hizo varios viajes de campo para apoyar la aplicación de la evaluación.

Los principales pasos de planificación necesarios para llevar a cabo la evaluación son los siguientes.

Selección de los/as participantes

Los/as participantes en el proceso de evaluación de *foto voz* fueron elegidos de acuerdo con los siguientes criterios:

- Participación activa de los jóvenes en los talleres artísticos descritos anteriormente: se priorizó a los asiduos para obtener una visión completa de los efectos del arte social
- La igualdad de elección de las niñas y los niños
- La voluntad de hablar frente a un grupo y el consentimiento informado para participar en la votación fotográfica
- Capacidad para entender el náhuatl y/o el español

Los grupos estaban formados por un mínimo de ocho y un máximo de 15 participantes, lo que es un número relativamente adecuado para este tipo de actividad de grupo. Una vez concluida la selección, el coordinador del proyecto compartió la lista de participantes con los dirigentes de cada comunidad y con los padres para obtener su consentimiento a fin de reunir a los/as participantes identificados por la selección. Se obtuvo el libre consentimiento de los padres y los hijos y se comunicó a los/as participantes información clave relacionada con la actividad artística (propósito, fecha, hora, lugar). El municipio proporcionó un aula para la actividad.

Selección de fotos

Antes de la actividad *foto voz*, los artistas y el equipo del proyecto seleccionaron las fotos que se utilizarían como base para la evaluación según los siguientes criterios:

- Fotos que representan escenas y participantes en la acción
- Fotos que representan cada uno de los talleres artísticos facilitados por los artistas en el marco del proyecto de arte social
- Fotos que representan los temas del proyecto: lavado de manos y tratamiento del agua
- Fotos que reflejan las emociones de los jóvenes que participan en el proyecto durante los talleres



*Desfile dentro de la comunidad.
Crédito de la foto: One Drop,
2016*

Se seleccionaron unas 20 fotos teniendo en cuenta el número de participantes (unas 10). Para cada una de las fotos seleccionadas se hicieron dos copias, especialmente en caso de que dos participantes quisieran seleccionar la misma foto. Cada foto fue identificada por un número en el reverso para facilitar el análisis. Finalmente, las fotos se imprimieron en un formato grande (8 ½ x 11) para permitir a los/as participantes visualizar mejor los detalles de cada foto.

Preparación de los recursos humanos

La cuestión de los recursos humanos es fundamental para esta actividad. Es útil contar con varias personas de recursos que puedan apoyar la actividad (traducción, toma de notas, etc.). Sin embargo, deben orientarse sobre la actitud a adoptar durante el proceso para que su presencia no cause molestias a los/as participantes.

La foto-vigilancia se llevó a cabo en dos comunidades a las que se dirigía el proyecto y cuatro especialistas asumieron cada uno las siguientes funciones:

- Animación
- Observación/observación durante el año
- Traducción
- Fotografía/giros

Estas personas tenían el mandato de apoyar la buena marcha del ejercicio. Por lo tanto, recibieron una breve sesión de capacitación previa que abarcó los siguientes puntos: 1) explicación del proceso de *foto voz*; 2) presentación del proceso y los pasos involucrados; 3) explicación de las cuestiones del ejercicio; 4) identificación y aclaración del papel de cada persona durante *foto voz*; 5) identificación de los papeles de cada persona durante el ejercicio.

Como parte de este proyecto, el coordinador de las actividades artísticas en el campo fue elegido para ser la facilitadora de *foto voz*. El objetivo era aprovechar la confianza y la complicidad existentes entre ella y los jóvenes, y fomentar su participación activa. En efecto, la proximidad del coordinador a los jóvenes creó un clima de confianza para que éstos se sintieran a gusto



*Espectáculo multidisciplinar con jóvenes.
Crédito de la foto: One Drop, 2016*

para expresarse libremente. El evaluador interno de One Drop estuvo presente durante el ejercicio como observador. Supervisó la actividad y observó el progreso y el comportamiento no verbal de los jóvenes. También orientó al facilitador en varias esferas, entre ellas la de velar por que todos los/as participantes hablaran, aclarar las preguntas y dar tiempo a los/as participantes más tímidos para que hablaran. Los artistas que facilitaron los talleres de arte no estuvieron presentes para permitir que los jóvenes se expresaran sobre su experiencia en estos talleres de manera abierta y con confianza.

Implementación

La actividad de foto voz en cada una de las dos comunidades tuvo lugar de la siguiente manera:

Organización del espacio de la sala (30 minutos)

La organización de la sala es muy importante para crear una atmósfera de compartir y escuchar. Las sillas se han colocado en el aula en un círculo para fomentar el intercambio, la interacción y la escucha activa. Además del área de discusión circular, se colocó una gran mesa en la parte de atrás de la sala y allí se exhibieron fotos seleccionadas. El aula fue una sabia elección porque los jóvenes estaban en un espacio familiar y tenían material escolar a su disposición.



Disposición de los jóvenes en un círculo para fomentar el diálogo, crédito de la foto: One Drop 2016

Presentación del trabajo al grupo (15 minutos)

Con el apoyo constante del traductor, la facilitadora se encargó de explicar y detallar con claridad y precisión los pasos del proceso a seguir, los métodos utilizados (incluida la importancia de escuchar y la posibilidad de expresarse libremente) y las limitaciones de tiempo. El objetivo era asegurar que cada participante comprendiera el proceso y la metodología y se sintiera seguro para participar activamente en la actividad. En esta introducción también se incluyó una presentación de las diferentes personas presentes en la sala y sus funciones, a fin de no crear incomodidad entre los jóvenes.

Elección individual de las fotografías (10 minutos)

La elección individual de las fotografías fue para permitir a los/as participantes expresarse en torno a las siguientes cuestiones de percepción:

- *Pregunta 1: Elige una foto que describa lo que más te gustó de los talleres.*
- *Pregunta 2: Elige una foto que describa lo que menos te gustó en los talleres.*

Para este paso, la facilitadora invitó a los jóvenes a pasar a la mesa donde se exhibían las fotografías para elegir individualmente la foto que mejor describiera lo que habían disfrutado en los talleres (Pregunta 1). Mientras se movían silenciosamente alrededor de la mesa, identificaron una foto relacionada con la primera pregunta. La facilitadora invitó entonces a los/as participantes a tomar la foto elegida y volver al espacio de debate.

Para los que habían elegido la misma foto, había copias disponibles para que todos pudieran tener su foto a mano para su discusión.

Después de completar el intercambio (descrito en el siguiente paso) en torno a la primera pregunta, se utilizó el mismo proceso para la pregunta número 2: ¿qué foto representa lo que menos le gustó en los talleres?



Los jóvenes escogen las fotos que representan la respuesta a la pregunta formulada, crédito de la foto: One Drop 2016

Trabajo en grupo sobre las preguntas 1 y 2 (duración: 1 hora)

Después de la elección de la foto relacionada con la primera pregunta, se formó el círculo de discusión para intercambiar sobre las reacciones y reflexiones de cada participante.

La facilitadora repitió la pregunta e invitó a cada joven a compartir sus pensamientos en relación con la pregunta presentando al grupo la foto

elegida, las razones de esta elección y la respuesta a la pregunta. Se alentó a los demás participantes a adoptar una actitud de escucha activa y a esperar su turno para hablar, a fin de no influir en las reflexiones de cada joven presente en el ejercicio.

Además, la facilitadora recordó a los/as participantes los objetivos del ejercicio para que cada uno de ellos se sintiera cómodo al expresarse.

Cada participante era libre de expresarse en español o náhuatl. Dependiendo del idioma elegido, las declaraciones se tradujeron en tiempo real. Cada participante pudo aprovechar su tiempo de conversación y todos expresaron el deseo de comunicar sus emociones con respecto a la elección de su foto. No había reticencia, ni había jóvenes que no quisieran expresarse.

Cuando procedía, la facilitadora hacía preguntas adicionales durante los intercambios a fin de promover la reflexión de cada joven.

La misma reflexión se hizo para la segunda pregunta.



Los jóvenes se expresan en las fotos elegidas, crédito de la foto: One Drop 2016

Trabajo individual sobre la pregunta 3 (duración: 30 minutos)

Esta fase apeló a la creatividad de los jóvenes y a su capacidad de “dibujar sus emociones”. En un proceso tradicional de foto voz o fotolenguaje, no se utiliza el dibujo. El equipo del proyecto decidió incluir el componente del dibujo en la evaluación de foto voz por dos razones: 1) era interesante incorporar otro método de recopilación de datos basado en el arte, y 2) dado que las artes visuales ocupan un lugar destacado en la cultura indígena, el dibujo promovería la expresión libre y diferente (no verbal) de los jóvenes para expresar sus emociones.

Durante esta fase, se pidió a los/as participantes que respondieran a la siguiente pregunta:

Pregunta 3: ¿Puedes hacer un dibujo que describa cómo te sientes desde esta experiencia?

El objetivo de esta pregunta era conseguir que los jóvenes expresaran sus emociones después de los talleres y otras actividades artísticas sociales que tenían lugar en su comunidad. En lugar de un intercambio de grupo como en las preguntas anteriores, los jóvenes se aislaron en el lugar de su elección, el patio de la escuela, otra habitación, alrededor de una mesa, para hacer su dibujo.

Por lo tanto, se proporcionó material de dibujo y se dio a cada participante aproximadamente 30 minutos para completar los dibujos. Los dibujos pueden ir acompañados de palabras u oraciones. Los jóvenes exploraron el tema individualmente, respetando sus espacios de trabajo y su individualidad.

Algunos retomaron las ideas que habían expresado durante las preguntas 1 y 2, mientras que otros expresaron nuevas ideas y produjeron imágenes fuertes y simbólicas.



Joven mirando hacia atrás en su experiencia, crédito de la foto: One Drop 2016

Conclusión de la actividad (10 minutos)



Joven mirando hacia atrás en su experiencia, crédito de la foto: One Drop 2016

El ejercicio de evaluación de *foto voz* concluyó con la presentación voluntaria de los dibujos realizados durante el ejercicio anterior. Se invitó a los/as participantes a compartir sus reflexiones basadas en su creación.

La facilitadora agradeció y felicitó a los jóvenes por su participación activa y el respeto a las instrucciones. Los jóvenes también tuvieron la oportunidad de compartir sus impresiones sobre la actividad y disfrutaron mucho al recordar su experiencia en los talleres de arte de una manera lúdica y artística.

En total, el proceso de votación fotográfica duró unas dos horas. Los/as participantes pudieron regresar a casa y los miembros del equipo hicieron un balance de la actividad y compartieron sus impresiones sobre cómo fue. También se decidió producir copias de las fotografías elegidas por los/as participantes para entregárselas. Para reforzar el sentimiento de orgullo en sus respectivas comunidades, las fotos de los jóvenes participantes se exhibieron y se hicieron visibles para todos.

Análisis de datos

Los datos reunidos durante un proceso de votación fotográfica son principalmente cualitativos y el proceso de análisis puede ser difícil. En este proyecto, el objetivo era recoger las percepciones y reacciones de los jóvenes ante sus experiencias durante los talleres artísticos y educativos. Las preguntas que se hicieron a los jóvenes fueron las siguientes:

Pregunta 1: Elige una foto que describa lo que más te gustó de los talleres.

Pregunta 2: Elige una foto que describa lo que menos te gustó en los talleres.

Pregunta 3: ¿Puedes hacer un dibujo que describa cómo te has sentido desde esta experiencia?

A continuación se presentan las estrategias de análisis utilizadas:

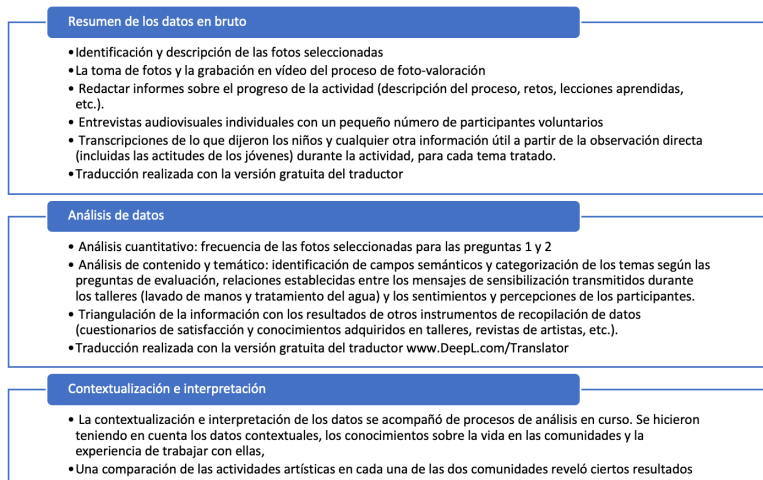


Figura 2 : Proceso de análisis de los datos recogidos

Resultados de la evaluación

Las reacciones de los/as participantes a las actividades de arte social del proyecto



Los jóvenes al final del ejercicio de foto voz, crédito de la foto: One Drop 2016

La mayoría de los jóvenes dijeron que disfrutaron **del ambiente de los talleres** (el entretenimiento, el trabajo en grupo y la complicidad con sus compañeros). Sin embargo, algunos también plantearon situaciones de conflicto experimentadas en algunos talleres y otros informaron de que disfrutaron mucho menos del taller de música, que les pareció menos

entretenido. Esto podría explicarse por la falta de experiencia de los facilitadores con los jóvenes aborígenes y su incapacidad para motivar a los niños.

La relación con los artistas (apego, complicidad, confianza) fue particularmente apreciada, sobre todo por su accesibilidad, su adaptabilidad al público y su capacidad para motivar, entretener, movilizar y sensibilizar a los jóvenes. Según Mackinnon y Ryan (2015), las actividades artísticas como foto voz son también una oportunidad para que los facilitadores demuestren su capacidad de liderazgo más allá del objetivo inicial.

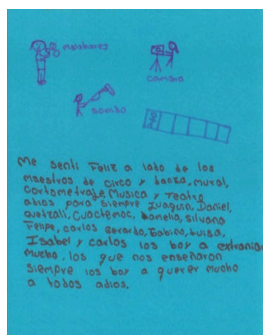
Al formar parte de un proceso creativo colectivo, el potencial de desarrollo individual de los jóvenes aumenta gracias **al reconocimiento del éxito colectivo**: “Lo que más me conmovió fue ver a mis amigos realizar acrobacias que no podían hacer al principio de la actividad” (participante joven).



Fotos de los jóvenes participantes en los talleres, crédito de la foto: One Drop 2016

Los jóvenes expresaron entonces **un sentimiento de abandono y tristeza** al final de las actividades (salida de los artistas en particular, pérdida de acceso al material artístico, ruptura del grupo creado durante los talleres). En general, la fotografía, el dibujo y el habla permitieron a los jóvenes expresar sus emociones más profundas y a veces mezcladas.

Aprender de los/as participantes en los talleres de arte social



*“Me sentí feliz al lado de los maestros de canto y danza, pintura mural, cortometrajes, música y teatro. Adiós para siempre Joaquín, Daniel, Quetzali, Cuactemoc, Pamela, Silvana, Felipe, Carlos Gerardo, Gabino, Luisa, Isabel y Carlos [nombres de pila de los artistas]. Los voy a extrañar mucho, lo que nos enseñaron. Siempre los voy a querer mucho. A todos adiós”.
Crédito de la foto: One Drop 2016*

Esta actividad confirmó que uno de los efectos de los talleres artísticos y educativos es **el fortalecimiento de las capacidades personales de los jóvenes y la liberación de su pleno potencial**. Los resultados de varios proyectos financiados por One Drop en África (Burkina Faso), la India (Odisha) y América Central (El Salvador y Honduras) en los que se involucra a los jóvenes en actividades de arte social describen procesos encaminados a convertir a los jóvenes en agentes de cambio en sus comunidades. En esos talleres, los jóvenes, especialmente las muchachas, demostraron un liderazgo individual y fueron ganando confianza poco a poco entre el comienzo y el final de las actividades, lo que fue confirmado por la experiencia mexicana.

El sentido de orgullo y autoeficacia aumentó, particularmente como resultado del aprendizaje de nuevas técnicas artísticas. Esta actitud positiva favorece la asimilación del aprendizaje de prácticas saludables como el lavado de manos y el tratamiento del agua para hacerla potable. Durante

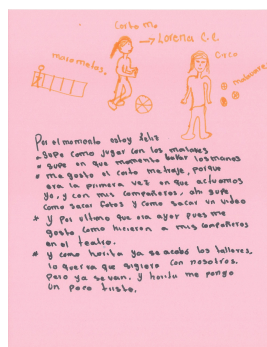
las actividades, los jóvenes pudieron practicar un lavado de manos adecuado en varias ocasiones. Sin embargo, la sostenibilidad de este aprendizaje puede verse comprometida si no se repite a diario. Para mitigar estos riesgos, se desarrollaron algunas actividades en estrecha colaboración con los maestros a fin de integrarlos en el programa de estudios de las escuelas y seguir aplicándolos en sus aulas. En la mayoría de los casos, los padres estaban presentes durante las actividades.

Percepciones de los/as participantes en el proceso de evaluación

Foto voz ha sido un éxito en ambas comunidades. Los/as participantes mostraron gran confianza y apertura al ejercicio a pesar de la presencia de oradores externos (camarógrafo, traductor). También fueron muy respetuosos con las instrucciones (silencio durante la elección, escucha, etc.), lo que creó una atmósfera de intercambio y diálogo. Este hallazgo está en consonancia con los resultados de varios estudios (Delgado 2015; Rania et al. 2014).

Dado que *foto voz* está orientada a los/as participantes, las preguntas preliminares que guían un estudio son maleables, lo que permite una mayor capacidad de respuesta a las necesidades identificadas de los/as participantes, así como la exploración de nuevas áreas de descubrimiento a medida que van surgiendo (Patton 2002, citado en Sutton-Brown 2014, pág. 171[4]).

Considerando la gran timidez de algunos participantes y teniendo en cuenta el desafío de la lengua materna (náhuatl) de estas comunidades, *foto voz* facilitó la comunicación de ideas a través del medio visual como soporte de la palabra. Los jóvenes fueron particularmente receptivos a las fotografías. Fueron un centro de curiosidad e interacción, así como el punto de partida de la reflexión cognitiva: las fotografías permitieron a los jóvenes recordar los momentos vividos durante los talleres e iniciar un análisis crítico de su experiencia (Strack et al. 2004). La foto seleccionada juega el papel de la “máscara” como en el teatro, permitiéndonos ser nosotros mismos detrás de una imagen que nos representa y a través de la cual se expresan los sentimientos. La foto es también un “espejo del alma” que permite a los jóvenes verse en acción para el cambio. Se consideró pertinente la elección de *foto voz* para mirar hacia atrás en las actividades de arte social, porque alentaba la reflexión personal del participante sobre los efectos que esas



“Por el momento estoy feliz: supe cómo jugar con las bolas de malabarismo. Supe en que momento lavar las manos. Me gustó el cortometraje, porque era la primera vez que actuamos yo y mis compañeros; ahora sé cómo tomar fotos y cómo hacer un video. Y por último, ayer me gustó como cómo mis amigos y yo hicimos teatro. Y ahorita como ya se acabaron los talleres, yo quería que siguieran con nosotros, y ahorita me siento un poco triste.”
 Crédito de la foto: One Drop 2016

actividades tenían en su individualidad y en su lugar dentro del grupo. Se promovió la conciencia y la responsabilidad personal por las cuestiones sociales (en este caso, las cuestiones de salud), en un contexto en el que la noción de pertenencia a la comunidad tiene prioridad sobre la individualidad. Foto voz reveló el empoderamiento de los/as participantes al proporcionarles el espacio para ser ellos mismos, para desarrollar y expresar sus opiniones personales en su interacción con el grupo en una atmósfera respetuosa de compartir y escuchar.

Foto voz resultó ser divertida y original. Todos los jóvenes estaban muy entusiasmados con la participación (Madrigal y otros 2014; Strack y otros 2004). A diferencia de las formas más tradicionales de evaluación (encuestas, grupos de discusión) a las que se enfrentaron durante la ejecución del proyecto, los/as participantes informaron de que se divertían, se entretenían y se emocionaban. Pudieron expresarse de manera creativa sin sentir que estaban evaluando las actividades en las que habían participado o que estaban siendo evaluadas. En su estudio, Madrigal y otros encontraron que la integración de las actividades artísticas en un proceso de votación fotográfica ayuda a crear una experiencia interesante y entretenida para los jóvenes, a la vez que cultiva su espíritu de equipo y su sentido de comunidad.

Los/as participantes tienen la confianza y la seguridad de expresar sus ideas, tanto si son aspectos positivos como los que deben mejorarse en los talleres. Los jóvenes también afirmaron su individualidad y demostraron un liderazgo colectivo durante el ejercicio, tal como lo identificó Fradet (2012) a través de la sigla VOZ: *Voicing Our Individual and Collective Experience*. En efecto, pudieron elegir personalmente, sin la influencia de sus compañeros de actividad, sus fotos y expresar opiniones consensuadas y a veces diferentes, incluso sobre una foto idéntica. A pesar de las barreras lingüísticas que a veces se presentan, los jóvenes se expresan libremente, sin vacilación ni inhibición.

El análisis reflexivo de *foto voz* como herramienta de evaluación

Foto voz fue usada por primera vez por One Drop en México. El proyecto de arte social se centró en el cambio de comportamiento y brindó la oportunidad de aplicar enfoques de evaluación innovadores y cualitativos

para recoger las percepciones de los/as participantes de las comunidades indígenas y comprender mejor los efectos del arte social, que son difíciles de cuantificar y medir con los instrumentos tradicionales.

El arte social es una intervención compleja, que depende en gran medida del contexto en el que se lleva a cabo. En este sentido, el uso de *foto voz* es apropiado, ya que se adapta a la cultura y al contexto en el que viven los/as participantes. Por ejemplo, no requiere la traducción de un cuestionario al idioma local y apela a la creatividad de los jóvenes al tiempo que fomenta su libre expresión. Esta primera experiencia nos ayudó a identificar los puntos fuertes y los aspectos a mejorar en la conducción del proceso; pudimos capitalizar este aprendizaje.

Cabe señalar también que la decisión de aplicar *foto voz* de manera ágil y adaptada permitió obtener información útil y aprovechable que contribuyó a movilizar y sensibilizar a la comunidad (que rara vez tiene la oportunidad de acoger actividades divertidas para los jóvenes), mejorar las actividades posteriores en otros proyectos, compartir las vías de mejora con los asociados, en particular los artistas, e integrar *foto voz* como una actividad plena en el modelo de intervención de la organización *Social Art for Behavioural Change*.

Lecciones aprendidas

Seleccionar e imprimir fotos: un paso importante para planificar

Durante el primer ejercicio de foto-voto en la primera comunidad, no se habían respetado las dimensiones requeridas (fotos suficientemente grandes para facilitar la selección por parte de los/as participantes). Por lo tanto, era necesario ser ingenioso y reimprimir las fotos tomadas a lo largo del proyecto con un tamaño lo suficientemente grande para una mayor visibilidad. Este contratiempo reveló la importancia de planear un paso tan inocuo como la impresión de las fotos, especialmente en duplicado y con mucha anticipación. Las dimensiones de las fotos son esenciales ya que la foto es el elemento principal de la actividad. Cuanto más grandes son las fotos, más permiten a los/as participantes comprender la escena y recordar los momentos descritos o identificarse con las fotos. La diversidad de las

fotos y su número (mínimo de 6 fotos, máximo de 10) también son importantes, a fin de ofrecer una amplia elección a los/as participantes y reflejar el mayor número posible de opiniones.

El idioma: un factor esencial para la participación y la implicación

El proyecto se ejecutó en comunidades que hablan náhuatl, una lengua indígena de México. En este contexto, las intervenciones sobre el terreno, incluida la recogida de datos, debían tener en cuenta la complejidad de la comunicación. Por consiguiente, todas las actividades que requerían la recopilación de percepciones, comentarios o sugerencias de las poblaciones se realizaban en náhuatl. Esto permitió una mayor fluidez y dio confianza a los/as participantes. Se observó en varias ocasiones que las personas que participaban en el proyecto encontraban mucho más fácil expresarse en su lengua materna que en español.

Así, durante el ejercicio de *foto voz*, se dio a los/as participantes la opción de expresarse en náhuatl o en español. La gran mayoría de los/as participantes decidieron expresarse en náhuatl. Dado que algunos de los facilitadores del ejercicio no hablaban con fluidez el náhuatl, uno de los/as participantes tradujo voluntariamente durante la intervención. Como resultado, este participante no pudo participar plenamente en la actividad. Para evitar esta situación, se contrató a un traductor independiente para que proporcionara traducción simultánea durante el ejercicio de *foto voz* en la segunda comunidad. Wang y Burris (1997) sugieren que *foto voz* se adapte a las necesidades de los/as participantes de manera flexible. Por consiguiente, la cuestión del idioma de comunicación elegido para llevar a cabo la actividad fue esencial en *foto voz* y otras actividades realizadas en esta región de México. Para que los/as participantes comprendan el ejercicio y participen plenamente, pudiendo expresarse con facilidad, es esencial crear un entorno que propicie su participación. Esto significa tener en cuenta el idioma en que los/as participantes se sienten más cómodos para expresarse y planificar un sistema de traducción simultánea si es necesario para facilitar la interacción. El hecho de poder hablar el propio idioma aumenta la calidad de los datos recopilados. Las emociones se expresan más fácilmente en la lengua materna, y por lo tanto la información recogida es más fiable. Sin duda, la traducción conlleva el riesgo de introducir un sesgo en la interpretación de lo que dicen los/as participantes. Una alternativa sería capacitar a los

agentes de la comunidad en el proceso e incluir este componente de creación de capacidad en la planificación del proyecto. De esta manera, la validez del enfoque, su anclaje en la comunidad y su sostenibilidad se verán fortalecidos.

El carácter innovador del método y su contexto de aplicación: elementos esenciales que deben tenerse en cuenta

En el contexto del proyecto, las comunidades estaban acostumbradas a participar en las actividades tradicionales de vigilancia y evaluación (encuestas demográficas) realizadas por los trabajadores, y nunca se habían enfrentado a un método innovador como *foto voz*. *Foto voz* hace hincapié en la creatividad, la reflexión basada en la memoria, e implica que los/as participantes expresen sus emociones y opiniones sobre las actividades descritas en las fotos, utilizando un ejemplo concreto.

No obstante, este método fue relativamente inquietante para algunos participantes al principio. Por lo tanto, los facilitadores prestaron especial atención a la explicación de los objetivos y el curso de la actividad. También era esencial escuchar activamente y ser capaz de estimular el pensamiento de los/as participantes para que pudieran llegar al final de sus pensamientos y no sólo describir la foto elegida.

Además, como los/as participantes eran jóvenes y no tuvieron la oportunidad de confrontarse con el material visual de la fotografía, al principio de la actividad les llevó tiempo familiarizarse con las fotos, reaccionar, intercambiar y a veces reírse, antes de que pudieran concentrarse de nuevo y responder a las instrucciones del ejercicio. Esta es una de las razones por las que, en la segunda comunidad, se organizó un juego de rompehielos con los jóvenes basado en las fotos para que se sintieran cómodos y se prepararan para la actividad antes de pedirles que se concentraran plenamente. Por consiguiente, al planificar la actividad es esencial tener en cuenta el contexto en el que se realiza la actividad y el público, que en este caso no está familiarizado con los métodos de evaluación innovadores y los materiales fotográficos. Requiere una buena capacidad de adaptación y ajuste en tiempo real por parte de la facilitadora. Un buen conocimiento de la realidad cultural y social en la que se desarrolla la actividad permite una mejor organización, planificación y participación en la actividad.

Análisis de la actividad: un proceso largo y complejo

El análisis de los datos de la sesión fotográfica (toma de notas, grabaciones de actividades, comentarios de los/as participantes, fotos seleccionadas) fue un gran reto, dada la duración del proceso y la importancia de obtener rápidamente datos válidos y utilizables.

Para poder analizar lo que dijeron los/as participantes, se favoreció la toma de notas en la primera comunidad. Sin embargo, esta tarea resultó ser compleja, en particular debido a la barrera del idioma (la persona que tomaba las notas no entendía el náhuatl) y a la velocidad de los intercambios. Por consiguiente, en el caso de la segunda comunidad, se decidió grabar y filmar la actividad con el acuerdo previo de los/as participantes y sus padres, a fin de poder observar las imágenes a posteriori e informar mejor sobre la marcha y los resultados de la actividad.

Posteriormente, y como se ha descrito anteriormente, se llevaron a cabo varias etapas de análisis con el fin de cumplir los objetivos fijados (recoger las percepciones de los/as participantes sobre los talleres de arte social y comprender mejor los efectos del arte social en la adquisición de conocimientos y aptitudes artísticas). El proceso requería tiempo, recursos humanos especializados (en investigación y evaluación participativas) y disponibilidad para llevar a cabo un análisis integrado y participativo que tuviera en cuenta la complejidad del contexto de la intervención y un buen conocimiento del proyecto.

Conclusión

Foto voz ayudó a cumplir el objetivo de la evaluación de comprender mejor las percepciones de los/as participantes sobre los talleres artísticos y educativos fundamentales para los proyectos de arte social de One Drop. Este proceso de evaluación, participativo, divertido y que fomenta el diálogo, fue respetuoso con los/as participantes y se adaptó al contexto de la aplicación, asegurando al mismo tiempo la continuidad de las actividades artísticas.

Puede ser difícil para los investigadores observar algunos de los comportamientos más importantes de una comunidad,

especialmente cuando esa comunidad desconfía de los forasteros, está aislada, desatendida, oprimida o funciona sin un lenguaje escrito. En estas situaciones, las investigaciones de observación realizadas por personas ajenas a la comunidad pueden dar lugar a una visión sesgada e inexacta de la misma, lo que a su vez puede dar lugar a un malentendido de sus necesidades, activos y valores culturales... [Foto voz] es muy útil cuando se utiliza al principio de un proyecto, así como durante la etapa de evaluación y análisis para medir el éxito y la validez. (Haldenby, T. fecha no indicada [5]).

Aunque los/as participantes no eran los/as autores/as de las fotografías, eran los actores de las mismas y, como tales, pudieron descubrirse en su entorno, en la acción y en colaboración con sus compañeros. La imagen provocó reacciones espontáneas antes de fomentar la reflexión y el debate. Se dio voz a los jóvenes participantes a través de la fotografía y el dibujo, canales privilegiados para el intercambio de experiencias y la expresión de percepciones, emociones y opiniones que a menudo no son tenidas en cuenta por los adultos. La aplicación de este enfoque presenta retos que no ponen en duda su potencial de aprendizaje y transformación individual y colectiva.

Como resultado de esta experiencia, One Drop optó por utilizar *foto voz* como una intervención además de aplicarla como un enfoque evaluativo, ya que tiene las ventajas de las intervenciones basadas en las artes (Jagiello, 2015; Coholic et al. 2012; Djurichkovic 2011). Así, se une a todas las actividades artísticas sociales (teatro de intervención, talleres artístico-pedagógicos, documentación y medios de comunicación) utilizadas en sus proyectos de agua, higiene y saneamiento.

Por último, esta actividad también generó productos visuales (fotos, vídeos, dibujos) y proporcionó información más general sobre la dinámica social en las comunidades destinatarias. Los resultados se difundieron entre los asociados del proyecto, la comunidad, pero también entre los agentes del sector del agua, la higiene y el saneamiento.

Referencias clave

Wang, C. et Burris, M. A. (1997). Photovoice : Concept, methodology, and use for participatory needs assessment. *Health education & behavior*, 24(3), 369-387.

En este artículo, Wang y Burris, considerados pioneros en el campo de foto voz, ofrecen una visión general de la metodología. Presentan sus orígenes, ventajas y limitaciones. A continuación, se analizan las consecuencias prácticas, los facilitadores y las limitaciones a la luz de un proyecto de evaluación participativa de las necesidades ejecutado en una comunidad china.

Fradet, L. (2012). *Guide de la méthode photovoix*.

<https://sites.google.com/site/rappratique/outils-et-documentation/guide-methode-photovoix>

En esta guía práctica, Lucie Fradet presenta paso a paso el método foto voz en el contexto de la investigación de acción participativa. Desde la preparación hasta la evaluación, pasando por la recopilación y el análisis de datos, foto voz se traduce en acción. La guía populariza los conceptos teóricos y metodológicos asociados a la fotovoice, al tiempo que ofrece “consejos y trucos” a todo aquel que desee utilizarla para que sea un éxito.

Mackinnon, S. et Ryan, L. (2015). Photovoice: Through the lens – A case study. *Youth Engagement in Health Promotion*. Vol1. No.1

En este estudio de caso de Ontario, Mackinnon y Ryan demuestran la pertinencia de utilizar foto voz como herramienta para promover el debate con los jóvenes. A través de la combinación de la fotografía y la narración de historias, sus percepciones se tienen en cuenta para definir lo que es una comunidad saludable. Foto voz se presenta también como una estrategia para promover el diálogo con los responsables de la toma de decisiones en la comunidad. El autor concluye que photo voice es un método de investigación de acción participativa particularmente eficaz en la promoción de la salud, en particular mediante el concepto de potenciación de la comunidad.

Referencias

- Catalani, C. et Minkler, M. (2010). Photovoice: A review of the literature in health and public health. *Health education & behavior*, 37(3), 424-451.
- Coholic, D., Eys, M. et Loughheed, S. (2012). Investigating the effectiveness of an artsBased and mindfulness-based group program for the improvement of resilience in children in need. *Journal of Children and Family Studies*, 21, 833-844.
- Danchin, É. et Sax, C. (2009). *Photolangage sur les sentiments d'insécurité et de sécurité*. Bruxelles. BE: Cultures & Santé asbl.
- Delgado, M. (2015). *Urban youth and photovoice: Visual ethnography in action*. Oxford : Oxford University Press.
- Desyllas, M. C. (2014). Using photovoice with sex workers: The power of art, agency and resistance. *Qualitative Social Work*, 13, 477-501.
- Djurichkovic, A. (2011). *Arts in prisons: A literature review of the philosophies and impacts of visual arts programs for correctional populations*.
https://opus.lib.uts.edu.au/bitstream/10453/19836/3/ArtinPrisons_Djurichkovic.pdf
- Duffy, L. (2010). Hidden heroines: Lone mothers assessing community health using Photovoice. *Health Promotion Practice*, 11(6), 788-797.
- Figueroa, M.E. et Kincaid, D.L. (2010). *Social, Cultural and behavioral correlates of household water treatment and storage*.
<http://ccp.jhu.edu/wp-content/uploads/Household-Water-Treatment-and-Storage-2010.pdf>
- Fradet, L. (2012). *Guide de la méthode photovoix*.
<https://sites.google.com/site/rappratique/outils-et-documentation/guide-methode-photovoix>
- Freire, P. (1970). *Pedagogy of the oppressed* (traduit par M. Ramos). New York : Herder & Herder.
- Jagiello, J. (2015). *A case study: art-based interventions into management education*.
<http://pin-c.sdu.dk/assets/art-based-interventions-into-management-education-jolanta-jagiello-16-23-pinc-2015.pdf>

- Goo-Kuratani, D. L. et Lai, E. (2011). *Photovoice literature review*.
[http://teamlab.usc.edu/PhotovoiceliteratureReview\(FINAL\).pdf](http://teamlab.usc.edu/PhotovoiceliteratureReview(FINAL).pdf)
- Mackinnon, S. et Ryan, L. (2015). Photovoice: Through the lens – A case study. *Youth Engagement in Health Promotion*, 1(1), p-p.
- Madrigal, D., Salvatore, A., Casillas, G., Casillas, C., Vera, I., Eskenazi, B. et Minkler, M. (2014). Health in my community: Conducting and evaluating photovoice as a tool to promote environmental health and leadership among Latino/a youth. *Progress in community health partnerships: research, education, and action*, 8(3), 317–329.
<http://doi.org/10.1353/cpr.2014.0034>
- Mayfield-Johnson, S., Rachal, J. R. et Butler, J. (2014). When we learn better, we do better: Describing changes in empowerment through photovoice among community health advisors in a breast and cervical cancer health promotion program in Mississippi and Alabama. *Adult Education Quarterly*, 64(2), 91-109.
- Moyer-Gusé, E. (2008). Toward a theory of entertainment persuasion: Explaining the persuasive effects of entertainment-education messages. *Communication Theory*, 18(3), 407-425.
- Onuekwé, C. E. (2015). *Entertainment-education for Health Behaviour Change*. Lieu d'édition : Friesen Press.
- Palibroda, B., Krieg, B., Murdock, L. et Havelock, J. (2009). *A practical guide to photovoice: sharing pictures, telling stories and changing communities*.
http://www.pwhce.ca/photovoice/pdf/Photovoice_Manual.pdf
- Rania, N., Migliorini, L., Reborá, S. et Cardinali, P. (2014). Enhancing critical dialogue about intercultural integration: The Photovoice technique. *International Journal of Intercultural Relations*, 41, 17-31.
- Singhal, A. et Rogers, E. M. (1999). *Entertainment-education: A communication strategy for social change*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Strack, R. W., Lovelace, K. A., Jordan, T. D. et Holmes, A. P. (2010). Framing photovoice using a social-ecological logic model as a guide. *Health promotion practice*, 11(5), 629–636.
- Strack, R. W., Magill, C. et McDonagh, K. (2004). Engaging youth through photovoice. *Health Promotion Practice*, 5(1), 49-58.

- Sutton-Brown, C. A. (2014). Photovoice : A methodological guide. *Photography and Culture*, 7(2), 169-186
- Vacheret, C. (2011). Le photolangage, une médiation thérapeutique : un bref historique des théories groupales. *Le Carnet PSY*, 141(1), 39-42.
- Valiquette-Tessier, S. C., Vandette, M. P. et Gosselin, J. (2015). In her own eyes: Photovoice as an innovative methodology to reach disadvantaged single mothers. *Canadian Journal of Community Mental Health*, 34(1), 1-16.
- Wang, C. C. (2006). Youth participation in photovoice as a strategy for community change. *Journal of Community Practice*, 14(1-2), 147-161.
- Wang, C. C. et Redwood-Jones, Y. A. (2001). Photovoice ethics : Perspectives from Flint photovoice. *Health education & behavior*, 28(5), 560-572.
- Wang, C. et Burris, M. A. (1997). Photovoice : Concept, methodology, and use for participatory needs assessment. *Health education & behavior*, 24(3), 369-387.
- Wang, C., Yuan, Y. L. et Feng, M. L. (1996). Photovoice as a tool for participatory evaluation : the community's view of process and impact. *Journal of contemporary health*, 4, 47-49.
- White, R., Sasser, D., Bogren, R. et Morgan, J. (2009). Photos can inspire a thousand words: Photolanguage as a qualitative evaluation method. *Journal of Extension*, 47(3), 1-5.
- World Health Organization et UNICEF. (2015). *Progress on sanitation and drinking water: 2015 update and MDG assessment*.
http://files.unicef.org/publications/files/Progress_on_Sanitation_and_Drinking_Water_2015_Update_.pdf

Resumen / Résumé / Abstract

Para la evaluación de programas, es esencial usar métodos evaluativos que sean adaptados al contexto de la intervención, al objetivo de la evaluación y a las poblaciones meta. Sin embargo, esto representa un desafío para intervenciones complejas y para comunidades marginadas. En este capítulo, presentamos un método de evaluación participativa que utiliza la foto como apoyo para la expresión, llamado fotovoz, el cual está particularmente adaptado a los jóvenes y a las comunidades indígenas de un

proyecto de acceso a agua, saneamiento e higiene en México. Se escogió este método también por sus características creativas y artísticas, alineado con el objetivo de la evaluación : evaluar las percepciones de los jóvenes en cuanto a las actividades artísticas en las cuales participaron en el contexto del componente de Arte Social para el Cambio de Comportamiento del proyecto, el cual promovía buenas prácticas relativas al agua potable y a la higiene. Se identifican las fortalezas y debilidades de este método y se describe el proceso metodológico en detalle. Se presentan brevemente los resultados de la intervención puestos en evidencia por este proceso y finalmente se describen las lecciones aprendidas de la aplicación de la fotovoz, con el fin de promover la utilización de este método que resultó muy útil y exitoso en el caso presentado.

Utiliser des méthodes évaluatives adaptées à la fois au contexte d'intervention, à l'objet de l'évaluation et aux populations ciblées est essentielle en évaluation de programmes. Toutefois, cela comporte des défis non négligeables dans le cas d'interventions complexes et de communautés marginalisées. Dans ce chapitre, nous présentons une méthode d'évaluation participative utilisant la photo comme support d'expression -la photovoix-, particulièrement adaptée aux jeunes et aux communautés autochtones ciblées par un projet d'accès à l'eau, l'hygiène et l'assainissement au Mexique. Cette méthode qualitative a aussi été choisie pour son caractère créatif et artistique permettant de répondre à l'objectif de l'évaluation : évaluer les perceptions des jeunes par rapport aux activités artistiques auxquelles ils ont participé dans le cadre de la composante Art Social pour le Changement de Comportement ; les encourageant à adopter de meilleures pratiques en lien avec l'eau potable et l'hygiène. Les forces et les faiblesses de cette méthode sont exposées et le processus méthodologique décrit en détail. Les résultats de l'intervention mis en évidence par ce processus sont brièvement présentés et enfin, les leçons tirées de l'application de la photovoix sont décrites afin d'encourager l'utilisation de cette méthode qui s'est avérée particulièrement utile et concluante dans le cas présenté.

In program evaluation, it is essential to use evaluation methods which are adapted to the intervention context, the evaluation objective and the target populations. However, this is quite challenging in the case of complex

interventions and of marginalized communities. In this chapter, we are presenting a participatory evaluation method, called photovoice, using photos as a support for expression. This choice was particularly adapted to the youth and the indigenous communities targeted by a project of access to water, sanitation and hygiene in rural Mexico. This qualitative method was also selected for its creative and artistic characteristics, well suited to the evaluation objective : to assess the youth's perceptions regarding the artistic activities they had participated in, as part of the Social Art for Behavior Change component of the project, to promote the adoption of adequate practices related to safe water and hygiene. The method's strengths and weaknesses are outlined and the methodological process is described in detail. Findings from the evaluation are briefly presented, as well as lessons learned from the use of photovoice, in an effort to promote the application of this method which proved to be very useful and successful in the case studied.

Lynda Rey se especializa en el uso de enfoques innovadores para la evaluación de programas y la planificación estratégica. Lleva a cabo una amplia gama de mandatos en materia de salud pública, desarrollo internacional y gestión de ONG para organizaciones públicas, privadas y comunitarias, fundaciones e instituciones académicas de todo el mundo. Recientemente, dirigió la unidad de supervisión, evaluación y gestión de conocimientos de One Drop, una fundación con sede en Montreal que invierte en proyectos de agua, higiene y saneamiento que integran el arte social para el cambio de comportamiento. Lynda Rey tiene un doctorado en salud pública y un diploma en análisis y evaluación de intervenciones sanitarias de la Universidad de Montreal. También obtuvo dos maestrías en Francia; una en Ciencias Políticas y Relaciones Internacionales de Sciences-Po Aix y la otra en Cooperación Internacional y Políticas de Desarrollo de la Sorbona.

Wilfried Affodégon es un ingeniero agro-económico. También tiene una maestría en administración. Actualmente es candidato al doctorado en Ciencias Políticas en la Universidad Laval, Quebec-Canadá. Ha trabajado durante aproximadamente una década en programas de asistencia al desarrollo internacional como especialista en supervisión y evaluación. En esta capacidad, ha utilizado el vídeo participativo, que es una variante del enfoque de fotovoz para la evaluación del apoyo de la cooperación danesa a

las comunidades agrícolas de Benin. También es especialista en instituciones parlamentarias, gestión pública, medición del rendimiento y evaluación de la acción pública y métodos de investigación de ciencias políticas.

Isabelle Viens es licenciada en arte dramático por la Universidad de Québec en Montreal. Durante más de 25 años ha colaborado en proyectos de desarrollo internacional que la han llevado a vivir experiencias ricas y humanas trabajando con el arte como fuente de transformación social. Durante los últimos 10 años, ha hecho de América Central su hogar. Actualmente, es la directora de intervenciones de arte social para el cambio de comportamiento en América Latina para la Fundación One Drop.

Hind Fathallah es especialista en la evaluación de proyectos de desarrollo internacional. Actualmente es consultora en materia de supervisión, evaluación y aprendizaje de la Fundación One Drop de Montreal, donde combina sus conocimientos especializados en materia de desarrollo internacional con su experiencia en la evaluación de proyectos de arte social y cambio de comportamiento relacionados con el agua, la higiene y el saneamiento en el África occidental, América Latina y la India. Le apasiona el desarrollo internacional y los métodos innovadores de evaluación participativa. Tiene una Maestría en Cooperación Internacional de Sciences Po Lyon (Francia) y ha trabajado con varias organizaciones sin fines de lucro en Francia y Canadá.

María José Arauz es especialista en evaluación de programas. Tiene una maestría en salud pública de la Universidad de Montreal y tiene más de 7 años de experiencia en evaluación de programas e investigación evaluativa. María ha trabajado en proyectos de desarrollo internacional, cambio de comportamiento, salud pública, salud mental, determinantes sociales de la salud y desarrollo social y con diversos grupos como las poblaciones rurales, indígenas e inmigrantes. María promueve enfoques de evaluación participativos, creativos y orientados a la utilización.

Cita

Lynda Rey, Wilfried Affodégon, Isabelle Viens, Hind Fathallah y Maria-Jose Arauz (2020). El método foto voz. Una intervención con poblaciones marginadas para el acceso al agua potable, la higiene y el saneamiento en México. En *Evaluación de las intervenciones sanitarias en salud global*.

Métodos avanzados. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 95-134. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

5. El análisis de un proyecto de investigación-acción

Combinación de enfoques en el sector de la salud en Burkina Faso

AKA BONY ROGER SYLVESTRE, VALÉRY RIDDE Y LUDOVIC QUEUILLE

El análisis de un proyecto de investigación-acción

Definición del método

El objetivo es analizar el proceso de aplicación de la investigación-acción combinando un análisis externo realizado por una persona ajena a la investigación-acción (IA) con un análisis interno (análisis reflexivo) realizado por los agentes que participan en la investigación-acción.

Fortalezas del método

- El análisis externo es factual y se basa en criterios objetivos
- El análisis interno (análisis reflexivo) permite a los actores de la investigación-acción llevar a cabo reflexiones para profundizar el análisis externo y aportar elementos adicionales
- La combinación de los dos tipos de análisis permite mejorar la calidad del análisis utilizando los dos análisis complementarios y utilizar la triangulación de métodos para recoger información adicional

Desafíos del método

- Poder diseñar una red de análisis externo capaz de comprender la realidad, que sirva de guía para el análisis externo
- Ser capaz de hacer que las partes interesadas se apropien del proceso de análisis y de distinguir entre "objeto de investigación" y "proceso de aplicación de la investigación-acción", utilizando un vocabulario apropiado para producir conocimientos útiles
- Poder reunir a todas las partes interesadas para realizar un análisis reflexivo, sin las limitaciones de tiempo y recursos financieros

Desde 2008, dos distritos sanitarios de la región del Sahel de Burkina Faso (Sebba y Dori) se han beneficiado de un proyecto experimental para eximir a los niños y niñas menores de cinco años y a las mujeres embarazadas y lactantes del pago de la atención, con el fin de mejorar el acceso a la atención de esas poblaciones vulnerables. Sin embargo, aunque la atención es gratuita para estos pacientes, se les debe reembolsar íntegramente los gastos incurridos por los centros de salud, con el propósito de no dejar el

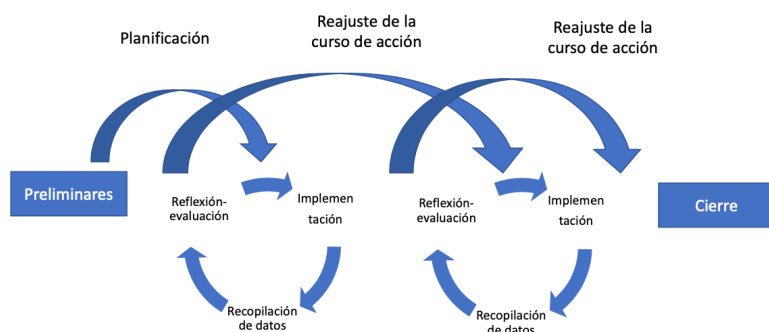
sistema de salud fuera de servicio. Sin embargo, el reembolso real mediante el cálculo de todos los costos de los servicios prestados por los centros de salud tiene muchas desventajas: representa una carga de trabajo para el personal sanitario y un alto costo de producción de los instrumentos para la ONG que los subvenciona.

Para abordar estas cuestiones, un equipo de investigadores/as ha estado realizando investigación-acción en el Distrito Sanitario de Sebba desde 2011, en colaboración con las partes interesadas en este proyecto. Su objetivo fue evaluar una innovación, a saber, el “reembolso a precio fijo” de los servicios de los centros de salud. Esto consiste en concederles una suma global por sus servicios mediante de reembolso en función del número de pacientes recibidos.

En este capítulo analizamos el enfoque utilizado para llevar a cabo esta investigación-acción con el fin de extraer lecciones que puedan servir de guía para futuros proyectos similares. El objetivo es proporcionar a los lectores un marco analítico original para mostrar cómo se puede evaluar y analizar un proceso de investigación-acción sanitaria mundial.

La investigación-acción es un proceso iterativo que implica la colaboración entre investigadores/as y profesionales para actuar juntos en un ciclo de actividades que implican la identificación de problemas, la planificación, la aplicación, la evaluación de soluciones y la reflexión (Avison 1999).

Roy y Prévost (2013) representaron las diferentes etapas de los ciclos experimentales de investigación-acción añadiendo una etapa preliminar y una etapa final al modelo original de Lewin (1940). Los ciclos experimentales llegan a su fin si las partes interesadas han resuelto todos los problemas identificados.



Los ciclos de investigación-acción. Fuente: Roy 2013

El enfoque utilizado para la aplicación de esta investigación-acción reunió un conjunto de información. El modelo de Roy (2013) nos permitió desglosarlo en varias etapas y ciclos para comprenderlo bien antes de centrarnos en el enfoque utilizado para implementar las diversas etapas. La noción de “enfoque” es definida por él (en Boudjaou y Bourassa, 2012) como “el proceso de realización de las intenciones de investigación y cambio en el que los actores tienen una gran autonomía sobre los fines y los medios”.

En el presente capítulo no se tiene en cuenta el objeto de la investigación (la experimentación con el reembolso global y sus efectos), que es objeto de otro estudio (no publicado). Se centra en el análisis del enfoque utilizado para llevar a cabo esta investigación-acción en el contexto de una colaboración o asociación entre los investigadores y las partes interesadas (Razón, 2011).

En el contexto de este proyecto, lo ideal era reunir a todas las partes interesadas (ONG, trabajadores sanitarios de diversos niveles, representantes de la población) para llevar a cabo un análisis reflexivo del enfoque utilizado para llevar a cabo la investigación-acción y extraer las lecciones aprendidas. Sin embargo, la utilización de ese análisis por sí sola plantea varios problemas. La mayoría de las partes interesadas tienen muy pocos conocimientos científicos sobre el enfoque de la aplicación de la investigación-acción para llevar a cabo ese análisis. Se consideró necesario un análisis externo basado en criterios objetivos para realizarlo de manera

adecuada e independiente. Sin embargo, según los enfoques utilizados en la esfera de la evaluación, la realización de un solo análisis externo también podría dar lugar a muchos problemas, como la difícil utilización de los resultados. Sobre este tema, Bonami (2005) señala:

Muchos informes de evaluación o de asesoramiento proporcionan los armarios de los responsables de la toma de decisiones antes de desaparecer en los archivos o en otros lugares. No es tan trivial como parece recordar que el resultado esperado de los procesos de evaluación tanto internos como externos radica en las decisiones que les siguen y que refuerzan, confirman, cambian y modifican las prácticas pedagógicas y de gestión en las escuelas.

A la luz de lo anterior, nuestro análisis se llevó a cabo en tres etapas: i) síntesis del conocimiento sobre el enfoque para llevar a cabo la investigación-acción e identificación de las mejores prácticas para llevarla a cabo, ii) análisis externo del enfoque de IA utilizando la red desarrollada de mejores prácticas, y iii) análisis reflexivo de las partes interesadas sobre el enfoque utilizado para llevar a cabo la IA.

Contexto y programa evaluados

Descripción del contexto en el que se desarrolla el programa

Desde 2008, los distritos sanitarios de Sebba y Dori se han beneficiado de un proyecto piloto para eximir a los niños menores de cinco años y a las mujeres embarazadas y lactantes, dos grupos con escaso acceso a la atención de la salud, del pago de la atención sanitaria. Los centros de salud (HC) se encargan de proporcionarles atención médica gratuita (consultas, partos, hospitalizaciones, atención médica y de enfermería, medicamentos, etc.). A cambio, la ONG alemana Help, que recibe una subvención de la Oficina Humanitaria de la Comunidad Europea (ECHO) de la Unión Europea, se ha comprometido a actuar como tercero pagador. Las dificultades inesperadas en la gestión de este proyecto pronto se hicieron evidentes. De hecho, todos los gastos relacionados con la atención de estos pacientes en los centros de salud se calculan manualmente con gran detalle antes de que se reclame su reembolso a la ONG. Al principio del proyecto, el personal

sanitario editaba y rellenaba documentos administrativos de varias páginas cuando los niños visitaban los centros de salud. Un efecto previsto de este servicio gratuito era el aumento de la tasa de asistencia a los centros de salud por parte de la población destinataria. El aumento de la cantidad de documentos que debían rellenarse condujo entonces, de manera inesperada, a un aumento de la carga de trabajo del personal sanitario. Los costos relacionados con la producción de los documentos también se volvieron exorbitantes para la ONG, lo que alargó el tiempo necesario para reembolsar a los centros de salud. Esta situación fue objeto de varias quejas expresadas por el personal sanitario, los miembros de los comités de gestión (COGES) de los centros de salud y los agentes de las ONG. Ante la recurrencia de estas quejas, se han intentado varias soluciones, sin mucho éxito, entre ellas la reducción del número de documentos a rellenar y la instalación de grandes comedores en los centros de salud para el archivo de documentos (más espacio de almacenamiento en las oficinas).

Ante la persistencia de los efectos inesperados del reembolso en la vida real y las quejas expresadas, las partes interesadas en el proyecto, con la ayuda de un equipo de investigación de la Universidad de Montreal, acordaron realizar una investigación-acción a partir de 2011 para experimentar una innovación, a saber, el “reembolso a tanto alzado”.

Descripción del programa

La investigación-acción se refería únicamente al cuidado de los niños y niñas. La medida propuesta consistía en pagar una suma global de 1300 francos CFA (2,7 dólares de los EE.UU.) por niño y mes al centro de salud. Esa suma se determinó sobre la base de los costos mensuales medios de la atención de los niños y niñas, durante el período comprendido entre enero de 2009 y junio de 2011 en los seis primeros centros de salud que experimentaron con el reembolso de una suma global. Sobre la base de esta cantidad, los centros de salud tenían que contar y multiplicar el número de niños y niñas tratados por la suma global y luego solicitar el reembolso a la ONG. Los trabajadores/as de la salud tenían que utilizar los documentos o instrumentos habituales de recogida de datos del Ministerio de Salud para registrar a los pacientes jóvenes y contribuir así a reducir su carga de trabajo y el costo de producción de los instrumentos. La ONG debía transferir la gestión de los reembolsos al distrito sanitario de Sebba y asegurar una

supervisión regular. En cuanto al componente de investigación, se centró en i) el estudio de la viabilidad técnica y operacional del reembolso de la suma global, ii) el cálculo de la evolución de los costos mensuales medios de la atención a los niños/as y iii) la evaluación del volumen de trabajo.

En esta investigación-acción se planificaron tres fases (véase el cuadro I). En la primera fase se reclutaron seis centros de salud de un total de 13. Para proporcionar confianza, se les advirtió que no había riesgo de pérdida económica. Los centros de salud fueron reembolsados sobre la base de una suma global de 1.300 francos CFA y el número de pacientes recibidos. Al final de esta fase, el cálculo de todos los costos reales de los servicios de los centros de salud debería permitir alcanzar el equilibrio financiero. Si los centros de salud habían gastado más de las sumas recibidas durante el reembolso, la ONG se comprometía a reembolsarles la diferencia. Si, por el contrario, los centros de salud habían recibido más reembolsos que sus gastos reales, se comprometieron a reembolsar la diferencia a la ONG. Tras 11 meses de experimentación, los otros siete centros de salud se introdujeron en la segunda fase de la investigación-acción “sin riesgo financiero”. Por otra parte, los seis primeros centros de salud y la ONG Help habían asumido en ese momento un “riesgo financiero”, es decir, más reembolsos en caso de pérdidas o ganancias financieras. Esta segunda fase duró 12 meses. En la tercera fase, los 13 centros de salud habían asumido un riesgo financiero. Esta fase también duró 12 meses. El Distrito de Salud de Dori fue elegido como distrito de control durante toda la investigación-acción.

Tabla 1 : Cita para el experimento de reembolso de la suma global (Fuente: Guía de aplicación de Action Research)

Investigación-acción	Fase 1	Fase 2	Fase 3
Duración de las fases	11 meses	12 meses	12 meses
Grupo 1 (6 CS)	Reembolso sin riesgo financiero	Reembolso con riesgo financiero	Reembolso con riesgo financiero
Grupo 2 (7 CS)	Reembolso en efectivo	Reembolso sin riesgo financiero	Reembolso con riesgo financiero
Distrito sanitario de Dori	Reembolso real	Reembolso real	Reembolso real

Método de análisis

El análisis del enfoque utilizado para llevar a cabo la investigación-acción se realizó en tres etapas.

1. Elaboración de una plantilla de “mejores prácticas”

El primer paso fue la elaboración de una plantilla de “mejores prácticas”, que se utilizará para analizar el proceso de investigación-acción. Según Patton (2001), “las prácticas óptimas basadas en principios para orientar la acción son útiles, mientras que las que son normativas y específicas son difíciles de aplicar en diversos contextos. Consideramos que una “práctica óptima” es un “know-how”, una lista de tareas para llevar a cabo eficazmente las diferentes etapas de los ciclos de investigación-acción. Sobre la base de los elementos fundamentales de la definición de la investigación-acción, hemos hecho, a partir de la literatura (científica y gris), una síntesis de las prácticas que describimos como “mejores prácticas”, además de las ya identificadas por de Roy y otros (2013). Los criterios para elegir estas prácticas residen en el hecho de que fueron descritas repetidamente por varios autores, demostrando al mismo tiempo la solidez de sus pruebas.

2. Análisis externo del enfoque IA

El segundo paso fue realizar un análisis externo del enfoque de esta investigación-acción. Utilizamos nuestra tabla de “mejores prácticas” para realizar: i) una revisión de los documentos de esta investigación-acción, ii) observaciones sobre el terreno, en particular la realización de reuniones de seguimiento sobre la investigación-acción y las actividades en los centros de salud, y iii) entrevistas individuales con determinados informantes clave (sólo hombres) utilizando una guía de entrevista semiestructurada para complementar la información no contenida en los documentos. Con el coordinador de esta investigación-acción, reclutamos una variedad de informantes de las diversas partes interesadas, elegidos según su antigüedad, la calidad de su participación en la investigación-acción y el período en que los centros de salud entraron en las diversas fases de la investigación-acción. Nuestra muestra se distribuyó de la siguiente manera:

- Equipo de investigación: el investigador principal y el coordinador de investigación-acción
- Ayuda de la ONG: el controlador financiero de la subvención y su ayudante, el coordinador médico y un antiguo director de proyectos.
- Dirección Regional de Salud del Sahel: el farmacéutico
- Equipo de Marco del Distrito de Sebba: el Gerente
- Los Centros de Salud (HC): 4 enfermeras jefas, 4 agentes de salud (enfermeras) y 3 gerentes de depósitos de drogas.
- 3 Presidentes de SMC

Los datos se reunieron utilizando una grabadora digital. Luego se transcribieron y se introdujeron en el software QDA Miner versión 4.1.6. Los resultados de esta segunda etapa se utilizaron para preparar la tercera etapa, que se dedicó al análisis reflexivo de las partes interesadas.

3. Análisis reflexivo del enfoque de la investigación-acción por las partes interesadas

En esta última etapa, los diversos interesados realizaron un análisis reflexivo del enfoque utilizado para llevar a cabo la investigación-acción. Organizamos un taller de dos días que reunió, además de los representantes de la población, a todas las partes interesadas (n=19). Al comienzo de este taller, se hizo una presentación oral. Se trató la definición, los ciclos experimentales y las etapas de la investigación-acción y los objetivos del taller. A continuación, se realizó un ejercicio para ayudar a las partes interesadas a distinguir entre el tema de la investigación-acción y el enfoque utilizado para llevar a cabo la investigación-acción. Durante el taller se organizaron tres actividades.

La primera actividad reunió a las partes interesadas en tres subgrupos homogéneos para reflexionar sobre lo que más les gustaba y lo que menos les gustaba del enfoque y para formular recomendaciones para futuros proyectos de investigación-acción. A continuación, se hizo un resumen de las actuaciones, seguido de los debates en el pleno.

La segunda actividad consistió en administrar un cuestionario cuantitativo a las partes interesadas a fin de recabar su grado de acuerdo con la aplicación de las “prácticas óptimas” establecidas (89% de los encuestados).

La tercera actividad consistió en la presentación de los resultados preliminares del análisis externo a las partes interesadas, seguida de debates.

Al final de este proceso, presentamos una síntesis de la contribución de este método.

Resultados del análisis

Los resultados de la revisión de la literatura

Presentamos los resultados obtenidos en una serie de cuadros que proponen, para cada etapa del ciclo de investigación-acción, una tabla de análisis de las “mejores prácticas”.

Tabla 2a : Etapa preliminar del proceso de investigación-acción

Mejores prácticas	Descripción y pruebas de la práctica
Realización de un análisis de la situación	Se rastrean las preocupaciones expresadas por las partes interesadas para justificar la investigación de la acción. Su presentación pública a las partes interesadas y los debates dan lugar a 1) la intención de actuar (necesidad de hacer) y 2) la intención de investigar (necesidad de saber) (Demange 2012; Faure, Hocdé y Chia 2011; Faure y Chia 2011; Faure 2007; Roy 2013; Tousignant 2012; Waterman 2007).
Identificar a las partes interesadas	Esta práctica tiene como objetivo tener una diversidad de perspectivas en la investigación de la acción y una expresión de diferentes intereses. La elección debe ser 1) ampliada para incluir una variedad de actores y 2) representativa de los actores afectados por la cuestión. Los criterios de elección son diversos (interés, función, capacidad de negociación o de trabajo en común). La presencia de los encargados de adoptar decisiones y de los líderes de opinión es fundamental y facilita la difusión de los resultados (Champagne 2007; Demange 2012; Detardo-Bora 2004; Dickens 1999; Faure, Hocdé y Chia 2011; Gauthier 2010; McVicar 2012; Mikolasek, Chia, Pouomogne, Tabi 2009; Roy 2013).
Establecer órganos de gobernanza de la investigación-acción	Estos órganos tienen como objetivo la participación activa de las partes interesadas. Deben ser democráticos, especificando: 1) el papel de cada interesado y 2) los órganos de gestión (comité directivo, equipo de investigación, etc.). El comité directivo debe ser representativo de las partes interesadas (Casabiaca 1997; Champagne 2007; Faure, Hocdé y Chia 2011; Vespiren 2012).
Formalizar el compromiso recíproco de todas las partes interesadas	Establecer un marco de asociación equitativo para crear confianza. El contenido del marco de asociación debería ser transparente y abordar la propiedad intelectual y la gestión de conflictos, la confidencialidad, los mecanismos de toma de decisiones, etc. El contenido del marco de asociación debe ser transparente y abarcar la gestión de la propiedad intelectual y los conflictos, la confidencialidad, los mecanismos de adopción de decisiones, etc. Se materializa con la firma del marco de asociación (Casabiaca 1997; Faure, Hocdé y Chia 2011; Le May 2001; Mikolasek, Chia, Pouomogne y Tabi 2009; Roy 2013).
Definir el rol del investigador o investigadora desde el comienzo	El investigador o investigadora tiene tres responsabilidades científicas según Bourassa (2012): <ul style="list-style-type: none"> • estructurar la investigación de la acción • acompañamiento (animación, mediación, formación, asesoramiento) • producción de conocimientos (estimula la reflexión, el análisis de problemas y soluciones, y la acción) Esta posición no partidista permite a los investigadores ser co-actores y a las partes interesadas ser co-investigadores. Esto fomenta la confianza entre las partes interesadas evitando los conflictos de intereses y conduce a una mayor objetividad (Le May 2001; Mikolasek, Chia, Pouomogne y Tabi 2009; Roy 2013).

Tabla 2b : Planificación de la investigación-acción

Mejores prácticas	Descripción y pruebas de la práctica
Identificar el objetivo de la investigación de acción	Formulado sobre la base del análisis de la situación, este objetivo debería ser una prioridad para que el/la investigador/a y otras partes interesadas se comprometan con ellos. Los desafíos que se deben examinar ya deben ser identificados (Casabiaca 1997; D'Alonzo 2010; Desclaux 2010; Faure, Hocdé y Chia 2011; Gauthier 2010; Roy 2013; Zachariah 2009; Zachariah 2010).
Analizar las cuestiones relacionadas con el propósito y la aplicación de la investigación-acción	Según Lui y otros (1997), la presentación del objetivo de la investigación-acción a los interesados, seguida de debates, permite analizar las cuestiones relacionadas con el objeto de la investigación (viabilidad, financiación, participación de las partes, etc.), para obtener su acuerdo a fin de crear confianza antes de comenzar. En el caso de Le May et al (2001), el análisis de los problemas permite asegurar la colaboración para crear confianza entre los participantes.
Identificar las preguntas de la investigación	Formuladas por los/as investigadores/as sobre la base de los objetivos de la investigación-acción, las preguntas de investigación deben ser abiertas para producir conocimiento científico de manera inductiva. Deben ser priorizadas de acuerdo a la lógica de las entradas y salidas. Deben ser tratables en el curso de un ciclo de investigación-acción (Casabiaca 1997; Chevalier 2013; Faure, Hocdé y Chia 2011; Mikolasek, Chia, Pouomogne y Tabi 2009; Vespiren 2012).
Proponer soluciones adaptadas y realistas	Identificar escenarios para la evolución de las preocupaciones expresadas en el análisis de situación (Faure 2007). Tres características fundamentales de las soluciones para mantener la confianza y el interés de las partes interesadas son: 1) la innovación, 2) la adaptación o integración en el contexto para facilitar la aplicación, y 3) las pruebas o la experiencia (Casabiaca 1997; Faure, Hocdé y Chia 2011; Waterman 2007).
Planificar las actividades a realizar	<p>El protocolo de investigación debe ser realista, flexible, no cerrado, y progresivo de acuerdo con la evolución de la investigación-acción (Roy, 2013). Incluye los siguientes elementos (D'Alonzo 2010; Demange 2012; Faure, Hocdé y Chia 2011; Mshelia 2013; OMS 2008; Waterman 2007):</p> <ul style="list-style-type: none"> • especificaciones de estudio adaptadas (tipo de investigación) • selección de objetivos • representatividad de los sitios experimentales de transferencia • métodos apropiados de recogida de datos • consideraciones éticas • identificación de los indicadores de medición que se deben recoger • cronograma de actividades <p>Demange (2012) identificó dos principios para su desarrollo: 1) desarrollo por un grupo pequeño (investigador/a, facilitador de la investigación, trabajador/a de campo) y 2) presentación al comité directivo para su discusión.</p>

Tabla 2c : Puesta en marcha de la investigación-acción

Mejores prácticas	Descripción y pruebas de la práctica
Elaborar guías para la ejecución de las actividades	Dado que las actividades realizadas en el marco de la investigación-acción son experimentales, deben definirse, describirse y normalizarse claramente en guías para facilitar la ejecución en todos los lugares. Se debe dar poder a los interesados (Champagne 2007; Demange 2012; Faure, Hocdé y Chia 2011; Faure 2007; Peters y Adam, 2013).
Organizar sesiones de trabajo sobre el terreno	Organización de sesiones de trabajo en los lugares de trabajo de los trabajadores sobre el terreno para examinar las actividades realizadas, rellenar los instrumentos de reunión de datos, detectar posibles dificultades y aportar soluciones, y fortalecer las capacidades de los agentes sobre el terreno si es necesario. El seguimiento también puede hacerse por correo electrónico, teléfono (Champagne 2007; D'Alonzo 2010; Demange 2012; Detardo-Bora 2004; Vespiren 2012).

Tabla 2d : Recogida de datos de la investigación-acción

Mejores prácticas	Descripción y pruebas de la práctica
Determinar los métodos de recogida de datos	Según Dickens y otros (1999), dado que las acciones de investigación en acción son experimentales, es preciso reunir datos útiles de manera exhaustiva para generar conocimientos (Dickens y otros 1999; Faure 2007; Mikolasek, Chia, Pouomogne y Tabi 2009). La triangulación (métodos, datos, tiempo e investigadores) aumenta el rigor científico de la investigación de acción (Champagne 2007). Esto incluye entrevistas individuales, grupos de discusión, observaciones y revisión de documentos (Gauthier 2010; McVicar 2012; Roy 2013). La retroalimentación de los datos estimula la reunión de datos (Mshelia 2013).
Desarrollar herramientas de recolección estandarizadas	Las herramientas de recolección se estandarizan adaptando su contenido y forma (Demange 2012). Esto permite la explotación científica de los datos, al tiempo que limita el sesgo, comparando los datos entre diferentes sitios experimentales. Pueden ser guías de observación, guías de entrevistas, cuestionarios, hojas de seguimiento, cuadernos de bitácora, etc. Se debe hacer una prueba previa para evaluar el contenido. La utilización de los instrumentos existentes puede facilitar la reunión de datos.
Asegurar la trazabilidad de la transmisión de datos haciendo responsables a los actores	Todos los datos que se recogen se recopilan y transmiten según un circuito determinado, mencionando las fechas de transmisión. Esto hace que sea fácil encontrar los datos perdidos. El archivo de datos es un activo que permite volver a los datos en cualquier momento. Esta buena práctica se ha observado en el campo.

Tabla 2e : Reflexión-evaluación de la investigación-acción

Mejores prácticas	Descripción y pruebas de la práctica
Organizar reuniones periódicas de seguimiento	<ul style="list-style-type: none"> • Organizar reuniones periódicas en el marco del comité directivo para el seguimiento y la evaluación de la investigación-acción • Tener una diversidad de opiniones en la validación e interpretación de los resultados (Día de Todos los Santos 2012)
Validar los resultados en conjunto	<ul style="list-style-type: none"> • Compartir los resultados obtenidos tras el análisis del comité científico con el comité directivo • Participar en el debate entre las partes interesadas para interpretar los resultados • Criterios de validación de los resultados según Chevalier (2013) • Fiabilidad (véase la triangulación de datos) • Consenso (validez del significado, significado dado por los actores)
Hacer un balance de cada ciclo del proceso	<ul style="list-style-type: none"> • Discutir y encontrar juntos soluciones a las dificultades que enfrentan los actores en el campo • Redefinir las preguntas de investigación innovadoras que vienen de los actores en el campo para un nuevo ciclo • Criterios para poner fin a un ciclo: satisfacción de los actores con la resolución del problema (Roy 2013)
Formalizar el conocimiento	<ul style="list-style-type: none"> • Escribir notas en forma de diario • Criterios de validación de conocimientos según Casabiaca y Roy (Casabiaca 1997; Roy 2013) • Nivel de resolución de los problemas identificados al comienzo • Criterios para la replicación del conocimiento: transferibilidad, relevancia, fertilidad, operatividad, aprendizaje

Una breve descripción de los resultados del análisis externo

Etapa preliminar

Las quejas expresadas por los actores durante la ejecución del proyecto de exención del pago de la atención fueron los elementos esenciales para justificar esta investigación-acción. Debido a su persistencia (varias soluciones infructuosas) y a su pertinencia, crearon una fuerte intención

de acción (necesidad de actuar) entre todas las partes interesadas en el proyecto, así como una intención de investigación (necesidad de saber). Aunque la búsqueda de una solución era una prioridad para las partes interesadas, la idea de hacer una investigación-acción, propuesta por el equipo de investigación, no fue aprobada inicialmente por los demás agentes del proyecto, porque había demasiados intereses financieros en juego. Frente a esta reticencia, los investigadores e investigadoras, tuvieron paciencia durante seis meses para identificar gradualmente a las partes interesadas que eran representativas de los diversos agentes afectados por el problema a nivel local. En el plano nacional, algunos responsables de la adopción de decisiones no participaron, en particular los Programas Nacionales de Salud Maternoinfantil y la Dirección de Investigación del Ministerio de Salud.

Fase de planificación

Las soluciones propuestas eran flexibles, graduales y adaptadas al contexto para facilitar su aplicación. Era necesario adoptar varias medidas para fomentar la confianza entre las diversas partes interesadas, incluidos los proveedores de servicios de salud y los miembros de los comités de gestión de los centros de salud. Estos pasos representaban los diferentes ciclos de investigación-acción experimental.

Etapas de reflexión-evaluación

La proyección de los resultados de la investigación durante las reuniones de seguimiento permitió a las partes interesadas ver los méritos de la investigación-acción y compartir conocimientos. Sin embargo, la ausencia de normas de funcionamiento no permitía canalizar los efectos indirectos en las declaraciones de las partes interesadas, como las rivalidades que se manifestaban y la tendencia a estigmatizar a los agentes que no habían obtenido buenos resultados durante las reuniones. El equipo de investigación entabló un debate abierto con las partes interesadas para interpretar los resultados.

Breve descripción de los resultados del análisis reflexivo de las partes interesadas

Las partes interesadas apreciaron sobre todo su participación en el proceso de aplicación de la investigación-acción, ya que cada uno tiene sus propias especificidades y la participación de todos ellos ayudó a reducir los obstáculos durante la aplicación. La firma de los acuerdos creó confianza y estableció su compromiso. La elección de los grupos de comparación les permitió observar mejor los efectos de la investigación-acción a lo largo del tiempo y disipar sus preocupaciones. La vigilancia y la evaluación permitieron compartir las experiencias y los niveles de información y tomar decisiones conjuntamente. Por unanimidad, todos los partes interesados valoraron enormemente la aplicación de todas las prácticas óptimas identificadas durante la etapa de reflexión-evaluación.

Sin embargo, apreciaron menos la falta de control sobre la duración de la investigación-acción, que estaba vinculada a la financiación externa, lo que dio lugar a una multiplicidad de acuerdos. La demora en la firma de los acuerdos debido a la lentitud administrativa fue a veces una fuente de desmotivación. También informaron de insuficiencias en la reunión de datos, entre otras cosas, de instrumentos mal informados o mal ubicados. El retraso en la transmisión del mandato de las reuniones de seguimiento y la corta duración de las mismas dieron lugar a un análisis parcial y tardío de los resultados.

Análisis reflexivo del proceso de análisis

Al final de nuestro trabajo, realizamos un análisis reflexivo de nuestro enfoque y nuestro proceso de análisis.

Lecciones aprendidas del proceso de evaluación

Al final del proceso de análisis, las lecciones aprendidas pueden resumirse de la siguiente manera:

- el análisis de la situación al principio creó un intento de acción e investigación entre las partes interesadas

- la identificación y participación de todas las partes interesadas desde el principio legitimó el proceso y dio a la investigación-acción todas las posibilidades de éxito
- el debate de las cuestiones y la firma de los acuerdos marcaron la confianza y el compromiso de las diversas partes implicadas en el proceso
- las soluciones propuestas fueron flexibles, progresivas, innovadoras y adaptadas al contexto
- la elaboración de una guía de aplicación, las visitas sobre el terreno y los instrumentos de reunión de datos facilitaron las acciones e investigaciones experimentales.
- la organización de reuniones periódicas y democráticas de seguimiento, enmarcadas en normas de funcionamiento, estimuló la reflexión y la evaluación de los resultados por consenso
- la escasa participación de ciertos actores en el Ministerio de Salud podría constituir un desafío para la aplicación de las recomendaciones de esta investigación.
- hay dificultades para reunir y organizar reuniones de seguimiento sobre la investigación-acción.

Fortalezas y debilidades de nuestro proceso de análisis

La elaboración de una tabla analítica de “mejores prácticas” con la descripción de sus características, así como la fuerza de sus pruebas ha permitido aplicarlas, estudiarlas y hacer un análisis externo fáctico basado en criterios objetivos mediante una triangulación de métodos. Las prácticas identificadas provienen de diversos autores y se han experimentado en diferentes contextos (países industrializados y en desarrollo) y en diversos ámbitos (salud, educación, agricultura, social), lo que permite ver su adaptación.

La combinación de los dos tipos de análisis (externo e interno) ofrece muchas ventajas. Ha permitido estructurar el proceso de análisis en dos etapas complementarias, utilizando al mismo tiempo una triangulación de métodos para reunir información adicional.

Durante el análisis externo, dado que el estudio de la documentación de la investigación-acción no proporcionó toda la información relativa a

la cuadrícula de “prácticas óptimas”, fue esencial una triangulación de los métodos (observaciones, entrevistas) para reunir datos complementarios. A este respecto, las observaciones permitieron observar las cuestiones de poder entre las partes interesadas durante las reuniones de seguimiento. Este análisis externo fue útil porque durante el análisis reflexivo, ningún subgrupo se refirió a cuestiones de poder, cuya gestión sigue siendo un factor determinante para el éxito de la investigación-acción (Bernier 2014). Sin embargo, no fue suficiente para reunir toda la información necesaria. Sin el análisis reflexivo, ciertas realidades no podrían haber sido detectadas por un punto de vista externo. Este análisis reflexivo permitió a las partes interesadas llevar a cabo reflexiones profundas sobre ciertos elementos ya identificados durante el análisis externo y aportar elementos adicionales. A este respecto, las partes interesadas expresaron preocupaciones detalladas sobre la organización práctica de la investigación-acción, como el retraso en el establecimiento de la financiación, en la organización de reuniones de seguimiento y en la firma de acuerdos, lo que a veces provocó desmotivación. Los valores culturales de las partes implicadas les han obligado ciertamente a actuar con moderación al tratar esta cuestión, especialmente en presencia de personas de diferentes orígenes. Sin embargo, de forma anónima, una persona se expresó sobre el tema en los siguientes términos:

necesitamos mejorar la motivación de los participantes durante las reuniones de investigación-acción porque no es un trabajo como cualquier otro, porque ponemos mucho esfuerzo intelectual en ello (menos de 15 euros/día como cuota de asistencia durante las reuniones de seguimiento). Debemos continuar la investigación-acción y mejorar las condiciones.

A la luz de lo anterior, observamos que la combinación de los dos análisis ha mejorado la calidad del estudio, como mencionaron el Barón y otros:

En primer lugar, la participación de los grupos afectados por un programa público en su proceso de evaluación permite una mejor calidad del juicio de valor que se hará sobre ese programa, porque el evaluador tendrá acceso a más información a través de sus intercambios con los participantes y el juicio se construirá a partir de una multiplicidad de opiniones informadas (Baron y Monnier, 2003).

A pesar de los puntos fuertes de ese análisis combinado, cabe señalar algunas limitaciones. Puede haber limitaciones para reunir las “prácticas óptimas” esbozadas por varios autores en el modelo de Roy y Prévost (2013). Ciertamente, nuestra revisión de la literatura no fue completa y no tuvo en cuenta todos los artículos científicos más relevantes, lo que dio lugar a tres versiones de la herramienta cuando se desarrolló. La falta de consenso en cuanto a la definición de la investigación-acción en la literatura científica afectó el posicionamiento de las “mejores prácticas” en las diferentes etapas. En la vida, no todas las etapas del ciclo de acción-investigación son siempre distintas, incluidas las etapas preliminares, de planificación y de reflexión-evaluación.

Principales cuestiones relacionadas con la realización del análisis

La primera cuestión es la capacidad de diseñar una plantilla de análisis externo capaz de comprender la realidad, que sirva de brújula para el análisis externo. En nuestro caso, la falta de consenso sobre la definición de la investigación-acción (Gauthier 2010) y las “mejores prácticas” fue una dificultad. La segunda cuestión se refiere a la capacidad de reunir a todos las partes interesadas para realizar un análisis reflexivo, sin tener que hacer frente a las limitaciones de tiempo y recursos financieros (Baron y Monnier 2003). En nuestro caso, el taller se pospuso dos veces debido a conflictos de programación. Dado que el presupuesto inicial para la investigación-acción no preveía tal análisis, fue necesario abogar por la organización del taller.

La tercera cuestión se refiere a la capacidad de conseguir que las partes interesadas se apropien del enfoque de análisis y de distinguir entre un objeto de investigación y un enfoque de aplicación de la investigación-acción, utilizando un vocabulario adaptado para producir conocimientos útiles. En nuestro caso, para no influir en el análisis de las partes interesadas, las sesiones fueron dirigidas por algunos de ellos. Sin embargo, como no participaron mucho en los preparativos, por varias razones, su participación no fue muy notable.

¿Cómo procedemos la próxima vez?

Si lo hiciéramos de nuevo, sería conveniente identificar claramente las palabras clave para facilitar la revisión de la literatura y establecer criterios más rigurosos para la selección de los artículos. Con respecto a nuestro enfoque de trabajo, la calidad de nuestro análisis podría verse sesgada por las fuertes relaciones que han existido entre el equipo de investigación-acción y las partes interesadas durante varios años en esta área.

En vista de las limitaciones, este instrumento de análisis debería difundirse, someterse a nuevas pruebas en diversos contextos y revisarse, de ser posible. El taller organizado no estaba suficientemente preparado con los cofacilitadores externos, en particular con los representantes de las ONG y el gerente del Distrito Sanitario. Podría haber habido más información a nivel de análisis reflexivo para generar más conocimiento. Las instrucciones dadas a los diferentes grupos de trabajo no eran lo suficientemente claras al principio. Si se volviera a hacer, se les debería haber dado la oportunidad de aclarar las instrucciones para que tuvieran el mismo nivel de información. Debido a la gran cantidad de documentos no archivados, ciertamente no consultamos todos los documentos que podrían haber arrojado más luz sobre el tema.

Conclusión

La realización de la investigación-acción implica diversos elementos que pueden constituir factores que limiten el logro de los objetivos si no se identifican y gestionan adecuadamente.

El análisis del enfoque utilizado para llevar a cabo la investigación-acción es útil para orientar los futuros proyectos de investigación-acción, pero todavía se enfrenta a muchos desafíos. Este trabajo ha permitido estructurar el proceso de análisis en tres etapas complementarias. La elaboración de una cuadrícula de “prácticas óptimas” proporcionó una brújula. La combinación del análisis externo con el análisis reflexivo de las partes interesadas ofrece muchas ventajas en cuanto a la complementariedad de los datos que se han de reunir y la calidad del análisis, al reunir los puntos de vista de las partes en mayor profundidad mediante una triangulación de métodos. Para llevar a cabo ese enfoque analítico, la

documentación debe ser cuidadosamente seleccionada a fin de constituir una cuadrícula de análisis externo basada en elementos de hecho. Para facilitar la reflexión de las partes interesadas, éstos deben participar desde el principio del proceso a fin de preparar el taller de intercambio de ideas y utilizar un vocabulario adaptado.

Referencias clave

Gauthier, B. (2010). La recherche-action. B. Gauthier (dir.), *Recherche sociale : de la problématique à la collecte des données* (p. 517-533). Québec : Presses de l'Université du Québec.

Un capítulo antiguo y excelente porque es uno de los primeros en abordar la investigación-acción como un enfoque científico en las ciencias sociales en lengua francesa.

Reason, P. et Bradbury, H. (dir.). (2011). *Handbook of action research: Participative inquiry and practice*. Londres : Sage Publications.

La biblia para aquellos/as que deseen comprender los detalles de los fundamentos epistemológicos, teóricos y conceptuales de la investigación-acción. El libro también está lleno de ejemplos de cómo se implementa la investigación-acción en una variedad de contextos y campos, incluso en países del Sur. También hay una sección sobre las aptitudes y la capacitación de los agentes que participan en la investigación-acción.

Roy, M. et Prévost, P. (2013). La recherche-action : origines, caractéristiques et implications de son utilisation dans les sciences de la gestion. *Recherches qualitatives*, 32(2), 129-151.

Un artículo que descubre las diversas etapas de la investigación de la acción y que sirvió de base para este capítulo, entre otras cosas.

Referencias

Avison, D., Lau, F., Myers, M. et Nielsen, P.A. (1999). Action Research: To make academic research relevant, researchers should try out their theories with practitioners in real situations and real organizations. *Communications of the ACM*, 42(1), 94-97.

- Baron, G. et Monnier, E. (2003). Une approche pluraliste et participative : coproduire l'évaluation avec la société civile. *Informations sociales*, 110, 120-129.
- Baum, F., McDougall, C., Danielle, S. (2006). Participatory action research. *J Epidemiol Community Health*, 60, 854-857.
- Bernier, J. (2014). La recherche partenariale comme espace de soutien à l'innovation. *Global Health Promotion* 2014 21: 58, 21(Suppl.1), 58-63.
- Boudjaou, M. et Bourassa, B. (2012). Introduction. In Bourassa, B and Boudjaoui, M. (dir.). *Recherches collaboratives en sciences humaines et sociales. Enjeux, modalités et limites*. Québec : Presses de l'Université Laval.
- Bourassa, B., Leclerc, C. et Fournier, G. (2012). Assumer et risquer une posture de recherche collaborative. In Bourassa, B. and Boudjaoui, M. (dir.) *Recherches collaboratives en sciences humaines et sociales. Enjeux, modalités et limites*. Québec : Presses de l'Université du Québec.
- Boyera, S., Koulla-Shirod, S., Abe, C., Spirea, B., Moattia, J-P. (2011). Implementing operational research to scale-up access to antiretroviral therapy for HIV infection: lessons learned from the Cameroonian experience. *Current Opinion in HIV and AIDS*, 6, 239-244.
- Casabiaca, F. et Albeladejo, C. (1997). Des multiples légitimités de la recherche action. *Etud. Rech. Syst. Agraires Dev*, 30, 11-25.
- Champagne, M. (2007). La pratique de la recherche-action. Entre utopie et nécessité. In Dorvill, H. (dir.) *Problèmes sociaux Tome III. Théorie et méthodologie de la recherche*, 463-490. Montréal : Presses de l'Université du Québec.
- Chevalier, J., Buckless, D. J et Bourassa, M. (2013). *Guide de la recherche action, la planification et l'évaluation participative, Module 2*. SAS2. Ottawa, Canada.
- D'Alonzo, K. T. (2010). Getting started in CBPR: lessons in bulding community partnerships for new researchers. *Nursing Inquiry*, 17(4), 282-288.
- Demange, E., Henry E. et Préau M. (2012). *De la recherche en collaboration à la recherche communautaire. Un guide méthodologique*. Paris. ANRS/Coalition Plus. Coll. Sciences sociales et sida. 2012.

- Desclaux, A., Kouanda, S., and Obermeyer, C.M. (2010). Stakeholders participation in operational research on HIV: insights from Burkina Faso. *AIDS*, 24(1), S79-85.
- Dickens, L and Watkins, K. (1999). Action Research: Rethinking Lewin. *Management Learning*, 30, 127.
- Ministère de l'économie et des finances. République du Burkina Faso (2011). *Stratégie de croissance accélérée et de développement durable 2011 – 2015*. http://www.unpei.org/sites/default/files/PDF/Burkina_Faso_PRSP_2011.pdf
- Faure, G., Hocdé, H., et Chia, E. (2011). Action research methodology to reconcile product standardization and diversity of agricultural practices: A case of farmers' organizations in Costa Rica. *Action Research*, 9, 242-260.
- Faure, G., Hocdé, H. et Meneses, D. (2007). Les organisations paysannes du Costa Rica construisent leur vision de l'agriculture familiale : une démarche de recherche-action marquée par une rupture. *Cahiers Agricultures*, 16(3), 205-2011.
- Hercot, D., Meessen, B., Ridde, V. et Gilson, L. (2011). Removing user fees for health services in low- income countries: a multi-country review framework for assessing the process of policy change. *Health Policy and Planning*, 26, ii5-ii15.
- Le May, A. et Lathlean, F. (2001). Action research: A design with potential. *Nursing Times Research*, 6, 502.
- Lui, M. (1997). La validation des connaissances au cours de la recherche action. *Etud. Rech. Syst. Agraires Dev*, 30, 183-196.
- McVicar, A., Munn-Giddings, C. et Abu-Helil, C. (2012). Exploring the development of action research in nursing and social care in the UK: A comparative bibliometric review of action research designs in social work (2000-2010). *Action Research*, 10, 79.
- Mikolasek, O., Barlet, B., Chia, E., Pouomogne, V. et Tabi, M.T.E. (2009). Développement de la petite pisciculture marchande au Cameroun: la recherche-action en partenariat. *Cahiers Agricultures*, 18(2-3), 270- 275.
- Mshelia, C., Huss, R., Mirzoev, T., Elsey, H., Baine, S.O., Aikins, M., et al. (2013). Can action research strengthen district health management and improve health workforce performance? A research protocol. *BMJ Open* 3.

- OMS. (2008). *Guide to operational research in programs support by the Global Fund*. Organisation mondiale de la santé. Fonds mondial pour la lutte contre Sida, la tuberculose et le paludisme.
- Patton, M.Q. (2001). Evaluation, knowledge management, best practices, and high quality lessons learned. *American Journal of Evaluation*, 22, 329.
- Peters, D. H., Tran, N. T. et Adam, T. (2013). *Implementation research in health: a practical guide*. Geneva: World Health Organization. Alliance for Health Policy and Systems Research
http://who.int/alliance-hpsr/alliancehpsr_irpguide.pdf
- Peters, D.H., Taghreed, A., Alonge, O., Agyepong, I.A. et Tran, N. (2013). Implementation research: what it is and how to do it. *BMJ* 347.
- Reason, P., et Bradbury, H. (2011). Introduction: Inquiry and participation in search of a world worthy of human aspiration. In P. Reason & H. Bradbury (Eds.), *Handbook of action research: Participative inquiry and practice* (pp. 1-14). London: SAGE.
- Roy, M. et Prévost, P. (2013). La recherche-action : origines, caractéristiques et implications de son utilisation dans les sciences de la gestion. *Recherches qualitatives*, 32(2), 129-151.
- Swann, C. (2002). Action research and the practice of design. *Design Issues*, 18(2).
- Tousignant, P., Remondin, M., Laurendeau, M-C., et Bergeron, P. (2012). Une approche participative pour une recherche ancrée dans la pratique en première ligne. *Réseau de recherche en santé des populations du Québec. Carnets-synthèse* 11.
<http://santepop.qc.ca/fr/activites/carnets.html>
- Vespiren, M. et Chia, E. (2012). Rôle d'une recherche-action sur la diffusion des savoirs et la modification du contexte social. Bourassa, B. and Boudjaoui, M. (dirs) *Recherches collaboratives en sciences humaines et sociales. Enjeux, modalités et limites*. Montréal : Presses de l'Université du Québec.
- Waterman, H., Marshall, M., Noble, J., Davies, H., Walshe, K., Sheaff, R. et Elwyn, G. (2007). The Role of Action Research in the Investigation and Diffusion of Innovations in Health Care: The PRIDE Project. *Qual Health Res*, 17, 373.

Zachariah, R., Harries, A.D., Ishikawa, N., Rieder, H.L., Bissell, K., Laserson, K., et al. (2009). Operational research in low-income countries: what, why, and how? *Lancet Infect Dis*, 9(11), 711- 717.

Zachariaha, R., Ford, N. Draguez, B., Yunc, O. et Reida, T. (2010). Conducting operational research within a non governmental organization: the example of Médecins Sans Frontières. *International Health* 2(1-8).

Resumen / Résumé / Abstract

Desde 2011, una investigación-acción se lleva a cabo como parte de un proyecto para subvencionar el acceso a la atención de los niños menores de cinco años y de las mujeres embarazadas y lactantes en el distrito de salud de Sebba, en Burkina Faso. Su objetivo es experimentar una innovación, el reembolso de los servicios de asistencia sanitaria gratuitos prestados a los pacientes, con el fin de resolver las dificultades inherentes al proyecto de subvención del acceso a la asistencia encontradas por los trabajadores de la salud y de la ONG financiadora. Los desafíos incluyen la carga de trabajo del personal de salud, los costos de producir herramientas de recolección de datos, etc. Como los resultados preliminares de esta investigación-acción fueron concluyentes, parecía necesario analizar el enfoque utilizado para llevarla a cabo. Para lograrlo, un análisis externo combinado con un análisis reflexivo de las partes interesadas proporcionó las lecciones apropiadas para guiar futuras investigaciones-acciones. El éxito de este enfoque de análisis se ha logrado a través de (i) el desarrollo de una tabla de análisis con elementos factuales para llevar a cabo el análisis externo, (ii) una buena preparación del análisis reflexivo y (iii) la participación de las partes interesadas en el proceso de análisis desde el principio.

Depuis 2011, une recherche-action est menée dans le cadre d'un projet de subvention de l'accès aux soins des enfants de moins de cinq ans, des femmes enceintes et allaitantes au district sanitaire de Sebba au Burkina Faso. Elle vise à expérimenter une innovation, le remboursement au forfait des prestations de soins de santé délivrés gratuitement aux patients afin de résoudre les difficultés inhérentes au projet de subvention de l'accès aux soins rencontrées par les agents de santé et l'ONG subventionnaire. Parmi les difficultés figurent entre autres la charge de travail pour le personnel de santé, les coûts de production des outils de collecte des données. Les

résultats préliminaires de cette recherche-action étant probants, il est apparu nécessaire d'analyser la démarche utilisée pour la mettre en œuvre. Pour y parvenir, une analyse externe combinée à une analyse réflexive des parties prenantes a permis de tirer les leçons idoines qui ont permis de guider d'autres recherches-actions. La réussite d'une telle démarche d'analyse est passée par i) l'élaboration d'une grille d'analyse comportant des éléments factuels pour mener à bien l'analyse externe, ii) une bonne préparation de l'analyse réflexive et iii) une implication des parties prenantes au processus d'analyse dès le début.

Since 2011, action research has been conducted as part of a project to subsidize access to care for children under five, pregnant and lactating women in Sebba health district in Burkina Faso. It aims to test an innovation, the reimbursement of health care benefits delivered free of charge to patients in order to solve the difficulties inherent to the project of grant of access to care met by the health workers and the subsidizing NGO. As a difficulty he had in other workload for the health staff, the costs of production of data collection tools. The preliminary results of this action research being conclusive, it appeared necessary to analyze the approach used to implement it. To achieve this, an external analysis combined with a reflective analysis of the stakeholders made it possible to draw the appropriate lessons that made it possible to guide other action research. The success of such an analysis has gone through i) the development of an analysis grid with factual elements to carry out the external analysis, ii) a good preparation of the reflexive analysis and iii) Stakeholder involvement in the analysis process from the beginning.

Aka Bony Roger Sylvestre es médico de profesión y posee una maestría en salud pública con opción a la salud mundial en 2014 de la Universidad de Montreal (Canadá) con orientación a la evaluación. En calidad de pasante del Centro de Investigación del Centro Hospitalario de la Universidad de Montreal, diseñó un instrumento y analizó el proceso de aplicación de la investigación para la acción en materia de salud en Burkina Faso. Los resultados de este trabajo han sido objeto de varias publicaciones, en particular en la revista de la Société Québécoise d'Évaluation de Programme au Canada y han obtenido un premio a la excelente en el concurso

“Evaluation d’Or 2015” en Costa de Marfil. Desde enero de 2018, es jefe de un proyecto de investigación sobre la tuberculosis en los niños en el Programa PACC-CI en Costa de Marfil. De 2015 a 2017, estuvo a cargo de la supervisión y evaluación del Programa Ampliado de Inmunización del Ministerio de Salud de Costa de Marfil.

Valéry Ridde es Director de Investigación en el CEPED, una Unidad Conjunta de Investigación asociada a la Universidad de París Descartes y al Instituto de Investigación para el Desarrollo (IRD). Es investigador habitual del *Institut de recherche en santé publique de l’Université de Montréal* (IRSPUM) e investigador asociado del *Centre interdisciplinaire de recherche en développement international et société* (CIRDIS) de Montreal. Es editor asociado de las revistas *BMJ Global Health* y *Global Health Promotion*, miembro del consejo editorial del *Canadian Journal of Program Evaluation* y *Qualitatif! Revue africaine de sciences sociales*. Su investigación se centra en la cobertura universal de la salud, la financiación de los servicios de salud, la evaluación de programas, las políticas de salud pública y la promoción de la salud.

Ludovic Queuille es Consultor en Gobernanza y Financiación de la Salud de la Organización Panamericana de la Salud (OPS)/Organización Mundial de la Salud (OMS) en Haití desde septiembre de 2014. En su calidad de intermediario del conocimiento (2008-2014) para el Centro de Investigación de la Universidad de Montreal, con sede en Uagadugú, se encargó, entre otras cosas, de esta investigación-acción sobre el pago de una suma global en el marco de un proyecto para subvencionar el pago de la atención a los niños y niñas, las mujeres embarazadas y las madres lactantes en la región del Sahel de Burkina Faso (<http://www.equitesante.org/helpburkina/>).

Cita

Aka Bony Roger Sylvestre, Valéry Ridde y Ludovic Queuille (2020). El análisis de un proyecto de investigación-acción. Combinación de enfoques en el sector de la salud en Burkina Faso. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 135-161. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

PARTE III

MÉTODOS MIXTOS

6. Revisiones sistemáticas mixtas

Un ejemplo sobre la financiación basada en los resultados

QUAN NHA HONG, ANNE-MARIE TURCOTTE-TREMBLAY Y PIERRE PLUYE

Revisiones sistemáticas mixtas

Definición del método

Las revisiones sistemáticas mixtas son revisiones de la literatura que utilizan un enfoque sistemático para combinar estudios cuantitativos, cualitativos y mixtos.

Fortalezas del método

- Utiliza un procedimiento explícito, transparente y reproducible
- Combina estudios cuantitativos, cualitativos y mixtos
- Permite explicar fenómenos complejos y multidimensionales
- Proporciona una imagen más completa de la literatura

Desafíos del método

- Requiere tiempo y recursos para reducir al mínimo los sesgos en la detección, selección, divulgación e interpretación
- Requiere la integración de datos cuantitativos y cualitativos

En los últimos años se ha producido un desarrollo considerable de los métodos de síntesis de los trabajos científicos y un aumento de los tipos de revisiones, incluidos las revisiones sistemáticas mixtas. En el presente capítulo se describen brevemente los pasos de este método, haciendo hincapié en la síntesis de datos, uno de los pasos del proceso de revisión que supone el análisis de los datos para construir un conjunto coherente (Mays, Pope y Popay 2005).

Las revisiones mixtas de la literatura utilizan un enfoque sistemático e incluyen estudios que utilizan métodos cuantitativos, cualitativos y/o mixtos (Pluye y Hong 2014). Se utilizan otros términos para referirse a este tipo de revisión: revisión integradora (*integrative review*) (Whittemore y Knafl 2005) o revisión de métodos mixtos (*mixed methods review*) (Harden 2010). Por ello, aún no existe un consenso sobre la terminología. En este capítulo, utilizaremos el término “revisión sistemática mixta”.

La revisión sistemática mixta se basa en los principios de los métodos mixtos, es decir, la integración de las fases, los resultados y los datos cualitativos y cuantitativos, a fin de aprovechar su complementariedad (Creswell y otros, 2011).



Figura 1: Tipos de estudios incluidos en una revisión sistemática mixta

En general, se pueden identificar tres tipos principales de estudios primarios: estudios cuantitativos, estudios cualitativos y estudios mixtos. Como se muestra en las zonas de intersección de la figura 1, una revisión que combine dos o tres tipos de estudios primarios puede considerarse mixta, siempre que se utilicen datos tanto cualitativos como cuantitativos (Pluye y Hong 2014). Además, dado que los estudios mixtos se caracterizan por la combinación de métodos cualitativos y cuantitativos (Creswell y otros, 2011), una revisión de esos estudios también puede considerarse mixta cuando se sintetizan datos cualitativos y cuantitativos. En otras palabras, la revisión de diferentes estudios dentro de la misma categoría no es mixta. Así, la

combinación de los resultados de diferentes tipos de estudios cuantitativos (por ejemplo, ensayos aleatorios y no aleatorios) constituye una revisión cuantitativa, mientras que la combinación de los resultados de diferentes tipos de estudios cualitativos (por ejemplo, estudios de casos y etnografía) genera una revisión cualitativa.

Al combinar estudios cualitativos, cuantitativos y mixtos, la revisión sistemática mixta permite abordar cuestiones complejas y profundizar en la comprensión de los fenómenos observados. Permite: i) una mejor comprensión de los resultados de los estudios cuantitativos, ii) una generalización de los resultados de los estudios cualitativos, sobre la base de inferencias estadísticas, o iii) una corroboración de los conocimientos obtenidos mediante métodos cuantitativos y cualitativos (Pluye et al. 2009). Además, la revisión sistemática mixta puede proporcionar un panorama más amplio y completo de la bibliografía sobre un tema determinado, en particular combinando preguntas complementarias como: ¿Cuán eficaz es una intervención? ¿Por qué la intervención es efectiva o no? ¿Cómo funciona la intervención? ¿Cuáles son los obstáculos y los facilitadores para la aplicación de la intervención? Las respuestas a estas preguntas complementarias pueden tener consecuencias prácticas, como facilitar la toma de decisiones por parte de quienes formulan políticas, gestionan o prestan servicios clínicos.

Se han propuesto varias taxonomías de revisiones de la literatura. En ellas, se han identificado una diversidad de tipos de revisiones que varían según varios criterios, como los objetivos perseguidos, los métodos (pasos) utilizados y el público destinatario (Cooper 1988; Grant y Booth 2009; Paré y otros 2015). La revisión sistemática mixta se considera el tipo más riguroso, ya que minimiza varios sesgos relacionados con la selección y el análisis de los estudios.

Todas las etapas de la revisión sistemática deben ser transparentes y explícitas para que pueda ser reproducido (Pluye et al. 2016). La revisión sistemática pretende responder a una pregunta claramente definida siguiendo un método sistemático y explícito para identificar los estudios pertinentes, evaluarlos y sintetizarlos (Chalmers, Hedges y Cooper 2002). En una revisión sistemática, hay múltiples fuentes de información. La búsqueda de literatura es realizada por bibliotecarios especializados y es exhaustiva

(incluyendo la literatura gris). La selección, la evaluación de la calidad de los estudios incluidos y la recuperación de los datos son llevadas a cabo por al menos dos especialistas de forma independiente.

Aplicación del método

El propósito de esta sección es esbozar los pasos a seguir para realizar una revisión sistemática mixta. Para ilustrar estos pasos, utilizamos el ejemplo de una revisión sistemática mixta sobre las consecuencias no deseadas de una intervención basada en la financiación basada en los resultados. En el Recuadro 1 se describe qué es el FBR y se explica la pertinencia de una revisión sistemática mixta en este ámbito.

Recuadro 1. La financiación basada en resultados (FBR)

La financiación basada en resultados (FBR) es una intervención destinada a proporcionar incentivos financieros a los trabajadores y las trabajadoras de la salud para mejorar la calidad y la cantidad de la atención sanitaria. Esta intervención se ha desarrollado en países de altos ingresos y se ha extendido rápidamente a los países de ingresos bajos y medianos (LMIC) con el apoyo de organizaciones internacionales que han invertido en la reducción de la mortalidad materno-infantil. Existen varias denominaciones y modelos de intervención de FBR. Sobre la base del modelo del Banco Mundial (Fritsche, Soeters y Meessen 2014), se establecen acuerdos de rendimiento para definir indicadores y objetivos de rendimiento medibles y verificables. Los centros de salud y/o los trabajadores y las trabajadoras de la salud son pagados por cada servicio de salud predeterminado. También se ofrecen bonos según el nivel de calidad de la atención. Los evaluadores independientes realizan auditorías, incluso de los registros e instalaciones, para promover la transparencia y la rendición de cuentas.

Un conjunto cada vez mayor de investigaciones cualitativas y cuantitativas sugiere que la FBR puede tener consecuencias no intencionadas más allá de los efectos positivos esperados, es decir, cambios deseables o no deseables para los que hay una falta de acción o causalidad intencional y que se producen en un sistema social como resultado de una innovación. Para sintetizar los resultados de los estudios existentes y complementarios sobre la FBR, llevaremos a cabo una revisión sistemática mixta.

Se seleccionó este tipo de revisión por su capacidad de tener en cuenta la heterogeneidad de los estudios y la complejidad de las intervenciones. La literatura sobre la FBR se caracteriza por una multiplicidad de contextos de aplicación, metodologías, métodos de investigación, modelos de intervención y resultados. La síntesis de estos estudios complementarios es necesaria para adquirir una comprensión completa del estado actual de los

conocimientos sobre esta compleja cuestión. Los resultados de esta revisión sistemática mixta permitirán, por una parte, crear un marco analítico preliminar para la investigación de las consecuencias imprevistas de la FBR en Burkina Faso y, por otra, ayudar a la adopción de decisiones sobre la aplicación de la FBR.

Como se presenta en el cuadro 1, las revisiones sistemáticas siguen principalmente ocho pasos. Los pasos de la revisión sistemática son similares a los de los estudios primarios, con la principal diferencia de que la unidad de análisis es un estudio (no participantes u organizaciones).

Tabla 1 : Etapas de la revisión sistemática y comparación con las etapas de un estudio primario

Revisión sistemática	Estudio primario
Formulación de una pregunta de investigación	Formulación de una pregunta de investigación
Definición de los criterios de elegibilidad	Definición de los criterios de elegibilidad
Identificación de las fuentes de los documentos	Identificación de las fuentes de datos
Realizar una búsqueda de literatura	Reclutamiento de participantes
Selección de estudios relevantes	Selección de los participantes
Evaluación de la calidad de los estudios seleccionados	No procede
Extracción de datos de estudios seleccionados	Recopilación de datos
Síntesis de los datos extraídos	Análisis de datos

A continuación, se describen brevemente los pasos de una revisión sistemática y se ilustran con el caso de la FBR. Al final del presente capítulo figura una lista de referencias metodológicas sobre el examen sistemático.

Paso 1: Formular una pregunta de investigación

En las revisiones sistemáticas mixtas, puede haber una o más preguntas para la búsqueda de literatura. Estas preguntas deben ser claras y precisas para poder centrarse en el tema. En nuestro caso sobre la FBR, formulamos la pregunta de la siguiente manera: ¿Cuáles son las consecuencias imprevistas de la FBR en los países de ingresos bajos y medios (LMIC)? Esta pregunta permite integrar los resultados cualitativos y cuantitativos. También podríamos haber añadido una pregunta complementaria como ¿Cuál es la importancia respectiva de estas consecuencias imprevistas? o ¿Qué estrategias se han puesto en marcha para hacer frente a estas consecuencias imprevistas? Nuestra elección de la redacción de la pregunta estuvo guiada por los recursos disponibles, así como por la utilidad y la relevancia. La redacción de la(s) pregunta(s) es importante porque influye en todos los pasos posteriores de la revisión.

Paso 2: Definir los criterios de elegibilidad

Los criterios de elegibilidad están asociados a los conceptos clave de la pregunta de investigación y consisten en criterios de inclusión y exclusión. La tabla 2 presenta los criterios desarrollados a partir de nuestra pregunta de investigación sobre la FBR. Es importante definir y justificar claramente cada criterio. Por ejemplo, decidimos excluir los estudios publicados antes del año 2000 porque la FRB se ha utilizado ampliamente en los países de ingresos bajos y medios desde el decenio de 2000. Además, limitamos la selección a los estudios científicos, es decir, estudios que consisten en una pregunta/objetivo de investigación, recopilación/análisis de datos (experimentales, de observación o de simulación) y resultados. La combinación de estudios cualitativos y cuantitativos aumenta el número de artículos que se han de sintetizar. Por consiguiente, es necesario mantener los criterios de selección tan reducidos como sea posible, es decir, centrarse en la(s) pregunta(s).

Tabla 2 : Criterios de inclusión y exclusión

	Criterio de inclusión	Criterio de exclusión
Problema	Baja calidad y cantidad de los servicios de salud	Problemas no relacionados con los servicios de salud
Intervención	Financiación basada en los resultados (incentivos del lado de la oferta)	No hay financiación o intervenciones basadas en los resultados que ofrezcan incentivos a los pacientes o a las compañías de seguros
Resultados	Consecuencias involuntarias	Sólo consecuencias intencionales
Lugar	Países de ingresos bajos y medios	Países de altos ingresos
Tipo de estudio	Estudios científicos con métodos cualitativos, cuantitativos o mixtos.	Estudios no científicos, protocolos de investigación, editoriales, directrices (es decir, texto sin recopilación de datos ni análisis utilizando una metodología de investigación)
Langues	Inglés, francés y español	Artículos escritos en un idioma distinto al inglés, francés y español
Años de publicación	2000 – 2016	Antes de 2000

Paso 3: Seleccionar las fuentes de información

Para garantizar que se identifiquen todos los estudios pertinentes al problema de investigación, se deben utilizar diversas fuentes de datos. Para este paso y el siguiente, se recomienda encarecidamente la colaboración con un bibliotecario o expertos en técnicas de búsqueda en bases de datos y otros tipos de fuentes. Para la revisión sistemática mixta sobre FBR, seleccionamos cinco bases de datos bibliográficas complementarias en los campos de la salud (MEDLINE, CINAHL, EMBASE), las humanidades y las ciencias sociales y económicas (FRANCIS, Business Source Premier). También se pueden utilizar otras fuentes de información para garantizar una búsqueda exhaustiva. Para nuestra revisión de la FBR, usamos otras cuatro fuentes de información adicionales:

a) Consultar la lista de artículos seleccionados de los exámenes sistemáticos sobre la FBR (Gorter y Meessen 2013; Witter y otros 2012) y la lista de referencias de los artículos seleccionados.

b) Buscar la literatura gris consultando los sitios de Internet sobre el tema (por ejemplo, www.rbfhealth.org, www.healthfinancingafrica.org, www.theglobalfund.org) y las tesis en la base de datos Proquest Dissertations & Theses Global.

c) Enviar la lista de artículos seleccionados a los investigadores con experiencia en la BFR para identificar los artículos pertinentes que faltan; y

d) Realizar una búsqueda de citas (es decir, buscar artículos que citen cada artículo seleccionado) en una base de datos bibliográfica especializada como Scopus o Google Scholar.

Paso 4: Desarrollar una estrategia de búsqueda de literatura

Para desarrollar una estrategia de búsqueda de literatura, es necesario identificar los términos relacionados con los conceptos clave en la pregunta de búsqueda. Para la revisión sistemática mixta de la FBR, desarrollamos una estrategia de búsqueda de bibliografía en torno a tres conceptos principales: consecuencias imprevistas, financiación basada en los resultados y países de ingresos bajos y medios (véase un ejemplo en el cuadro 3). Esta estrategia se adaptó para cada base de datos seleccionada. Una vez completada la búsqueda de literatura, todas las referencias deben ser exportadas a un software de gestión de documentos como Zotero, EndNote, RefWork o Mendeley. Dado que la búsqueda se realiza en varias bases de datos superpuestas, la misma referencia puede aparecer varias veces en el programa (duplicados). Estos duplicados se identifican y eliminan mediante métodos apropiados, como el método de Bramer et al (2016) o Rathbone et al (2015). Estos métodos identifican los duplicados seleccionando campos específicos (por ejemplo, autor, título, año, revista, página, volumen y número) en un orden determinado.

Tabla 3 : Ejemplo de una estrategia de búsqueda de literatura en PUBMED

Search	Query
#1	Search unintent* OR intend* OR unpredict* OR predict* OR unanticipat* OR anticipat* OR spillover* OR indirect* OR direct* OR unaware* OR aware* OR unexpect* OR expect* OR desirable* OR Undesirable* OR “cherry picking” OR “cream skimming” OR gaming* OR inequit* OR equit* [ALL]
#2	Search “pay for performance” OR “pay for value” OR “pay for quality” OR “P4P” OR “PBF” OR “value-based purchasing” OR (“monetary incentive” AND quality) OR (“monetary incentives” AND quality) OR (“financial incentive” AND quality) OR (“financial incentives” AND quality) OR (bonus AND quality) OR (bonuses AND quality) OR (reward AND quality) OR (rewards AND quality) OR “performance-based payment” OR “performance-based pay” OR “performance-based reimbursement” OR “performance-based financing” OR “performance-based contracting” OR “results-based financing” OR “output-based payment” OR “incentive reimbursement” OR “incentive reimbursements” OR “quality-based purchasing” OR “quality incentive” OR “quality incentives” [TIAB]
#3	Search “Developing Countries” OR “Africa” OR “Africa, Northern” OR “Africa South of the Sahara” OR “Africa, Central” OR “Africa, Eastern” OR “Africa, Southern” OR “Africa, Western” OR “Asia” OR “Asia, Central” OR “Asia, Southeastern” OR “Asia, Western” OR “Caribbean Region” OR “West Indies” OR “South America” OR “Latin America” OR Central America [MH]
#4	Search #1 AND #2 AND #3
#5	Limit #4 to ([english OR french OR spanish [LA]] AND [2000:2016 [DPJ]])

Los investigadores están interesados en desarrollar y probar filtros de búsqueda para identificar documentos relevantes sobre diferentes conceptos. Por ejemplo, se ha creado un filtro en Medline para identificar los estudios cuantitativos, cualitativos y mixtos (El Sherif *et al.* 2016), mientras que otro filtro identifica los documentos sobre los países de ingresos bajos y medios (Cochrane Effective Practice and Organisation of Care 2012). La utilización de filtros de búsqueda ayuda a facilitar la elaboración de estrategias de búsqueda de literatura y a orientar mejor los documentos potencialmente pertinentes.

Paso 5: Seleccionar los estudios

Este paso implica filtrar los resultados de la búsqueda de literatura para seleccionar estudios para su inclusión en la revisión sistemática mixta. La selección de los estudios se hace generalmente en dos partes. Primero se

leen los títulos y resúmenes para juzgar la elegibilidad de los documentos en base a los criterios preestablecidos en el paso 2. En segundo lugar, los textos se leen en su totalidad. En una revisión sistemática, la selección de todo el material es llevada a cabo por dos revisores independientes a fin de reducir el sesgo de selección. Si hay desacuerdo, los revisores se reúnen para establecer un consenso y una tercera persona participa para decidir si el desacuerdo persiste después del debate. El acuerdo entre revisores se puede estimar usando el Kappa de Cohen. Al final de este proceso de selección, se presenta un diagrama de flujo que muestra el número de documentos retenidos y excluidos en cada etapa y las razones de la exclusión (Figura 2) (adaptado de las directrices de PRISMA de Moher y otros (2009), traducidas en Urrutia y Bonfill [2010]). La creación del diagrama de flujo requiere una gran organización desde el principio. Todo el proceso debe ser documentado (principio de transparencia).

Para reducir el número de títulos y resúmenes que deben filtrarse, se desarrolló un clasificador automático de textos sobre estudios científicos (CATES) a raíz de un estudio que comparaba el rendimiento de varios clasificadores automáticos (Langlois et al. 2018). Este clasificador clasifica automáticamente las publicaciones indexadas en (1) estudios científicos (cualitativos, cuantitativos y mixtos) y (2) otros tipos de trabajos (por ejemplo, comentarios, editoriales, revisión de la literatura, artículos metodológicos, descripción de programas y directrices). El CATES está disponible en esta dirección web y puede ser utilizado de forma gratuita: <https://babel.iro.umontreal.ca/>.

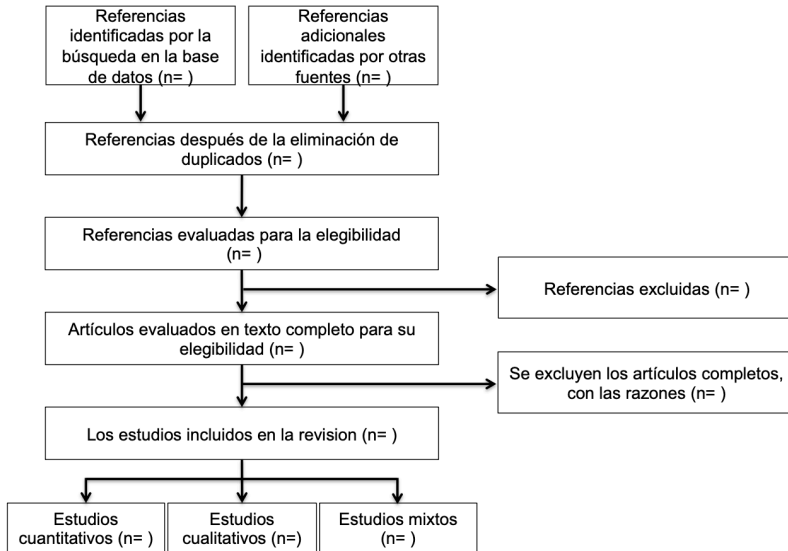


Figura 2. Diagrama de flujo

Paso 6: Evaluar la calidad de los estudios

En una revisión sistemática es importante evaluar la calidad de los estudios incluidos teniendo en cuenta los posibles sesgos, a fin de sacar conclusiones válidas sobre el estado de los conocimientos. La evaluación de la calidad de los estudios suele realizarse mediante instrumentos de evaluación crítica. Se dispone de una variedad de instrumentos para diversos métodos de investigación (Bai et al. 2012; Crowe y Sheppard 2011; Heyvaert et al. 2013). En el caso de la revisión sistemática mixta de FBR, la calidad de los estudios se evalúa utilizando una herramienta desarrollada específicamente para este tipo de revisión: la Herramienta de Evaluación de Métodos Mixtos (MMAT) (Pluye et al. 2011; Pluye et al. 2009; Souto et al. 2015; Pace et al. 2012). El MMAT permite la evaluación simultánea de cinco tipos de estudios: estudios cualitativos, ensayos controlados aleatorios, estudios analíticos cuantitativos, estudios descriptivos cuantitativos y estudios mixtos (Pluye et al. 2009). La última versión de esta herramienta incluye 25 ítems sobre la calidad metodológica de los estudios que se califican en una escala nominal (Sí/No/No sé) (Hong et al. 2018). La evaluación de la calidad se utiliza para

describir los artículos seleccionados e interpretar los datos de la síntesis. Esta herramienta puede ser descargada de este sitio web: <http://mixedmethodsappraisaltoolpublic.pbworks.com/>.

En la revisión sobre FBR, los resultados de esta evaluación no se utilizan para eliminar los estudios de baja calidad metodológica, ya que pueden seguir contribuyendo al conocimiento. Sus limitaciones se tienen en cuenta en los análisis de sensibilidad. Los análisis de sensibilidad exploran cómo se modifican los resultados obtenidos en la síntesis según la calidad de los estudios (Carroll y Booth 2015). Por ejemplo, estos análisis pueden utilizarse para comprobar si los resultados de la síntesis son similares cuando se eliminan los estudios de baja calidad. Por ejemplo, una similitud permite que los resultados se apoyen en un mayor número de estudios y hace que los resultados sean más precisos para los responsables de la toma de decisiones.

Paso 7: Extraer los datos

Este paso tiene por objeto reunir los datos (resultados de los estudios incluidos o datos primarios cuando se disponga de ellos) mediante un formulario de extracción que especifica todos los datos necesarios. Los datos extraídos se utilizan para describir los estudios incluidos (tabla de características), describir los resultados de los estudios incluidos (tabla de resultados) y sintetizar los datos.

Se requiere un marco conceptual o modelo teórico para orientar la selección de los datos que se han de extraer, a menos que el objetivo de la revisión sea construir dicho marco o modelo. Para nuestra revisión sistemática mixta sobre la FBR, desarrollamos un marco conceptual basado en la teoría de Rogers (2003) sobre la difusión de las innovaciones. Esta teoría sugiere que la interacción entre varias variables, como las características de los actores, la naturaleza del sistema social y la naturaleza y el uso de una nueva intervención, puede tener consecuencias diferentes a lo largo del tiempo. Así pues, de los artículos seleccionados extraemos los siguientes datos: autor, año, país, metodología y métodos, muestra (población y número), nombre de la intervención de la FBR, consecuencias imprevistas, naturaleza y uso de la FBR, características de los actores y naturaleza del sistema social. Si faltan datos de los artículos analizados, se contacta con el

primer autor. Dos personas hacen la extracción de forma independiente para asegurar una comprensión coherente de los datos y reducir las fuentes de sesgo y error.

Paso 8: Sintetizar los datos extraídos

La síntesis tiene como objetivo analizar y combinar los datos extraídos y evaluados en un todo coherente. En la revisión sistemática mixta, la síntesis es más compleja debido a la inclusión de datos tanto cualitativos como cuantitativos. Se pueden adoptar diferentes formas de sintetizar los datos cualitativos y cuantitativos. En el recuadro 2 se definen las construcciones necesarias para comprender las diferentes formas de sintetizar los datos.

Recuadro 2: Definiciones de los constructos utilizados en las especificaciones de síntesis

Datos: En una revisión, los datos son los resultados de los estudios incluidos (o los datos primarios cuando se dispone de ellos). Los datos pueden ser cualitativos (QL) (por ejemplo, los temas identificados en las entrevistas) o cuantitativos (QT) (por ejemplo, los porcentajes identificados en una encuesta).

Síntesis: En una revisión, el propósito de la síntesis es analizar los datos. Existen diversos métodos de síntesis que pueden definirse en dos categorías principales. La primera categoría incluye métodos de síntesis cuantitativa (QT) que analizan los datos sobre una variable específica utilizando métodos estadísticos básicos o más avanzados (por ejemplo, meta-análisis, síntesis bayesiana y meta-regresión). El segundo son los métodos de síntesis cualitativa (QL), que implican el análisis de datos para generar categorías, temas, construcciones o teorías como resultados (por ejemplo, meta-etnografía, síntesis temática y síntesis realista). En la actualidad, es posible identificar más de 20 métodos cualitativos y cuantitativos de síntesis (Barnett-Page y Thomas 2009; Dixon-Woods et al. 2005; Popay 2006; Sutton y Higgins 2008; Tricco et al 2016). En una revisión sistemática mixta, los métodos de síntesis utilizados en la misma revisión pueden ser sólo QL, sólo QT o una combinación de ambos (QL y QT).

Resultados: En una revisión, los resultados son los de la síntesis de datos. Según los métodos de síntesis utilizados, los resultados del examen pueden ser QL (por ejemplo, temas o marco conceptual), o QT (por ejemplo, estadísticas descriptivas o tamaño del efecto), o ambos (QL y QT).

Discusión: La discusión es una sección que tiene por objeto interpretar los resultados obtenidos y compararlos con las principales ideas publicadas en artículos de investigación o no investigación (por ejemplo, editoriales o comentarios influyentes o libros, guías de práctica, declaraciones de política pública, informes institucionales, etc.). Además, en esta sección se suelen exponer las limitaciones de la revisión y las recomendaciones.

Secuencia: La secuencia se define desde una perspectiva de desarrollo entre las fases de síntesis. En una revisión mixta, se presenta una secuencia cuando los resultados de una fase informan la síntesis de una fase subsiguiente.

Integración: La integración se define como la combinación de componentes cuantitativos y cualitativos dentro de un mismo estudio (Plano Clark e Ivankova 2015). En una revisión mixta, la integración puede producirse a nivel de los datos, los resultados de la síntesis, la interpretación o entre fases.

Diseño de la síntesis

Se realizó un análisis de 459 revisiones mixtas para examinar la forma en que se analizan los datos cualitativos y cuantitativos y para sugerir especificaciones de síntesis (Hong et al. 2017). Basándose en la literatura sobre métodos mixtos (Creswell et al., 2011), el análisis se realizó sobre los métodos de síntesis utilizados, la secuencia e integración de los datos cualitativos y cuantitativos. Se identificaron dos tipos principales de diseño para la síntesis: convergente y secuencial (figura 3). Estas especificaciones permiten comprender cómo se puede llevar a cabo la síntesis en una revisión sistemática mixta. Para cada una de estas especificaciones, es posible utilizar métodos de síntesis únicamente cualitativos, únicamente cuantitativos, o combinar métodos de síntesis cuantitativos y cualitativos (enfoque de síntesis mixta). A continuación, se presenta cada uno de estos diseños.

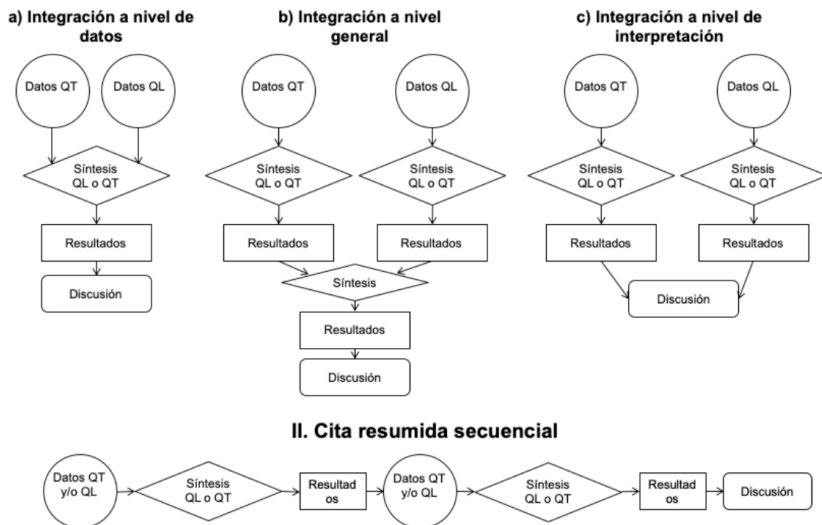


Figura 3. Diseños de síntesis en revisiones sistemáticas mixtas

Diseño de síntesis convergente

En un diseño de síntesis convergente, los estudios cualitativos y cuantitativos se analizan simultáneamente. Dentro de estas especificaciones, es posible identificar tres niveles de integración, ya sea a nivel de datos, de resultados de síntesis o de interpretación.

Diseño de síntesis convergentes con integración a nivel de datos

Todos los estudios se analizan utilizando el mismo método de síntesis, que puede ser únicamente cualitativo o cuantitativo (figura 3). Esto implica la transformación de los datos, es decir, los datos cuantitativos se transforman en categorías o temas (síntesis cualitativa), o los datos cualitativos se transforman en variables y valores numéricos (síntesis cuantitativa).

Por ejemplo, la pregunta de investigación para nuestro caso sería “¿Cuáles son las consecuencias no deseadas de la FRB en los países de ingresos bajos y medios?”. Para responder a esta pregunta se puede utilizar un método de síntesis cualitativa, por ejemplo, una síntesis temática que extraiga los temas principales de todos los estudios seleccionados (Thomas y Harden 2008). En este ejemplo, la transformación consiste en interpretar los datos de los estudios cuantitativos (frecuencias, tamaños de los efectos) cualitativamente (en términos de temas).

Si un investigador quisiera proporcionar estadísticas sobre las consecuencias imprevistas de la FBR, tendría que realizar, por ejemplo, un análisis cuantitativo del contenido, que implica asignar códigos (unas pocas variables clave y los valores numéricos asociados) a los resultados del estudio basados en un esquema de codificación validado y preciso (Neuendorf 2002). De esta manera, los códigos asignados a los estudios pueden ser analizados estadísticamente. La transformación se produce porque se asignarán valores numéricos a los datos de los estudios cualitativos.

Diseño de síntesis convergentes con integración a nivel de los resultados de la síntesis

Los datos cualitativos y cuantitativos se analizan y presentan por separado, pero se integran utilizando un método de síntesis diferente (figura 3). La integración puede consistir en comparar o yuxtaponer los resultados de las dos síntesis mediante tablas o matrices, o puede implicar un análisis secundario o de subgrupos.

En este diseño de síntesis, suele haber una pregunta de investigación general y subpreguntas. Por ejemplo, para responder a la pregunta “¿Cuáles son las consecuencias no deseadas de la FBR?” puede haber dos subpreguntas, “¿Cuáles son las consecuencias no deseadas en la satisfacción del cuidador?” y “¿Cuáles son las consecuencias no deseadas en la cantidad de atención médica proporcionada?”.

Para responder a estas preguntas, pueden ser necesarias tres síntesis que combinen métodos cualitativos y cuantitativos: a) una síntesis cualitativa de las consecuencias imprevistas en la satisfacción de los cuidadores, b) una síntesis meta-analítica cuantitativa de ensayos aleatorios de la cantidad de atención sanitaria, y c) una yuxtaposición de los resultados de las dos síntesis para examinar la relación entre la satisfacción de los cuidadores y la cantidad de atención prestada.

Diseño de síntesis convergentes con integración a nivel de la interpretación

Los datos cualitativos y cuantitativos se sintetizan y presentan por separado. La integración tiene lugar sólo cuando se interpretan los resultados en la discusión (Figura 3).

En este diseño de síntesis, se abordan en general cuestiones de investigación separadas y complementarias. Por ejemplo, “¿Es efectiva la FBR?”, “¿Cuál es la rentabilidad de la FBR?”, “¿Cuáles son las consecuencias no deseadas de la FBR?” y “¿Cuáles son las cuestiones éticas relacionadas con la FBR?”.

Para responder a estas preguntas de investigación complementarias, se necesitan métodos de síntesis tanto cualitativos como cuantitativos. Por ejemplo, las síntesis cuantitativas pueden responder a las dos primeras preguntas sobre la eficacia y la rentabilidad. Las síntesis cualitativas pueden analizar los estudios cualitativos sobre cuestiones éticas y consecuencias no deseadas. En la discusión, se resumirán las conclusiones generales sobre la eficacia, la rentabilidad, las consecuencias imprevistas y las cuestiones éticas. Además, se formularán recomendaciones sobre la evaluación y la aplicación de los marcos reguladores. Este tipo de especificación se observa, entre otros, en los informes de evaluación de la tecnología sanitaria, en los que la evaluación debe tener en cuenta diversas dimensiones (por ejemplo, la eficacia, la eficiencia, la seguridad, las repercusiones éticas, psicológicas, organizativas y económicas) (Busse et al. 2002).

Diseño de síntesis secuencial

Una cita de resumen secuencial implica al menos dos fases de síntesis dependientes. Los resultados de una primera fase en la que se sintetizan los datos cualitativos y/o cuantitativos informan una segunda fase de síntesis (figura 3).

En un diseño de síntesis secuencial, puede haber una pregunta de investigación o preguntas dependientes. Por ejemplo, las preguntas pueden ser “¿Cuáles son las consecuencias no deseadas de la FBR?” y “¿Cuál es la importancia relativa de estas consecuencias no deseadas?”. Para responder a estas preguntas, se requieren dos fases. Una síntesis cualitativa inicial (Fase 1) desarrollará una lista de consecuencias no deseadas. A partir de esta lista (resultados de la fase 1), se analizan los estudios cuantitativos mediante un método de síntesis cuantitativa (fase 2) para determinar la importancia relativa de las consecuencias imprevistas. De este modo se cuantificará la importancia de las consecuencias imprevistas y se identificarán las lagunas de conocimiento, es decir, las consecuencias que no se han estudiado en los estudios cuantitativos.

En los últimos ejemplos presentados se utiliza un enfoque de síntesis mixto (es decir, el uso de métodos de síntesis tanto cualitativos como cuantitativos en el mismo examen). En función de las preguntas de investigación, es posible utilizar sólo métodos de síntesis cualitativos o sólo

métodos de síntesis cuantitativos. Por ejemplo, los métodos de síntesis cualitativa pueden utilizarse en ambas fases de un diseño de síntesis secuencial. En la primera fase, se utiliza un método de síntesis inductiva para definir los temas principales. Luego, en una segunda fase, los estudios se analizan de forma deductiva en base a los temas identificados.

Análisis reflexivo

La robustez de una revisión sistemática es su enfoque explícito, transparente y reproducible. Se han elaborado varias directrices para la presentación de informes sobre una revisión sistemática, la más conocida de las cuales es PRISMA (Moher et al. 2009).

Varios sesgos de investigación pueden minimizarse siguiendo este enfoque. Un primer sesgo potencial es el de detección, que está relacionado con la búsqueda de literatura. Con el fin de asegurar que se identifiquen todos los estudios pertinentes a la pregunta de investigación (búsqueda exhaustiva), es importante diversificar las fuentes documentales, utilizar varias bases de datos y elaborar estrategias rigurosas de búsqueda de documentos con bibliotecarios especializados. Un segundo sesgo es el de selección. Es importante evitar la selección arbitraria de los estudios definiendo criterios claros y hacer participar a dos personas en la selección. Un tercer sesgo que debe considerarse es el de la divulgación de los resultados; el más conocido es el de publicación. De hecho, es más probable que se publique un estudio con resultados positivos y más rápidamente (Dubben y Beck-Bornholdt 2005). Esto lleva a una sobrerrepresentación de los estudios primarios que demuestran efectos positivos y puede afectar a las conclusiones de la revisión. Se recomienda diversificar las fuentes de datos y los tipos de documentos que se incluirán. Por ejemplo, realizar una búsqueda de la literatura gris para incluir informes científicos y tesis inéditas. Además, es posible que exista un sesgo de publicación selectiva (Martin, Renaud y Dagenais 2013), es decir, los autores informan de más consecuencias intencionadas y esperadas de las intervenciones que de consecuencias no intencionadas e inesperadas, lo que constituye un reto en nuestro ejemplo sobre la FBR. Por último, el sesgo relacionado con el evaluador y su interpretación de los estudios puede reducirse al mínimo si se involucra a por lo menos dos personas en la evaluación y síntesis de

los estudios. En una revisión sistemática mixta, se recomienda contar con un equipo con conocimientos complementarios en metodologías y técnicas de investigación cualitativas y cuantitativas, para facilitar la síntesis y la evaluación de la calidad de los estudios (Pluye y Hong 2014, Pluye et al., 2016).

Las desventajas de un enfoque sistemático son principalmente que requiere tiempo y recursos. Según el objetivo de la revisión, el número de personas que participan, el número de estudios que se analizan y la síntesis que se realiza, la duración de una revisión sistemática puede variar entre 6 y 24 meses (Khangura y otros, 2012). Así pues, es importante garantizar la disponibilidad de recursos y justificar la elección de llevar a cabo una revisión sistemática mixta, por ejemplo, queriendo aprovechar la complementariedad de los datos cualitativos y cuantitativos para explicar un fenómeno complejo y multidimensional.

Los exámenes sistemáticos pueden realizarse utilizando programas informáticos básicos (como Word o Excel) o especializados, según los métodos de síntesis elegidos (como QDA Miner, NVivo, SPSS, RevMan). Además, se han desarrollado varios paquetes de software especializados para revisiones sistemáticas como DistillerSR, Covidence y EPPI Reviewer. Estos programas informáticos están concebidos para facilitar la gestión de los datos y pueden utilizarse en las diferentes etapas de la revisión sistemática, desde la selección de los artículos hasta la extracción y el análisis de los datos. Exceptuando algunos programas informáticos disponibles gratuitamente en línea (por ejemplo, QDA Miner Lite), generalmente se exigen cuotas mensuales o anuales. La utilización de estos derechos depende de los recursos disponibles y de la complejidad y el tamaño de la revisión.

No siempre es posible realizar una revisión sistemática. En una revisión para una tesis doctoral o una revisión rápida para un gerente, es raro que se disponga de recursos y tiempo suficientes para realizar una revisión sistemática. Como resultado, un estudiante de posgrado o un profesional puede comenzar por realizar una revisión de las revisiones existentes. Si no hay revisiones, se puede hacer una búsqueda de estudios pertinentes con la ayuda de un bibliotecario en una base de datos bibliográfica para obtener una muestra de estudios. Luego, la selección, la evaluación de la calidad de los estudios incluidos y la recuperación de los datos son realizadas por un solo investigador. Seguir estos pasos ahorra mucho tiempo en comparación

con una búsqueda bibliográfica no estructurada y produce un capítulo o informe de tesis convincente (es decir, un examen no exhaustivo pero transparente y reproducible).

Una de las dificultades encontradas en la revisión sistemática mixta es la integración, es decir, la combinación o conexión de componentes cualitativos y cuantitativos. Por ejemplo, en una revisión se podrían examinar cuestiones como la eficacia de la FBR, su rentabilidad y sus consecuencias no deseadas. Estas cuestiones implican diferentes tipos de estudios y pueden ser objeto de revisiones separadas. Lo que caracteriza una revisión mixta es que todas las conclusiones se integran para proporcionar una comprensión más profunda de la FBR y hacer recomendaciones más específicas. Sin esta integración, una revisión no podría considerarse como mixta, sino más bien como un estudio que incluye varias revisiones independientes. Por lo tanto, es esencial describir cómo se integran los datos. Los diseños de síntesis presentados ofrecen soluciones prácticas para integrar los datos. Además, en el capítulo 7 se presentan diversas estrategias de integración para los estudios primarios. Estas estrategias pueden explorarse para su uso en revisiones sistemáticas. En los artículos que informen sobre revisiones sistemáticas mixtas, la discusión debería, como mínimo, resumir la forma en que los resultados de los datos cualitativos y cuantitativos se comparan o complementan entre sí. Debería reflejar el valor añadido de esta integración y sus limitaciones.

La revisión sistemática mixta es una forma de síntesis de los conocimientos en evolución. El número de artículos que reportan revisiones sistemáticas mixtas ha aumentado exponencialmente desde 2010. Los beneficios de una revisión sistemática mixta pueden ser considerables, en particular para responder a preguntas complejas, aprovechar los datos cuantitativos y cualitativos complementarios y tener una comprensión profunda y completa de un fenómeno. Las investigaciones sobre revisiones sistemáticas mixtas son todavía jóvenes y los próximos años prometen desarrollos conceptuales y metodológicos.

Referencias clave

Heyvaert, M., Hannes, K. et Onghena, P. (2016). *Using mixed methods research synthesis for literature reviews: The mixed methods research synthesis approach* (vol. 4). Thousand Oaks, CA: Sage Publications.

Libro metodológico sobre la revisión mixta, escrito por investigadores en ciencias sociales, psicología y pedagogía de la Universidad de Lovaina, Bélgica. Se detallan los diferentes pasos para realizar una revisión mixta, comenzando por la redacción de un protocolo, seguido de la búsqueda de documentos pertinentes, la evaluación de la calidad de los estudios, la extracción y síntesis de datos y la publicación de una revisión mixta.

Martin, V., Renaud, J. et Dagenais, P. (2013). *Les normes de production des revues systématiques : guide méthodologique*.

https://www.inesss.qc.ca/fileadmin/doc/INESSS/DocuMetho/INESSS_Normes_production_revues_systematiques.pdf

Guía metodológica, escrita en francés, que propone 33 normas para la elaboración de revisiones sistemáticas. La presente guía no trata específicamente de las revisiones mixtas, pero ofrece una excelente descripción de los pasos a seguir para realizar una revisión sistemática. Fue producida por investigadores del Instituto Nacional de Excelencia en Salud y Servicios Sociales (INESSS), organización francófona cuya misión es promover la excelencia clínica y la utilización eficaz de los recursos en el sector de la salud y los servicios sociales en Quebec, Canadá.

Suri, H. (2014). *Towards methodologically inclusive research synthesis*. Abingdon, UK : Routledge.

Libro metodológico sobre revisiones mixtas escrito por una investigadora en educación de la Universidad de Deakin (Australia). El libro tiene 11 capítulos, el primero de los cuales trata de los avances metodológicos en la síntesis, los principios de la síntesis y las orientaciones epistemológicas. En los capítulos finales se explica cómo realizar una revisión mixta, es decir, la formulación de un objetivo, la búsqueda de bibliografía, la evaluación, la síntesis y la interpretación de los datos.

Collaboration Cochrane : <http://www.cochrane.org/>

La Colaboración Cochrane es una organización internacional sin ánimo de lucro dedicada a promover la práctica de la salud basada en la evidencia.

Tiene más de 30.000 colaboradores en más de 120 países. Ofrece apoyo metodológico y capacitación en materia de revisión sistemática. El Centro Cochrane Iberoamericano (CCIb) es uno de los 14 centros internacionales de Cochrane (<https://es.cochrane.org/>). Su ámbito de responsabilidad comprende España, Portugal, Andorra y todos los países hispanohablantes de América Latina. Además, dispone de un libro de referencia metodológica sobre revisiones sistemáticas de forma gratuita: *Cochrane Handbook for Systematic Reviews of Interventions* (<http://handbook.cochrane.org/>).

Collaboration Campbell : <http://www.campbellcollaboration.org/>

La Colaboración Campbell es una organización similar a la Colaboración Cochrane, especializada en las ciencias sociales, específicamente en las áreas de justicia, educación, desarrollo internacional y bienestar social.

Toolkit for Mixed Studies Reviews :
<http://toolkit4mixedstudiesreviews.pbworks.com>

Este sitio web sobre la revisión mixta es producido por investigadores de la Universidad McGill en Canadá. Se pretende que sea un sitio web de colaboración en el que se invite a todos los interesados en este método a participar en su mantenimiento. Los pasos de la revisión sistemática mixta se describen en profundidad.

Referencias

- Bai, A., Shukla, V. K., Bak, G. et Wells, G. (2012). *Quality assessment tools project report*. Ottawa : Canadian Agency for Drugs and Technologies in Health.
- Barnett-Page, E. et Thomas, J. (2009). *Methods for the synthesis of qualitative research: A critical review*. *BMC Medical Research Methodology*, 9(1), 59.
doi: 10.1186/1471-2288-9-59
- Bramer, W.M., Giustini, D., de Jonge, G. B., Holland, L. et Bekhuis, T. (2016). *De-duplication of database search results for systematic reviews in EndNote*. *Journal of the Medical Library Association : JMLA*, 104(3), 240-243.
doi: 10.3163/1536-5050.104.3.014
- Busse, R., Orvain, J., Velasco, M., Perleth, M., Drummond, M., Jørgensen, T., ... et Wild, C. (2002). *Best practice in undertaking and reporting health*

- technology assessments. *International Journal of Technology Assessment in Health Care*, 18(2), 361-422.
doi: 10.1017/S0266462302000284
- Carroll, C. et Booth, A. (2015). Quality assessment of qualitative evidence for systematic review and synthesis: Is it meaningful, and if so, how should it be performed? *Research Synthesis Methods*, 6(2), 149-154.
doi: 10.1002/jrsm.1128
- Chalmers, I., Hedges, L. V. et Cooper, H. (2002). A brief history of research synthesis. *Evaluation and the Health Professions*, 25(1), 12-37.
doi: 10.1177/0163278702025001003
- Cochrane Effective Practice and Organisation of Care. 2012. LMIC Filters.
<http://epoc.cochrane.org/lmic-filters>
- Cooper, H.M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1), 104-126.
- Creswell, J. W., Klassen, A. C., Clark, V. L. P. et Smith, K. C. (2011). *Best Practices for Mixed Methods Research in the Health Sciences*. Site du National Institutes of Health, Office of Behavioral and Social Sciences Research.
http://obssr.od.nih.gov/mixed_methods_research
- Crowe, M. et Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. *Journal of Clinical Epidemiology*, 64(1), 79-89.
doi: 10.1016/j.jclinepi.2010.02.008
- Dixon-Woods, M., Agarwal, S., Jones, D., Young, B. et Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. *Journal of Health Services Research and Policy*, 10(1), 45 – 53.
doi: 10.1258/1355819052801804
- Dubben, H.-H. et Beck-Bornholdt, H.-P. (2005). Systematic review of publication bias in studies on publication bias. *British Medical Journal*, 331(7514), 433-434.
- El Sherif, R., Pluye, P., Gore, G., Granikov, V. et Hong, Q. N. (2016). Performance of a mixed filter to identify relevant studies for mixed studies reviews. *Journal of the Medical Library Association : JMLA*, 104(1), 47-51.
- Fritsche, G.B., Soeters, R. et Meessen, B. (2014). *Boîte à outils: financement basé sur la performance*. Montreuil, FRA : Pearson France.

- Gedda, M. (2015). Traduction française des lignes directrices PRISMA pour l'écriture et la lecture des revues systématiques et des méta-analyses. *Kinésithérapie, la Revue*, 15 (157), 39-44.
- Gorter, A. C. et Meessen, B. (2013). *Evidence review, results-based financing of maternal and newborn health care in low-and lower-middle-income countries*. Germany: Study commissioned and funded by the German Federal Ministry for Economic Cooperation and Development through the sector project PROFILE at Deutsche Gesellschaft für Internationale Zusammenarbeit, Bonn et Eschborn.
- Grant, M.J. et Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91-108.
doi: 10.1111/j.1471-1842.2009.00848.x
- Harden, A. (2010). Mixed-methods systematic reviews: Integrating quantitative and qualitative findings. *Focus*, (25), 1-7.
- Heyvaert, M., Hannes, K., Maes, B. et Onghena, P. (2013). Critical appraisal of mixed methods studies. *Journal of Mixed Methods Research*, 7(4), 302-327.
doi: 10.1177/1558689813479449
- Hong, Q. N., Pluye, P., Bujold, M. et Wassef, M. (2017). Convergent and sequential synthesis designs: Implications for conducting and reporting systematic reviews of qualitative and quantitative evidence. *Systematic Reviews*, 6(1), 61.
doi: 10.1186/s13643-017-0454-2
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M.-C. et Vedel, I. 2018. « Mixed Methods Appraisal Tool (MMAT), version 2018 ». Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.
- Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J. et Moher, D. (2012). Evidence summaries: The evolution of a rapid review approach. *Systematic Reviews*, 1(1), 1-9.
doi: 10.1186/2046-4053-1-10

- Langlois, A., Nie, J. Y., Thomas, J., Hong, Q. N. et Pluye, P. 2018. « Discriminating Between Empirical Studies and Nonempirical Works Using Automated Text Classification ». *Research Synthesis Methods*.
doi: 10.1002/jrsm.1317
- Martin, V., Renaud, J. et Dagenais, P. (2013). *Les normes de production des revues systématiques : guide méthodologique*. Montréal, QC : Institut national d'excellence en santé et en services sociaux (INESSS).
- Mays, N., Pope, C. et Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy*, 10(Suppl. 1), 6 – 20.
doi: 10.1258/1355819054308576
- Moher, D., Liberati, A., Tetzlaff, J. et Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Public Library of Science : Medicine*, 6(7), :e1000097.
doi: 10.1371/journal.pmed.1000097
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. et Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49(1), 47-53.
doi: 10.1016/j.ijnurstu.2011.07.002
- Paré, G., Trudel, M.-C., Jaana, M. et Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. *Information et Management*, 52(2), 183-199.
- Clark, V. L. P. et Ivankova, N. V (dir.). (2015). *Mixed methods research: A guide to the field* (vol. 3). Thousand Oaks, CA: Sage Publications.
- Pluye, P., Gagnon, M. P., Griffiths, F. et Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, 46(4), 529-46.
doi: 10.1016/j.ijnurstu.2009.01.009

- Pluye, P. et Hong, Q. N. (2014). Combining the power of stories and the power of numbers: Mixed methods research and mixed studies reviews. *Annual Review of Public Health*, 35, 29-45.
doi: 10.1146/annurev-publhealth-032013-182440
- Pluye, P., Hong, Q. N., Bush, P. L. et Vedel, I. (2016). Opening-up the definition of systematic literature review: The plurality of worldviews, methodologies and methods for reviews and syntheses. *Journal of Clinical Epidemiology*, 73, 2-5.
doi:dx.doi.org/10.1016/j.jclinepi.2015.08.033
- Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., ... et Rousseau, M.C. (2011). Proposal: A mixed methods appraisal tool for systematic mixed studies reviews. Montréal : McGill University, 1-8.
<http://mixedmethodsappraisaltoolpublic.pbworks.com/>
- Popay, J (dir.). (2006). *Moving beyond effectiveness in evidence synthesis: Methodological issues in the synthesis of diverse sources of evidence*. Londres, UK : National Institute for Health and Care Excellence (NICE).
- Rathbone, J., Carter, M., Hoffmann, T. et Glasziou, P. (2015). Better duplicate detection for systematic reviewers: Evaluation of Systematic Review Assistant-Deduplication Module. *Systematic Reviews*, 4(1), 6.
- Rogers, E. M. (2003). *Diffusion of Innovations* (5^e éd.). New York : Free Press.
- Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I. et Pluye, P. (2015). Systematic mixed studies reviews: Updating results on the reliability and efficiency of the Mixed Methods Appraisal Tool. *International Journal of Nursing Studies*, 52(1), 500-501.
doi: 10.1016/j.ijnurstu.2014.08.010
- Sutton, A.J. et Higgins, J. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5), 625-650.
doi: 10.1002/sim.2934
- Thomas, J. et Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, 8, 45.
doi: 10.1186/1471-2288-8-45
- Tricco, A. C., Soobiah, C., Antony, J., Cogo, E., MacDonald, H., Lillie, E., ... et Perrier, L. (2016). A scoping review identifies multiple emerging knowledge

synthesis methods, but few studies operationalize the method. *Journal of Clinical Epidemiology*, 73, 19-28.

doi: 10.1016/j.jclinepi.2015.08.030

Whittemore, R. et Knafl, K. (2005). The integrative review: Updated methodology. *Journal of Advanced Nursing*, 52(5), 546-553.

doi: 10.1111/j.1365-2648.2005.03621.x

Witter, S., Fretheim, A., Kessy, F. L. et Lindahl, A. K. (2012). Paying for performance to improve the delivery of health interventions in low-and middle-income countries. *Cochrane Database of Systematic Review*, (2), CD007899.

Resumen / Résumé / Abstract

Las revisiones sistemáticas de estudios mixtos son revisiones de la literatura que utilizan un enfoque sistemático para combinar estudios cuantitativos, cualitativos y de métodos mixtos. Su popularidad va en aumento debido a su potencial para proporcionar una comprensión profunda de las intervenciones y los problemas de salud complejos. En este capítulo se presentan ocho pasos de revisiones sistemáticas de estudios mixtos: (a) formular una pregunta de revisión, (b) definir los criterios de elegibilidad, (c) identificar las fuentes de información, (d) seleccionar los estudios pertinentes, (e) evaluar la calidad de los estudios, (f) extraer los datos y (g) sintetizar los datos extraídos de los estudios incluidos. Estos pasos se ilustran con un ejemplo de revisión sistemática de estudios mixtos sobre las consecuencias no deseadas de la intervención financiera basada en el desempeño. Además, para proporcionar orientación sobre cómo combinar la evidencia cualitativa y cuantitativa, se presentan cuatro diseños de síntesis: a) diseño de síntesis convergente basado en datos, b) diseño de síntesis convergente basado en resultados, c) diseño de síntesis convergente basado en resultados paralelos, y d) diseño de síntesis secuencial. Este capítulo termina con algunos de los desafíos que se pueden encontrar al realizar una revisión sistemática de estudios mixtos.

Les revues systématiques mixtes sont des revues de littérature utilisant une approche systématique pour combiner des études quantitatives,

qualitatives et mixtes. Elles sont de plus en plus populaires, car elles permettent une compréhension approfondie des phénomènes et interventions complexes. Dans ce chapitre, huit étapes de la revue systématique mixte sont présentées : (a) formulation d'une question de recherche, (b) définition des critères d'éligibilité, (c) identification des sources documentaires, (d) (élaboration de stratégie de recherche documentaire, (e) sélection des études pertinentes, (f) évaluation de la qualité des études retenues, (g) extraction des données des études retenues et (h) synthèse des données extraites. Ces étapes sont illustrées à partir d'un exemple tiré d'une revue systématique mixte sur les conséquences non intentionnelles d'une intervention reposant sur le financement basé sur les résultats. De plus, afin de guider l'intégration des données quantitatives et qualitatives, quatre devis de synthèse sont présentés : (a) devis de synthèse convergent avec intégration au niveau des données, (b) devis de synthèse convergent avec intégration au niveau des résultats de synthèse, (c) devis de synthèse convergent avec intégration au niveau de l'interprétation et (d) devis de synthèse séquentiel. Enfin, quelques défis des revues systématiques mixtes sont présentés dans l'analyse réflexive.

Systematic mixed studies reviews are literature reviews using a systematic approach to combine quantitative, qualitative and mixed methods studies. They are growing in popularity owing to their potential to provide a deep understanding of complex health interventions and problems. In this chapter, eight steps of systematic mixed studies reviews are presented: (a) formulate a review question, (b) define eligibility criteria, (c) identify sources of information, (d) select relevant studies, (e) appraise the quality of studies, (f) extract data and (g) synthesize extracted data from included studies. These steps are illustrated using an example of systematic mixed studies review on unintended consequences of performance-based financing intervention. Also, to provide guidance on how to combine qualitative and quantitative evidence, four synthesis designs are presented: (a) data-based convergent synthesis design, (b) results-based convergent synthesis design, (c) parallel-results convergent synthesis design, and (d) sequential synthesis design. This chapter ends with some challenges that can be encountered when conducting a systematic mixed studies review.

Quan Nha Hong es candidata al doctorado en el Departamento de Medicina Familiar de la Universidad McGill y becaria de los Institutos Canadienses de Investigación en Salud (CIHR). Tiene formación como terapeuta ocupacional. Tiene una Maestría en Ciencias Clínicas de la Université de Sherbrooke y una Maestría en Evaluación y Gestión de Tecnologías de la Salud de la Université de Montréal. Le interesan los exámenes sistemáticos, los métodos de síntesis, los métodos mixtos y la evaluación de la calidad metodológica de los estudios. Su proyecto de doctorado se centra en la actualización de un instrumento de evaluación de la calidad de los estudios cuantitativos, cualitativos y de métodos mixtos, el Mixed Methods Appraisal Tool (MMAT).

Anne-Marie Turcotte-Tremblay es candidata al doctorado en salud pública con especialización en organización de la atención de la salud en la École de santé publique de l'Université de Montréal (ESPUM). También está afiliada al Instituto de Investigación en Salud Pública de la Universidad de Montreal (IRSPUM). Su tesis doctoral se centra en las consecuencias imprevistas de la financiación basada en los resultados combinada con medidas de protección social en Burkina Faso. Anne-Marie es becaria del Instituto Canadiense de Investigación de la Salud (CIHR). Ha coordinado varios proyectos internacionales de investigación sobre la salud de la población. Anne-Marie está interesada en la evaluación de las intervenciones para reducir las desigualdades sociales en materia de salud.

Pierre Pluye es profesor titular del Departamento de Medicina Familiar de la Universidad McGill, investigador principal del Fondo de Investigación de Quebec Salud (FRQS) y director del componente de Desarrollo Metodológico de la Unidad SOUTIEN SRAP de Quebec. Es miembro asociado de la Escuela de Ciencias de la Información de la Universidad McGill. Es especialista en medicina familiar y salud pública. Es doctor en promoción de la salud (Universidad de Montreal). Ha recibido premios a su carrera de los Institutos Canadienses de Investigación de la Salud (CIHR) y del FRQS. En 2017 fue elegido miembro de la Academia Canadiense de Ciencias de la Salud y recibió el premio “Investigador del Año” del Colegio de Médicos de Familia del Canadá. Tiene experiencia en métodos mixtos. Su investigación estudia los efectos de la información de los recursos electrónicos en los profesionales, los pacientes y el público.

Cita

Quan Nha Hong, Anne-Marie Turcotte-Tremblay y Pierre Pluye (2020). Revisiones sistemáticas mixtas. Un ejemplo sobre la financiación basada en los resultados. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 165-194. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

7. Integración en métodos mixtos

Marco conceptual para la integración de fases, resultados y datos cualitativos y cuantitativos

PIERRE PLUYE

La integración en métodos mixtos

Definición del método

Los métodos mixtos integran métodos cualitativos y cuantitativos. Pueden utilizarse para responder a preguntas complejas de investigación y evaluación.

Fortaleza del método

Proponemos tres tipos de integración y nueve estrategias prácticas, y todas las combinaciones posibles de estrategias para integrar fases, resultados y datos cualitativos y cuantitativos.

Desafíos del método

Nuestra propuesta puede cubrir la necesidad de a) investigadores para planificar mejor, realizar y publicar estudios utilizando métodos mixtos, y b) del profesorado para formar estudiantes en métodos mixtos.

Los métodos mixtos combinan métodos cualitativos (QUAL) y cuantitativos (QUAN) en la evaluación de programas, la investigación primaria y la revisión de la literatura (Creswell y Plano Clark 2011; Johnson, Onewuegbuzie y Turner 2007; Pluye y Hong 2014; Tashakkori y Teddlie 2010). Se utilizan cada vez más en las ciencias sociales y de la salud. A lo largo de los años, se han propuesto varias estrategias para integrar las fases, resultados y datos de QUAL y QUAN, pero rara vez se han conceptualizado y nunca se han probado (Greene 2008).

Un reto para cualquier investigador que utilice métodos mixtos es planificar, realizar e informar de manera sencilla y clara cuáles son las estrategias de análisis de los métodos mixtos específicos y las combinaciones de estas estrategias. En el presente capítulo se ofrece un marco conceptual de las principales combinaciones de estrategias que se utilizan en la investigación primaria con métodos mixtos. En el próximo capítulo, presentaremos una vigilancia colaborativa de las estrategias de

análisis de métodos mixtos, que ha puesto a prueba esta propuesta y proporcionaremos ilustraciones basadas en estudios de salud global que utilizan métodos mixtos.

Los métodos mixtos son cualquier combinación de metodologías y métodos (incluyendo diseños de investigación y técnicas de recolección y análisis de datos QUAL y QUAN) cuando estas combinaciones satisfacen tres condiciones (Creswell y Plano Clark 2011) : a) se combinan al menos un método QUAL y un método QUAN (por ejemplo, etnografía y experimentación), b) cada método se utiliza rigurosamente con arreglo a los criterios generalmente aceptados en la metodología o tradición de investigación invocada, y c) la combinación de métodos se logra como mínimo mediante un diseño de métodos mixtos (previstos o emergentes) y una integración de las fases, resultados y datos QUAL y QUAN. Obsérvese que el equipo de investigación también puede combinar (pero no necesariamente) los métodos QUAL y QUAN, a nivel de la recopilación de datos (instrumentación mixta), la revisión de la literatura (revisión mixta) y la interpretación científica (epistemología, ontología, teleología y metodología).

Por el contrario, no se consideran métodos mixtos: a) Un método QUAN con una recopilación o análisis de información cualitativa que no sea “datos de investigación”, ya que no se refiere a un diseño y un método QUAL; b) Un método QUAL con una recopilación o análisis de información cuantitativa que no sea “datos de investigación”, ya que no se refiere a una cita y un método QUAN; c) combinaciones de métodos de QUAN; d) combinaciones de métodos de QUAL, y e) yuxtaposiciones de métodos de QUAL y QUAN sin integración en cuanto a enfoques, diseños, técnicas, fases, resultados y datos de QUAL y QUAN.

En este capítulo, presentaremos los principales elementos de los métodos mixtos (contexto) que sustentan el marco conceptual y propondremos tres tipos de integración (mapa), nueve estrategias específicas de los métodos (aplicación) y siete combinaciones de estrategias (análisis reflexivo) para integrar las fases cualitativas y cuantitativas, los resultados y los datos. En el próximo capítulo, presentaremos ejemplos de investigación en el campo de la salud global.

Contexto: Pluralismo, diseño, datos y calidad

Pluralismo de los métodos mixtos

Ya en 2003, Teddlie y Tashakkori afirmaron la coexistencia de diferentes interpretaciones del mundo y la ciencia por parte de los investigadores en métodos mixtos (Teddlie y Tashakkor, 2003). Este pluralismo es ilustrado por numerosas publicaciones (Niglas 2010). En este capítulo, los métodos mixtos se basan en la interpretación que hacen los investigadores de la ciencia (explícita o tácitamente) en términos de epistemología, ontología, teleología y metodología (Gendron 2001; Ridde y Dagenais 2012). Cuando un equipo incluye investigadores de métodos mixtos con diferentes interpretaciones de la ciencia, la combinación de métodos requiere discusiones de epistemología, ontología, teleología y metodología. Por ejemplo, el equipo puede combinar de manera transparente los enfoques o tradiciones científicas que reconocen los métodos QUAL y QUAN, como el pospositivismo de Campbell, el construccionismo social de Hacking, el pragmatismo, el realismo crítico y las teorías críticas, entre otros (Campbell 1988; De Waal 2005; Hacking 1999; Sayer 2000; Tyson 2014).

Principales diseños en métodos mixtos

Los investigadores pueden basarse en las recomendaciones existentes para planificar las investigaciones que utilizan métodos mixtos (Creswell y Plano Clark 2011). Una clasificación común de métodos mixtos se basa en dos tipos de diseño de investigación: secuenciales y convergentes. También hay tres variantes principales de estos diseños básicos: multietápico, multinivel y multietápico/multinivel.

Los diseños de investigación secuencial utilizan un método QUAL seguido de un método QUAN (por ejemplo, los resultados QUAN se movilizan para generalizar estadísticamente ciertos resultados QUAL), o un método QUAN seguido de un método QUAL (por ejemplo, los resultados QUAL se movilizan para interpretar ciertos resultados QUAN). En cualquier diseño secuencial, los resultados de la fase 1 informan a la fase 2. En este capítulo, consideramos que un proyecto de investigación puede conceptualizarse como un proceso organizacional (por ejemplo, un proyecto colectivo en

el que participan investigadores de QUAL y QUAN). La literatura sobre organizaciones (en gestión y administración de proyectos) proporciona entonces una definición útil para comprender mejor estas secuencias. Inspirándonos en Van de Ven (1992), proponemos que se defina una secuencia a) desde una perspectiva de desarrollo (la fase 2 se basa en los resultados de la fase 1), b) como un cambio en la orientación del proyecto a lo largo del tiempo (los resultados de una primera recogida/análisis de datos [fase 1] informan una segunda recogida/análisis de datos [fase 2] que es diferente de la primera en cuanto a metodología y método) y c) como una transición cognitiva de los investigadores al pasar a una nueva fase (de una metodología y métodos QUAL a una QUAN, o de QUAN a la metodología y métodos QUAL) (Van de Ven 1992).

Por el contrario, los diseños convergentes combinan los métodos QUAL y QUAN en la etapa de planificación de la investigación o en el momento de la recopilación y análisis de los datos de QUAN y QUAL. En un diseño convergente, los métodos QUAL y QUAN son a menudo (pero no necesariamente) concurrentes. Una vez más, la literatura sobre organizaciones y procesos colectivos de adopción de decisiones (por ejemplo, decisiones adoptadas por un equipo de investigadores utilizando métodos mixtos) ofrece definiciones de convergencia. Inspirados por Langley y colegas, proponemos que la convergencia se defina como un proceso de mejora progresiva, sucesiva y continua en la recogida y análisis de datos QUAL y QUAN (convergencia de datos) y en la interpretación de los resultados (convergencia de resultados). Los investigadores trabajan de manera prospectiva y no lineal, guiados por una representación cognitiva de los datos adicionales, de las bases de datos, del análisis de datos, o de los resultados a crear (Langley, Mintzberg, Pitcher, Posada y Saint-Macary 1995).

Las variantes de estos diseños, simplemente implican múltiples fases/etapas o niveles de recogida y análisis de datos. Por un lado, el diseño multietápico incluye tres o más fases secuenciales (por ejemplo, QUAL, luego QUAN y luego QUAL) [Lisle 2013]. La viabilidad de este diseño puede ser problemática, ya que cada fase depende de la finalización oportuna de la anterior. Por otra parte, el diseño multinivel incluye dos o más niveles de análisis (por ejemplo, QUAN a nivel individual y QUAL a nivel de organización) (Dagenais, Nault-Brière, Dupont y Dutil 2008). Este diseño se

basa en la convergencia de los resultados de los análisis realizados en cada nivel. Youngs y Piggot-Irvine combinaron las dos variantes y utilizaron una estimación multietápica y multinivel (Youngs y Piggot-Irvine 2011).

En un diseño mixto – MM (secuencial, convergente, multietápico, multinivel o multietápico/multinivel), los métodos QUAL que se combinan más frecuentemente con QUAN son: la investigación cualitativa descriptiva o interpretativa, el estudio de casos exploratorios, la etnografía, la teoría fundamentada, la fenomenografía y las historias de vida o biografías (Schwandt 2007). A su vez, los métodos QUAN que se combinan más frecuentemente con QUAL son: los estudios descriptivos (por ejemplo, encuestas de prevalencia o incidencia), estudios no aleatorios (por ejemplo, encuesta analítica transversal, cohorte, comparación de casos o cuasi-experimental) y experimentos aleatorizados y controlados o pragmáticos (Porta 2008). Cabe destacar que los diseños de “estudio de casos” y de “teoría fundamentada”, pueden ser en sí mismos QUAL y QUAN; por lo tanto, cada una de ellos puede ser suficiente (no tiene por qué combinarse necesariamente con otro diseño) para planificar un diseño de métodos mixtos.

Técnicas de recogida de datos QUAL y QUAN en métodos mixtos

En estos diseños, cualquier combinación de técnicas de recogida de datos QUAL y QUAN es posible en métodos mixtos. Los investigadores de MM son creativos y utilizan diversas posibilidades de recogida de datos experimentales, observacionales o datos simulados. No obstante, predominan las técnicas tradicionales de recogida. El cuestionario estructurado tiende a predominar en el lado de las técnicas QUAN, y la entrevista abierta o semiestructurada tiende a predominar en el lado de las técnicas QUAL (Bryman 2006). Cabe destacar, el cuestionario estructurado que combina preguntas cerradas (medidas estándar cuya validez y fiabilidad se ha comprobado) y preguntas abiertas, pues se trata de una pregunta recurrente durante la formación en MM. Estas preguntas abiertas pueden ser parte de los métodos QUAL o QUAN, dependiendo de la forma en que se diseñen y utilicen. Por ejemplo, las respuestas a las preguntas abiertas constituyen datos QUAL cuando se obtienen mediante una metodología QUAL rigurosa y un proceso de investigación científica (explícito, transparente y reproducible) que produce resultados QUAL plausibles

(creíbles, contextuales, que pueden confirmarse y son transferibles). Los investigadores conocen a los participantes e interactúan con ellos (reformulando las respuestas o estimulando el desarrollo de las mismas) para aprender más sobre el contexto y comprender mejor el significado de las palabras y los acontecimientos.

Por otra parte, una respuesta escrita en un recuadro de comentarios, a una pregunta abierta y opcional, que se formula al final de un cuestionario auto-administrado en línea no puede considerarse un dato QUAL. En los estudios epidemiológicos, las palabras obtenidas de esta manera tradicionalmente proporcionan algunas ilustraciones para discutir los resultados de QUAN. Estas palabras constituyen información, pero no constituyen datos QUAL porque no se han obtenido mediante una metodología y un proceso de investigación científica QUAL rigurosos, y no pueden utilizarse para producir resultados QUAL plausibles. Los investigadores no pueden saber quién escribió estas palabras y por qué (anónimo); no pueden interactuar con los que escribieron y los que no (para aprender más sobre el contexto y comprender mejor el significado de las palabras escritas o por qué no se escribió nada). En el próximo capítulo mostraremos que alrededor del 20% de los estudios de salud se denominan “MM”, pero no cumplen las tres condiciones de la definición porque utilizan información QUAL (o QUAN) que no puede considerarse como datos obtenidos mediante una metodología y un proceso de investigación científica QUAL (o QUAN) rigurosos.

Evaluación de la calidad de los métodos mixtos

Esto nos lleva a la evaluación de la calidad de los métodos mixtos. Combinando una revisión de la literatura y un estudio piloto con talleres y consultas con expertos (Pace et al. 2012; Pluye, Gagnon, Griffiths y Johnson-Lafleur 2009), desarrollamos una herramienta para evaluar la calidad de los estudios que utilizan métodos QUAL, QUAN y MM: la MMAT (Herramienta de Evaluación de Métodos Mixtos) (Pluye et al. 2011). El MMAT utiliza diferentes criterios para diferentes métodos. Así pues, permite: a) evaluar la calidad metodológica de los componentes QUAL de los MM a partir de los criterios utilizados para evaluar los métodos QUAL; b) evaluar los componentes QUAN de los MM a partir de los criterios utilizados en epidemiología para evaluar los métodos QUAN; y c) evaluar la calidad general los MM en cuanto a la

justificación del diseño, la integración de los métodos QUAL y QUAN, y el valor añadido y las limitaciones de esa integración. El MMAT está disponible gratuitamente en Internet (mixedmethodsappraisaltoolpublic.pbworks.com). Esta herramienta es fácil de utilizar y cabe en una sola página, incluye un manual de uso. Trabajos recientes han puesto a prueba la fidelidad del MMAT y han demostrado la validez y eficiencia de esta herramienta (Souto et al. 2015).

Estado del conocimiento sobre la integración en métodos mixtos

Los análisis de MM “consisten en técnicas analíticas aplicadas a los datos QUAL y QUAN” (procedimientos estándar utilizados en epidemiología/estadística e investigación cualitativa), y técnicas que integran las fases, resultados y datos de QUAL y QUAN “en un solo punto del proceso de investigación de MM o en múltiples puntos en el tiempo” (Creswell y Plano Clark 2011, pág. 212). Éstas se agrupan en estrategias específicas de los MM que se presentarán en la sección siguiente “Aplicación”, en forma de propuesta. En la sección “Análisis” presentaremos combinaciones de estas estrategias.

Esta propuesta se basa en una síntesis de las principales publicaciones de referencia metodológica. Esta síntesis es innovadora por dos razones: integra todas las estrategias específicas conocidas y todas las combinaciones posibles de estrategias. Además, esta propuesta está asociada a un sistema de vigilancia que permite ponerla a prueba e identificar estrategias emergentes (imprevistas).

Recientemente se ha publicado una visión general de los análisis en MM según 13 dimensiones (Onwuegbuzie y Combs 2010), que incluye la secuencia temporal de los análisis, la prioridad dada a un análisis particular, la orientación de los análisis (casos, variables, procesos), la interdependencia de los análisis, los vínculos entre los análisis y otros aspectos de las especificaciones (filosofía, objetivos y formas de datos) y la forma en que se generalizan los resultados (estadística o teóricamente). Aparte de este panorama general, las publicaciones metodológicas de MM y las obras de referencia proponen estrategias básicas no combinadas (cada estrategia se presenta como un proceso necesario y suficiente para obtener resultados)

y las publicaciones suelen ser prescriptivas (no probadas con métodos científicos como la recopilación y el análisis de datos de investigación) y se limitan a unas pocas especificaciones o estrategias.

Para planificar nuestra propuesta, hemos agrupado estas estrategias básicas no combinadas en tres grandes categorías de estrategias específicas: (a) las que conectan las fases QUAL y QUAN, (b) aquellas que comparan los resultados de los análisis QUAL y QUAN, y (c) las que asimilan los datos QUAL y QUAN. En la siguiente sección, definiremos cada una de estas categorías. Se destacan tres ventajas: se han definido claramente aplicando los principios de armonización de los términos utilizados en documentación (ciencias de la información) (Organización Internacional de Normalización [ISO] 2009; Roche 2012), integran las terminologías anteriores (Bazeley 2009; Greene 2007; Teddlie y Tashakkori 2009), y se refieren a los tres tipos comunes de integración de los métodos QUAL y QUAN (conexión de fases, comparación de resultados y asimilación de datos).

La conexión de fases se denomina “desarrollo secuencial” (Bazeley), o “correlación y comparación” (Greene), o “análisis mixto secuencial” (Teddlie y Tashakkori). La comparación de los resultados se denomina “triangulación y expansión” (Bazeley), o “análisis inferencial conjunto” (Greene), o “análisis mixto paralelo” (Teddlie y Tashakkori). La asimilación de datos se denomina “transformación” (Bazeley), “transformación y consolidación” (Greene) o “análisis de conversión mixta” (Teddlie y Tashakkori). Además, hemos incluido dos de las tres categorías propuestas por Creswell, pero sin limitar cada categoría a un diseño en MM. En efecto, Creswell y sus colaboradores limitan cada categoría a un diseño así: conectar las fases de QUAL y QUAN (diseño secuencial), comparar los resultados de QUAL y QUAN, y transformar los datos de QUAL o QUAN (diseño convergente) (Creswell y Plano Clark 2011; Guetterman, Fetters y Creswell 2015).

Tres tipos de integración en MM

Greene, Teddlie y Tashakkori (Greene 2007; Teddlie y Tashakkori 2003, 2009, 2010) sugieren tres principios de MM (complementariedad, tensión dialéctica y unificación) que justifican conceptualmente los tres tipos de integración propuestos (conexión de fases, comparación de resultados y asimilación de datos). Estos tres tipos de integración no se excluyen

mutuamente (pueden combinarse) y no están asociados a un juicio de valor (por ejemplo, la conexión de fase no es una integración mejor o peor que la asimilación de datos).

- **Conexión de fases.** El principio de complementariedad se deriva de la literatura que sugiere que las interpretaciones de las ciencias (epistemología, ontología, metodología y teleología) asociadas a los métodos QUAL son diferentes y separadas de las asociadas a los métodos QUAN. Los métodos de recopilación y análisis de datos de QUAL y QUAN están separados. Los métodos y resultados de QUAL y QUAN se presentan por separado en la literatura, pero al menos una frase sugiere que los resultados de QUAL y QUAN son complementarios. La integración (transición cognitiva) se produce cuando hay una conexión entre dos fases (por ejemplo, entre una fase QUAL y una fase QUAN).
- **Comparación de resultados.** El principio de la tensión dialéctica proviene de la literatura que sugiere que las interpretaciones de las ciencias asociadas a los métodos QUAL y QUAN son diferentes e interdependientes (su yuxtaposición genera tensiones creativas de descubrimiento e innovación). Los métodos de recogida y análisis de datos de QUAL y QUAN están separados o interconectados en uno o más puntos en el tiempo, y los resultados están interconectados a través de un proceso de comparación. Se explicitan las similitudes, diferencias y contradicciones entre los resultados de QUAL y QUAN (guiados por una representación cognitiva de los resultados a crear). Por ejemplo, las discrepancias entre los resultados de QUAN y QUAL se mencionan en la literatura, y al menos una frase sugiere cómo resolverlas.
- **Asimilación de datos.** El tercer principio, la unificación, se centra en una cuestión o teoría o interpretación de la ciencia asociada a los métodos QUAL y QUAN. Corresponde a dos corrientes de pensamiento: por un lado, los métodos responden a las preguntas de la investigación y movilizan teorías que unifican el uso de MM sobre una pregunta o teoría (independientemente de las interpretaciones de las ciencias). Por otra parte, varias interpretaciones de las ciencias permiten directamente el uso de MM (unificación en una sola interpretación). Este principio justifica la asimilación de datos (guiada por una representación cognitiva de los resultados a crear). Los datos de QUAL y QUAN pueden

transformarse en una sola forma de QUAL (temas) o QUAN (variables) o fusionarse caso por caso.

Aplicación

Marco conceptual

Proponemos un marco conceptual que incluye estos tres tipos de integración, nueve estrategias específicas y siete combinaciones de estrategias, para integrar las fases, resultados y datos de QUAL y QUAN en MM. Esta conceptualización se resume en el cuadro 1 y la figura 1 y sugiere que cualquier estudio que utilice MM puede combinar varias estrategias. Se inspira en el construccionismo social (Hacking 1999): cualquier estrategia produce un objeto de tipo mixto (*mixed kind*) por efecto de bucle (*looping effect*) entre las fases, resultados o datos de QUAL y QUAN. En otras palabras, cada estrategia se define por un recurso (una contribución en el proceso de gestión del proyecto de investigación mediante MM), un proceso (efecto de bucle) y un producto (objeto mixto). Cualquier combinación de estrategias multiplica los efectos de bucle y los objetos mixtos.

Para cada tipo de integración, las estrategias específicas que proponemos (sin limitar cada estrategia a un único diseño) corresponden a las estrategias básicas descritas por los autores y las autoras a menudo citadas en métodos mixtos. En la evaluación se suelen describir cuatro estrategias básicas: desarrollo de tipología o taxonomía, análisis de casos extremos, transformación de datos y consolidación o fusión de datos (Caracelli y Greene 1993). En las ciencias de la salud, las estrategias básicas que se suelen describir son la comparación de resultados (triangulación), seguimiento de un hilo conductor (*follow a thread*), la fusión de los datos de QUAL y QUAN para cada caso (meta-matriz), y la transformación de los datos de QUAL en QUAN (cuantificación) o de los datos de QUAN en QUAL (cualitativo) (O’Cathain, Murphy y Nicholl 2010; Sandelowski 2000). A continuación, se presentan las nueve estrategias de integración específicas de la MM.

Primer tipo de integración: conexión de fases (3 estrategias

específicas)

Estrategia 1.1 Conectar una fase QUAL a una fase QUAN

En esta estrategia, el recurso consiste en los resultados de la Fase 1 QUAL (Creswell y Plano Clark 2011). El proceso es la conexión de los resultados de QUAL con la recopilación y el análisis de datos de la Fase 2 QUAN; por ejemplo, los investigadores utilizan los resultados de la Fase 1 QUAL para construir el cuestionario de la Fase 2 QUAN. El producto mixto es una recopilación/análisis de datos de QUAN que está informado (estructurado) por los resultados de QUAL. Con esta estrategia se logran, entre otros, los siguientes objetivos a) Desarrollar una tipología (fase 1-QUAL) y una taxonomía (fase 2-QUAN); b) Crear un instrumento (fase 1-QUAL) y utilizarlo para la medición (fase 2-QUAN); c) Validar un instrumento (validación de contenido QUAL y luego validación de construcción QUAN); o validación del contenido ecológico QUAL -representatividad de las preguntas- luego QUAN -relevancia de las preguntas-; d) diseñar una intervención (fase-1 QUAL) y evaluar su implementación (fase-2 QUAN); y e) construir un marco conceptual o teórico (fase-1 QUAL) y probarlo (fase-2 QUAN).

Estrategia 1.2 Conectar una fase QUAN a una fase QUAL

El recurso consiste en los resultados de la fase 1 QUAN (Creswell y Plano Clark, 2011). El proceso es la conexión de los resultados de QUAN con la recopilación y análisis de datos de la Fase 2 QUAL; por ejemplo, los investigadores utilizan los resultados de la Fase 1 QUAN para identificar las categorías de informantes clave de la Fase 2 QUAL. El producto mezclado es una recolección/análisis de datos de QUAL que se basa en los resultados de QUAN. Con esta estrategia se logran, entre otros, los siguientes objetivos a) validar un instrumento (los resultados QUAL pueden explorar explicaciones para los resultados QUAN); b) explicar los resultados QUAN utilizando los resultados QUAL obtenidos con una muestra intencionada de informantes clave, por ejemplo, para comprender mejor las diferencias entre grupos (intervención frente a control; favorecido frente a perjudicado); y c) ayudar a explicar los casos extremos.

Caso especial de 1.1 y 1.2: Seguir un hilo conductor entre las fases QUAL y QUAN

Este caso en particular se llama la estrategia de “hilo conductor” (*follow the thread*) (O’Cathain et al. 2010). El recurso se compone de los resultados de una fase QUAL o QUAN. El proceso consiste en conectar los resultados de esta fase con la recopilación de datos y el análisis de la siguiente fase (QUAN o QUAL), identificar un tema QUAL o una variable clave QUAN durante estas fases, y volver a analizar en profundidad los datos relacionados con este tema o variable de forma prospectiva y retrospectiva (metafóricamente como cuando se desenrolla un sendero de migas de pan en el camino a una cueva, y luego se sigue este sendero de nuevo al aire libre). El producto mezclado consiste en una recolección/análisis de datos de QUAL o QUAN que es informado por los resultados de QUAN o QUAL.

Segundo tipo de integración: comparación de resultados (3 estrategias específicas)

Estrategia 2.1. Comparación de los resultados obtenidos por separado

En esta estrategia, el recurso consiste en los resultados de QUAL y QUAN obtenidos mediante la recolección y análisis de datos por separado (Creswell y Plano Clark 2011). El proceso es una comparación por parte de los investigadores de las similitudes y diferencias entre los resultados de QUAL y QUAN, por ejemplo, yuxtaponiéndolos en una tabla con cada fila correspondiente a un tema principal de QUAL y la(s) variable(s) correspondiente(s). El producto mixto consiste en una interpretación o decisión. Normalmente, una tabla de comparación consta de cuatro columnas: subtítulo (columna 1), resultados QUAL (columna 2), resultados QUAN (columna 3) e interpretación o decisión (columna 4).

Estrategia 2.2. Comparación de los resultados obtenidos de manera interdependiente

El recurso consiste en los resultados de QUAL y QUAN obtenidos mediante la recopilación y el análisis interdependiente de datos (Creswell y

Plano Clark 2011). Al igual que en la estrategia anterior, el proceso es una comparación por parte de los investigadores de las similitudes y diferencias entre los resultados de QUAL y QUAN. El producto mixto es una interpretación o decisión que tiene en cuenta explícitamente las interdependencias, por ejemplo, cuando una variable QUAN se deriva directamente de un tema QUAL (o participantes QUAL reclutados entre los participantes QUAN). Cuando existen interdependencias, los investigadores esperan principalmente ver similitudes, como la corroboración de los resultados de QUAL por los resultados de QUAN (o viceversa). Se pueden hacer varias hipótesis para explicar las diferencias, si las hubiera (cada diferencia puede suponer una amenaza para la validez interna de los resultados de la QUAN o la credibilidad de los resultados de la QUAL).

Caso especial de las estrategias 2.1 y 2.2: Divergencia de los resultados QUAL y QUAN

En este caso particular, los investigadores comparan los resultados QUAL y QUAN, centrándose en las discrepancias (contradicciones, discordias o disonancias). Como antes, el proceso es una comparación por parte de los investigadores de los resultados QUAL y QUAN. El producto mixto consiste en una interpretación o decisión para cada discrepancia. En la literatura de MM, encontramos ejemplos que ilustran dos estrategias principales que los investigadores utilizan para tratar las discrepancias entre los resultados de QUAL y QUAN: la reconciliación o el inicio de un nuevo proyecto de investigación (Pluye, Grad, Levine, y Nicolau 2009).

Tercer tipo de integración: asimilación de datos (3 estrategias específicas)

Estrategia 3.1. Transformación de datos QUAL en datos QUAN (cuantificación)

En esta estrategia, el recurso consiste en todos o algunos de los datos de QUAL de la investigación cualitativa, y los datos de QUAN. El proceso es la transformación o conversión de los datos QUAL en datos QUAN para

que los datos QUAL transformados puedan ser analizados con los datos QUAN. Por ejemplo, los investigadores utilizan la antigua técnica de análisis cuantitativo del contenido (Neuendorf, 2002): transforman los temas clave en un pequeño número de variables, y los correspondientes datos QUAL en valores numéricos, utilizando una rejilla de codificación válida y precisa y un manual. El producto mixto consiste en datos QUAN (combinando datos QUAN y datos QUAL transformados en datos QUAN) que pueden ser analizados usando técnicas estadísticas estándar. El rigor de esta estrategia se basa en la validez y la fidelidad inter-jueces del análisis cuantitativo de contenido (siendo la hoja de registro y la guía de codificación los instrumentos de medición). En nuestra opinión, muchos investigadores están criticando incorrectamente esta estrategia con el argumento de que “reduciría” los datos QUAL. De hecho, no hay nada que impida que todos los datos QUAL sean analizados en paralelo. La transformación de los datos de QUAL en datos de QUAN produce un valor añadido en términos de análisis y resultados adicionales.

Estrategia 3.2. Transformación de los datos de QUAN en datos de QUAL

El recurso consiste en todos o parte de los datos QUAN de una investigación epidemiológica o estadística, y los datos QUAL. El proceso es la transformación o conversión de los datos QUAN en datos QUAL para que los datos QUAN transformados puedan ser analizados con los datos QUAL. Los investigadores utilizan análisis interpretativos (por ejemplo, temáticos) y transforman los resultados estadísticos en historias o narraciones (por ejemplo, organizados por temas) (Sandelowski, 2000). El producto mixto consiste en datos QUAL (combinando datos QUAL y datos QUAN transformados en datos QUAL) que pueden ser analizados de forma narrativa. Como en la estrategia anterior, no hay nada que impida que todos los datos QUAN sean analizados estadísticamente en paralelo. La transformación de los datos QUAN en datos QUAL proporciona un valor añadido (análisis y resultados adicionales).

Estrategia 3.3. Fusión de los datos de QUAL y QUAN

En esta última estrategia, el recurso consiste en casos (por ejemplo, individuos, organizaciones, eventos, etc.) y datos de QUAL y QUAN (O’Cathain et al. 2010). El proceso consiste en la fusión de los datos QUAL y QUAN caso por caso (análisis intra-caso) para permitir análisis entre casos. El producto mezclado consiste en un nuevo conjunto de datos que pueden analizarse de manera estadística (por ejemplo, la descripción de un caso y la comprobación de hipótesis) o narrativa (por ejemplo, la comparación de historias de casos). El rigor de esta estrategia se basa en la definición de los casos y en la recopilación de datos similares QUAL y QUAN para cada caso. Por ejemplo, los clínicos evaluaron los efectos cognitivos y el uso de la información encontrada en los recursos electrónicos especializados utilizando un cuestionario validado (estudio longitudinal QUAN); los casos se definieron como búsquedas de información en las que se utilizó la información clínica para un paciente específico. En cada caso se entrevistó a los clínicos y se describieron los efectos del uso de esta información en la salud de los pacientes (estudios de casos múltiples de QUAL), los datos QUAN y QUAL se fusionaron en viñetas clínicas (cada viñeta describía un caso), y estas viñetas se utilizaron para construir una “meta-matriz mixta” y estimar la proporción de pacientes cuya salud se benefició de las búsquedas de información de los clínicos (Pluye, Grad, Johnson-Lafleur et al. 2013; Pluye, Grad, Repchinsky et al. 2013).

Análisis reflexivo: Combinaciones de estrategias para integrar fases, resultados y datos QUAL y QUAN

Una combinación se define por uno o más elementos de un conjunto, independientemente del orden de selección. En la figura 1 se muestran las siete combinaciones posibles de las estrategias propuestas.

En este capítulo se describió el pluralismo de los métodos mixtos (MM) en cuanto a las interpretaciones de las ciencias (epistemología, ontología, metodología y teleología) y se recordó que los MM suelen requerir la colaboración de expertos en métodos QUAL y QUAN (por ejemplo, un supervisor de QUAL y un co-supervisor de QUAN para los estudiantes de doctorado) que tienen que conciliar su interpretación de las ciencias cuando

éstas difieren. Además, en este capítulo se mencionan los principales tipos de especificaciones y criterios de calidad en la MM. A este conocimiento básico añadimos tres tipos de integración, nueve estrategias específicas de MM y todo un nuevo mundo a explorar con múltiples combinaciones posibles de estrategias para integrar las fases, resultados y datos de QUAL y QUAN.

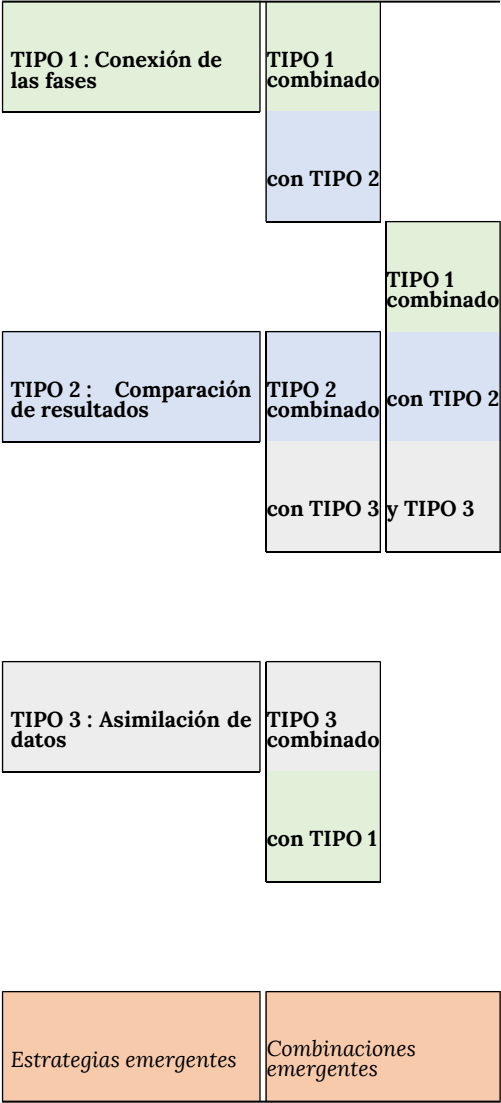
Este capítulo ha contribuido al conocimiento de MM en términos de definición, práctica e innovación. Ha definido los conceptos de secuenciación y convergencia en relación con la literatura sobre gestión de proyectos, en este caso la gestión de proyectos de investigación o evaluación utilizando MM. Nuestra conceptualización contribuye al conocimiento sobre MM y es práctica. Puede guiar a los estudiantes, investigadores y evaluadores en la planificación, realización y evaluación de estudios y evaluaciones de programas utilizando la MM, particularmente en la salud global. En el siguiente capítulo ilustraremos todas estas estrategias y combinaciones con las investigaciones sobre la salud mundial. Discutiremos las fortalezas y debilidades de nuestras propuestas a la luz de esta investigación.

Tabla 1 : Nueve estrategias específicas para integrar las fases, resultados o datos cualitativos (QUAL) y cuantitativos (QUAN) en métodos mixtos (MM)

3 TIPOS DE INTEGRACIÓN Y 9 ESTRATEGIAS ESPECIFICAS	DEFINICIONES
TIPO 1 : Conexión de fases	Conectar fases QUAL y QUAN
Fase QUAL a fase QUAN	Conectar los resultados de una fase-1 QUAL con la recogida y análisis de una fase-2 QUAN [1].
Fase QUAN a fase QUAL	Conectar los resultados de una fase-1 QUAN con la recogida y análisis de una fase-2 QUAL [1].
Caso particular de 1.1 et 1.2 : “Follow a thread”	Analizar los datos QUAL (o QUAN) e identificar los principales temas (o variables) que necesitan una profundización ; escoger un tema (o variable) y reanalizar a través de los componentes QUAN (o QUAL) [2].
TIPO 2 : Comparación de resultados	Comparar resultados QUAL y QUAN
2.1. Resultados QUAL y QUAN obtenidos de manera separada	Comparar las similitudes y diferencias entre los resultados QUAL et QUAN obtenidos de la recogida y el análisis de datos separados [1].
2.2. Resultados QUAL et QUAN obtenidos de manera interdependiente	Comparar las similitudes y diferencias entre los resultados QUAL y QUAN obtenidos de la recogida y el análisis de datos interdependientes [1].
Caso particular de 2.1 y 2.2 : Divergencia de resultados QUAL y QUAN	Comparar, destacando las discrepancias (contradicciones, incoherencias o disonancias) entre los resultados QUAL y QUAN [3].
TIPO 3 : Asimilación des datos	Asimilar los datos QUAL et QUAN
3.1. Datos QUAL en datos QUAN	Asimilar los datos QUAL y QUAN transformando los datos QUAL en datos QUAN [1].
3.2. Datos QUAN en datos QUAL	Asimilar los datos QUAL y QUAN transformando los datos QUAN en datos QUAL [1].
3.3. Fusión de datos QUAL y QUAN	Asimilar los datos QUAL y QUAN fusionándolos para cada caso en una base de datos adicional [1].
TIPO 4 : Estrategias emergentes	Descúbrelo observando la evolución de las tendencias en MM con eSRAP

[1](Creswell et Plano Clark 2011); [2] (O’Cathain *et al.* 2010); [3](Pluye, Grad *et al.* 2009)

FIGURA 1. SIETE COMBINACIONES DE ESTRATEGIAS PARA LA INTEGRACIÓN DE LAS FASES, RESULTADOS O DATOS CUALITATIVOS (QUAL) Y CUANTITATIVOS (QUAN) EN MÉTODOS MIXTOS (MM)



Referencias clave

Méthodes mixtes francophonie :

<http://methodesmixtesfrancophonie.pbworks.com>

Esta nueva asociación internacional sin fines de lucro tiene como objetivo promover los métodos mixtos en francés. En 2018, ya cuenta con 121 miembros de 8 países diferentes (Alemania, Bélgica, Canadá, Francia, Rwanda, Senegal y Suiza). Ofrece apoyo metodológico, una bibliografía abierta de las publicaciones de los miembros y una lista de expertos, recursos y cursos de capacitación. En particular, ofrece acceso gratuito a la primera obra de referencia metodológica colectiva sobre métodos mixtos en francés (14 capítulos, 266 páginas): Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. y Pluye, P. (2018). *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé*. Montréal : 117e Cahiers scientifiques de l'ACFAS.

Creswell, J. W. et Clark, V. L. P. (2018). *Designing and conducting mixed methods research* (3^e éd.). Thousand Oaks, CA: Sage Publications.

Este es el libro metodológico más popular sobre métodos mixtos. Se detallan los pasos para llevar a cabo la investigación de métodos mixtos. Este libro es una guía práctica para los evaluadores, estudiantes e investigadores que se inician en la investigación de métodos mixtos. En él se ofrecen especificaciones para la realización de investigaciones con métodos mixtos, ejemplos de recopilación y análisis de datos y recomendaciones para redactar un protocolo o una publicación.

Tashakkori, A. et Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.

Este libro de referencia colectiva incluye capítulos sobre prácticamente todos los aspectos importantes de los métodos mixtos en las ciencias sociales y de la salud. Está escrito por investigadores de varias disciplinas y países reconocidos en sus campos. Presenta un estado del arte de los conocimientos epistemológicos y metodológicos sobre los métodos mixtos. Es un compañero de los estudiantes e investigadores que utilizan métodos mixtos.

Referencias

- Bazeley, P. (2009). Analyzing mixed methods data. Dans S. Andrew et E. Halcomb (dir.), *Mixed methods research for nursing and the health sciences* (p. x-x). Chichester : Wiley.
- Bryman, A. (2006). Integrating quantitative and qualitative research: how is it done?. *Qualitative Research*, 6(1), 97-113.
- Campbell, D. T. (1988). Qualitative knowing in action research. Dans S. Overman (dir.), *Methodology and epistemology for social science: selected papers of Donald T. Campbell* (p. 360-376). Chicago : The University of Chicago Press.
- Caracelli, V. J. et Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 15(2), 195-207.
- Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2^e éd.). Thousand Oaks, CA: Sage Publications.
- Dagenais, C., Nault-Brière, F., Dupont, D. et Dutil, J. (2008). Implementation and effects of a service coordination program for youths and their families in complex situations: A mixed evaluation design. *Children and Youth Services Review*, 30(8), 903-913.
- De Waal, C. (2005). *On pragmatism*. Belmont : Wadsworth.
- Gendron, S. (2001). *La pratique participative en santé publique : l'émergence d'un paradigme* (Thèse de doctorat). Université de Montréal, Montréal.
- Greene, J. C. (2007). *Mixed methods in social inquiry* (édition, volume). San Francisco: Jossey Bass.
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology?. *Journal of Mixed Methods Research*, 2(1), 7-22.
doi: 10.1177/1558689807309969
- Guetterman, T. C., Fetters, M. D. et Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine*, 13(6), 554-561.
- Hacking, I. (1999). *The social construction of what?*. Cambridge, MA: Harvard University Press.

- International Standards Organization (ISO). (2009). *Terminology work – Principles and methods*.
http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=38109
- Johnson, R. B., Onwuegbuzie, A. J. et Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Langley, A., Mintzberg, H., Pitcher, P., Posada, E. et Saint-Macary, J. (1995). Opening up decision making: the view from the black stool. *Organization Science*, 6(3), 260-279.
- Lisle, J. D. (2013). Exploring the value of integrated findings in a multiphase mixed methods evaluation of the continuous assessment program in the Republic of Trinidad and Tobago. *International Journal of Multiple Research Approaches*, 7(1), 27-49.
- Neuendorf, K. A. (2002). *The content analysis guidebook*. Thousand Oaks, CA: Sage Publications.
- Niglas, K. (2010). The multidimensional model of research methodology: an integrated set of continua. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2^e éd., p. 215-236). Thousand Oaks, CA: Sage Publications.
- O’Cathain, A, Murphy, E. et Nicholl, J. (2010). Three techniques for integrating data in mixed methods studies. *British Medical Journal*, 341, 1147-1150.
doi: 10.1136/bmj.c4587
- Onwuegbuzie, A. J. et Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: a synthesis. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 397-430). Thousand Oaks, CA: Sage Publications.
- Pace, R., Pluye, P., Bartlett, G., Macaulay, A. C., Salsberg, J., Jagosh, J. et Seller, R. (2012). Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *International Journal of Nursing Studies*, 49(1), 47-53.
- Pluye, P., Gagnon, M. P., Griffiths, F. et Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly

- appraising qualitative, quantitative and mixed methods primary studies in mixed studies reviews. *International Journal of Nursing Studies*, 46(4), 529-546.
- Pluye, P., Grad, R. M., Johnson-Lafleur, J., Granikov, V., Shulha, M., Marlow, B. et Ricarte, I. (2013). The Number Needed to Benefit from Information (NNBI): proposal from a mixed methods research study with practicing family physicians. *The Annals of Family Medicine*, 11(6), 559-567.
doi: 10.1370/afm.1565
- Pluye, P., Grad, R. M., Levine, A. et Nicolau, B. (2009). Understanding divergence of quantitative and qualitative data and results in mixed methods studies. *International Journal of Multiple Research Approaches*, 3(1), 58-72.
- Pluye, P., Grad, R. M., Repchinsky, C., Jovaisas, B., Johnson-Lafleur, J., Carrier, M. E., Légaré, F. (2013). Four levels of outcomes of information-seeking: a mixed methods study in primary health care. *Journal of the American Society for Information Science and Technology (JASIST)*, 64(1), 108-125.
- Pluye, P. et Hong, Q. N. (2014). Combining the power of stories and the power of numbers: mixed methods research and mixed studies reviews. *Annual Review of Public Health*, 35, 29-45.
- Pluye, P., Robert, E., Cargo, M., Bartlett, G., O'Cathain, A., Griffiths, F., Rousseau, M.C. (2011). Proposal: a mixed methods appraisal tool for systematic mixed studies reviews.
<http://mixedmethodsappraisaltoolpublic.pbworks.com>
- Porta, M. (2008). *A dictionary of epidemiology*. New York : Oxford University Press.
- Ridde, V. et Dagenais, C. (2012). Introduction générale à l'évaluation de programme. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme* (p. 11-29). Montréal : Presses de l'Université de Montréal.
- Roche, C. (2012, mai). *Ontologie : entre terminologie et connaissance de spécialité*. Communication présentée au GLAT – Terminologies : textes, discours et accès aux savoirs spécialisés, Gênes.
<https://hal.archives-ouvertes.fr/hal-01180282/document>

- Sandelowski, M. (2000). Combining qualitative and quantitative sampling, data collection, and analysis techniques in mixed-method studies. *Research in Nursing et Health*, 23(3), 246-255.
- Sayer, A. (2000). *Realism and social science*. Londres : Sage Publications.
- Schwandt, T. A. (2007). *Dictionary of qualitative inquiry* (3^e éd.). Thousand Oaks, CA: Sage Publications.
- Souto, R. Q., Khanassov, V., Hong, Q. N., Bush, P. L., Vedel, I. et Pluye, P. (2015). Systematic mixed studies reviews: updating results on the reliability and efficiency of the mixed methods appraisal tool. *International Journal of Nursing Studies*, 52(1), 500-501.
- Tashakkori, A. et Teddlie, C. (2010). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.
- Teddlie, C. et Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioural sciences. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 3-50). Thousand Oaks, CA: Sage Publications.
- Teddlie, C. et Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: Sage Publications.
- Teddlie, C. et Tashakkori, A. (2010). Overview of contemporary issues in mixed methods research. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 1-41). Thousand Oaks, CA: Sage Publications.
- Tyson, L. (2014). *Critical theory today*. New York: Routledge.
- Van de Ven, A. H. (1992). Suggestions for studying strategy process: a research note. *Strategic Management Journal*, 13(5), 169-188.
- Youngs, H. et Piggot-Irvine, E. (2011). The application of a multiphase triangulation approach to mixed methods: The research of an aspiring school principal development program. *Journal of mixed methods research*, 6(3), 184-198.

Resumen / Résumé / Abstract

Los métodos mixtos (MM) son cada vez más populares. Al utilizar los métodos mixtos, los investigadores integran métodos cualitativos (QUAL) y cuantitativos (QUAN) con respecto a las cosmovisiones, metodologías y diseños, preguntas de investigación, técnicas de recopilación y análisis de datos y resultados de estudios. Se han propuesto varias estrategias de integración, pero su conceptualización suele estar impulsada por el diseño, o ser fragmentada, o no haber sido probada empíricamente. Esto es un reto para la planificación y realización de estudios que utilizan métodos mixtos y para la formación de estudiantes de postgrado. Basado en la literatura metodológica, este capítulo presenta un marco conceptual que incluye tipos de integración, estrategias prácticas y posibles combinaciones de estrategias para integrar las fases, resultados y datos de QUAL y QUAN. En el siguiente capítulo, este marco se prueba utilizando una revisión de los estudios que utilizan métodos mixtos de 2015 con una descripción detallada del método. Esto apoya el marco, ilustra múltiples tipos de combinaciones y proporciona ejemplos de salud mundial. Estos dos capítulos complementarios contribuyen a avanzar en el conocimiento metodológico de los métodos mixtos a través de (a) un llamado a informar mejor sobre los estudios que utilizan métodos mixtos en artículos científicos, y (b) una conceptualización probada que comprende tres tipos de integración y nueve estrategias específicas, que explican las posibilidades actuales y futuras de combinar estrategias para integrar las fases, los resultados y los datos de QUAL y QUAN.

Les méthodes mixtes (MM) sont de plus en plus populaires. En utilisant les MM, les chercheuses et chercheurs intègrent des méthodes qualitatives (QUAL) et quantitatives (QUAN), par exemple au niveau des questions et devis de recherche, et des collectes et analyses des données. Plusieurs stratégies d'intégration ont été proposées, mais leur conceptualisation est généralement limitée, fragmentée, ou non testée scientifiquement. Cela constitue un défi pour planifier et effectuer des études utilisant les MM, et pour enseigner les MM. Fondé sur la littérature méthodologique, ce chapitre présente un cadre conceptuel incluant tous les types d'intégration connus, des stratégies pratiques et des combinaisons possibles de stratégies pour intégrer les phases, les résultats et les données QUAL et QUAN. Dans le

chapitre suivant, ce cadre est testé avec une revue des études utilisant les MM. Cette revue illustre plusieurs types de combinaisons avec des études en santé mondiale. Ensemble, ces deux chapitres contribuent à faire avancer les connaissances méthodologiques par (a) un appel à une meilleure description des MM dans les publications scientifiques, et (b) une conceptualisation testée proposant trois types d'intégration et neuf stratégies spécifiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données QUAL et QUAN.

Mixed methods (MM) are increasingly popular. In performing MM, researchers integrate qualitative (QUAL) and quantitative (QUAN) methods with respect to worldviews, methodologies and designs, research questions, data collection and analysis techniques, and study results. Several integration strategies have been proposed, but their conceptualization is usually design-driven, or fragmented, or not empirically tested. This is challenging for planning and conducting MM studies, and for training graduate students. Based on the methodological literature, this chapter presents a conceptual framework including types of integration, practical strategies, and possible combinations of strategies for integrating QUAL and QUAN phases, results and data. In the next chapter, this framework is tested using a review of 2015 MM studies with a method-detailed description. This supports the framework, illustrates multiple types of combinations, and provides global health examples. These two companion chapters contribute to advance methodological knowledge on MM via (a) a call for better reporting MM studies in scientific articles, and (b) a tested conceptualisation comprising three types of integration and nine specific strategies, which explain current and future possibilities for combining strategies to integrate QUAL and QUAN phases, results, and data.

Pierre Pluye es profesor titular en el Departamento de Medicina Familiar de la Universidad McGill, investigador principal del Fondo de Investigación de Quebec Salud (FRQS) y director del componente de desarrollo metodológico de la Unidad SOUTIEN SRAP de Quebec. Es miembro asociado de la Escuela de Ciencias de la Información de la Universidad McGill. Es especialista en medicina familiar y salud pública. Es doctor en promoción de la salud (Universidad de Montreal). Ha recibido

premios a su carrera de los Institutos Canadienses de Investigación de la Salud (CIHR) y del FRQS. En 2017 fue elegido miembro de la Academia Canadiense de Ciencias de la Salud y recibió el premio “Investigador del Año” del Colegio de Médicos de Familia del Canadá. Tiene experiencia en métodos mixtos. Su investigación estudia los efectos de la información de los recursos electrónicos en los profesionales, los pacientes y el público.

Cita

Pierre Pluye (2020). Integración en métodos mixtos. Marco conceptual para la integración de fases, resultados y datos cualitativos y cuantitativos. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 195-220. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

8. La práctica de la integración en métodos mixtos

Las múltiples combinaciones de las estrategias de integración

PIERRE PLUYE, ENRIQUE GARCÍA BENGOCHEA, DAVID LI TANG Y VERA GRANIKOV

La práctica de la integración en métodos mixtos

Definición del método

Los métodos mixtos integran métodos cualitativos y cuantitativos. Permiten responder a preguntas complejas de investigación y evaluación.

Fortalezas del método

Proponemos tres tipos de integración y nueve estrategias prácticas, y todas las combinaciones posibles de estrategias para integrar las fases, resultados y datos de CUAL y CUAN; éstas propuestas se ilustran con ejemplos en investigación en salud global.

Desafíos del método

Nuestra propuesta y estos ejemplos pueden dar respuesta a una necesidad: a) de que los investigadores/as planifiquen, realicen y publiquen mejor los estudios que utilizan métodos mixtos, y b) de que el profesorado forme a los/as estudiantes en métodos mixtos.

Los métodos mixtos (MM) se definen como cualquier combinación de métodos cualitativos (CUAL) y cuantitativos (CUAN) en la evaluación de programas, la investigación primaria y la revisión de la literatura. Surgieron en la década de 1950 en salud global (Pluye, 2012). En el capítulo anterior, resumimos los conocimientos básicos de los métodos mixtos y propusimos un marco conceptual para las técnicas de análisis de los métodos mixtos. En este capítulo, probamos e ilustramos esta conceptualización con investigaciones sobre salud global.

Antecedentes

Nuestro marco conceptual incluye tres tipos de integración, nueve estrategias específicas y siete combinaciones de estrategias para integrar las fases, resultados y datos de CUAL y CUAN. Esto sugiere que cualquier estudio que use MM puede combinar varias estrategias. El primer tipo de integración es la conexión de fases y corresponde a tres estrategias específicas: (1.1) la conexión de una fase CUAL con una fase CUAN; (1.2) la conexión de una fase CUAN con una fase CUAL; y (1.3) el caso especial de las estrategias anteriores que se denomina “camino de migas de pan” entre las fases CUAL y CUAN. El segundo tipo de integración es la comparación de resultados y se refiere a tres estrategias específicas: (2.1) la comparación de los resultados obtenidos por separado; (2.2) la comparación de los resultados obtenidos de forma interdependiente; y (2.3) el caso especial de las estrategias anteriores que se denomina “divergencia” de los resultados de CUAL y CUAN. El tercer y último tipo de integración es la asimilación de datos. También incluye tres estrategias específicas que son: (3.1) transformando los datos de CUAL en datos de CUAN (cuantificación); (3.2) transformando los datos de CUAN en datos de CUAL (cualitativos); y (3.3) fusionando los datos de CUAL y CUAN. Por las razones explicadas en el capítulo anterior, el conjunto de posibles combinaciones de estas estrategias (1; 2; 3; 1 y 2; 1 y 3; 2 y 3; 1 y 2 y 3) constituye un nuevo mundo que debe explorarse con métodos mixtos (una combinación que se define por uno o más elementos de un conjunto).

Mapa

Para poner a prueba e ilustrar esta conceptualización con investigaciones recientes sobre salud global, utilizamos el eSRAP, que es un sistema de vigilancia de las tendencias de la investigación (Recuadro 1) (Tang, Pluye y Bouthillier 2015). El eSRAP permite la colaboración de los miembros de las comunidades de investigación orientada al paciente RAP *por sus siglas en francés*) para evaluar y compartir los resultados de las investigaciones y los proyectos de manera fácil de utilizar. Según los Institutos Canadienses de Investigación Sanitaria (CIHR), la investigación orientada al paciente se refiere a una serie de investigaciones que involucran a los pacientes como

socios, se centran en las prioridades de los pacientes y mejoran los servicios de salud y la salud. Es multidisciplinar y se lleva a cabo en asociación con todas las partes interesadas. La palabra “paciente” representa a la población en general, incluidas las personas con un problema de salud o social y sus cuidadores.

Los autores de este capítulo contribuyeron a la elaboración del prototipo de eSRAP, que se diseñó específicamente para facilitar la vigilancia de las tendencias de la investigación tan pronto como las publicaciones se indizan en las bases de datos bibliográficas. En el momento de redactar este capítulo, 15 investigadores (profesores y ex alumnos de los cursos de métodos combinados que ofrece el Departamento de Medicina Familiar de la Universidad McGill desde 2008) están utilizando el eSRAP para analizar y comprender el entorno y las tendencias emergentes en lo que respecta a las estrategias para integrar las fases, los resultados y los datos de CUAL y CUAN en los métodos combinados.

eSRAP : Vigilancia en colaboración de las tendencias de las investigaciones orientadas al paciente

¿Por qué las comunidades RAP necesitan eSRAP?

eSRAP es un sistema de Vigilancia de Tendencias de Investigación (STR *por sus siglas en francés*) que permite a los miembros de las comunidades de investigación orientada al paciente (RAP) colaborar para analizar y comprender el entorno y las tendencias emergentes en sus áreas de investigación de una manera que sea relevante para sus necesidades. eSRAP tiene como objetivo mantener informadas a las comunidades PAR del avance del conocimiento en sus respectivos campos. Facilita la vigilancia de los acontecimientos de vanguardia tan pronto como aparecen en forma de publicaciones indizadas en bases de datos bibliográficas.

¿Qué es un sistema STR?

STR aplica los principios de la inteligencia competitiva y la vigilancia del medio ambiente para actualizar los conocimientos más recientes de manera rápida, constante y estructurada. El STR puede complementar las revisiones sistemáticas de la literatura y reemplazar los mecanismos de alerta tradicionales. Por un lado, las evaluaciones sistemáticas son muy completas y producen recomendaciones valiosas, pero son costosas, y no se actualizan constantemente. Por otra parte, los sistemas de alerta tradicionales (fuentes RSS) proporcionan una lista sencilla de títulos y resúmenes no clasificados y a menudo irrelevantes.

Por su parte, el STR ofrece una selección de documentos y resultados de análisis pertinentes de calidad, por y para una comunidad de usuarios. La STR puede requerir pocos recursos (los recursos necesarios disminuyen a medida que aumenta el número de miembros de la comunidad) y conserva las ventajas de capitalizar y aprovechar un conjunto cada vez mayor y más amplio de conocimientos y proyectos científicos. Además, el STR tiene una visión de futuro, ya que se centra en los últimos protocolos y publicaciones a medida que se van publicando.

Cada parte del eSRAP es adaptable a las necesidades de una comunidad RAP

La estrategia de vigilancia se determina con expertos en el tema/campo (por ejemplo, los líderes de una comunidad de RAP). Se valida y pone en práctica un enfoque estructurado para difundir y organizar los nuevos conocimientos. Los análisis pueden ser automatizados (con algoritmos), semiautomáticos o manuales. Se pueden crear alertas para difundir los nuevos conocimientos y los resultados de los análisis a los miembros de la comunidad. La comunidad del RAP puede utilizar este depósito estructurado de conocimientos (que se actualiza continuamente) para generar análisis que puedan facilitar la capacitación, los protocolos y las publicaciones en los medios tradicionales y sociales. El eSRAP está financiado por la *Unité SOUTIEN SRAP du Québec* y lo desarrollan expertos académicos e industriales en inteligencia competitiva y vigilancia de la ciencia y la tecnología.

Referencia : Tang D., Pluye P. et Bouthillier F. (2015). eSRAP : Système de surveillance des tendances en recherche permettant la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour évaluer et partager les résultats et les projets de recherche de manière adaptée. Registration of Copyright (# 1126124), Canadian Intellectual Property Office, Industry Canada.

Para escribir este capítulo, eSRAP seleccionó artículos que describen en detalle las estrategias utilizadas en la investigación de métodos mixtos centrados en el paciente en las ciencias sociales y de la salud. Nuestros criterios de selección fueron los siguientes: Estudio empírico utilizando métodos mixtos en investigaciones centradas en el paciente publicado en 2015. Nuestra fuente de información fue la base de datos bibliográfica de Scopus. Utilizamos la siguiente búsqueda: “TÍTULO (método mixto PRE/ 5*) Y SUBJAREA (MEDI O SOCI)”. Esto permitió identificar una serie de documentos (autores, revista, año, título y resumen) relacionados con la gestión de la movilidad con una alta precisión (baja recordación) que

podieron ser manejados por dos revisores (Pierre Pluye y Enrique García Bengoechea). Los documentos y artículos pertinentes se seleccionaron utilizando un manual de codificación que incluía la definición de los métodos mixtos y la investigación centrada en el paciente.

Sólo se incluyeron estudios que utilizaban métodos mixtos (MM) y que proporcionaban una descripción detallada de este uso. Utilizamos las seis recomendaciones GRAMMS (Good Reporting of A Mixed Methods Study) para definir este criterio de selección (O’Cathain, 2010). Estas recomendaciones son para describir al menos:

1. el objetivo y la especificación de MM;
2. la toma de muestras, la recogida y el análisis de los datos de CUAL y CUAN;
3. la integración de las fases, resultados y datos de CUAL y CUAN;
4. una justificación para el uso de MM;
5. una descripción de los beneficios asociados a la integración de los métodos;
6. una descripción de las limitaciones de esta integración.

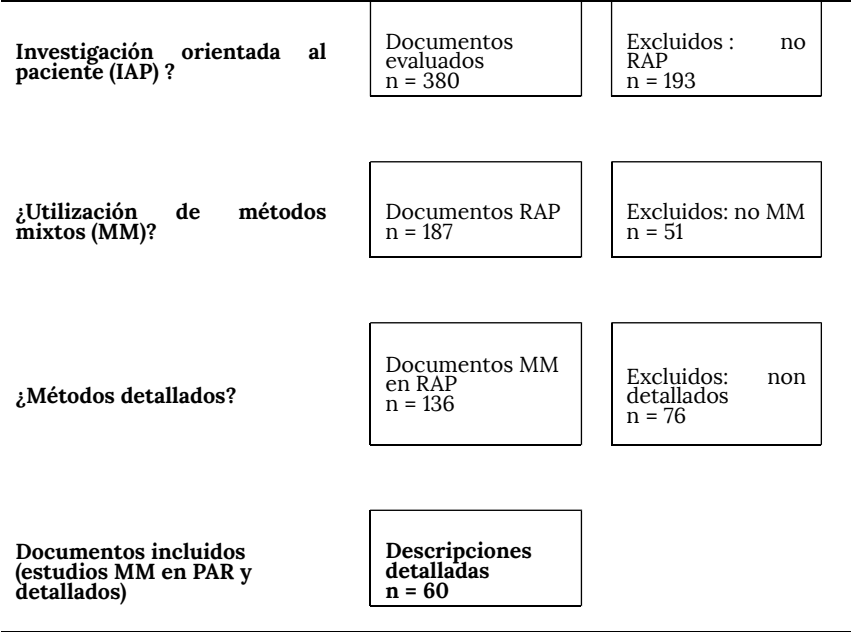
Para cumplir con este criterio de selección (descripción detallada), se juzgó que, como mínimo, los tres componentes principales de MM (métodos CUAL y CUAN e integración de los métodos CUAL y CUAN) deberían cumplir las tres recomendaciones GRAMMS (recomendaciones 1-2-3). Utilizamos las tres últimas recomendaciones del GRAMMS (recomendaciones 4-5-6) para describir los estudios incluidos, pero no como criterios de exclusión.

Aplicación

De los 380 documentos sobre “métodos mixtos” evaluados en el momento de la presentación de este capítulo, 187 (49,2%) estaban relacionados con la RAP (incluidos 35 en el ámbito de la salud global) (figura 1). De ellos, 136 (35,8%) informaron de investigaciones empíricas primarias que se ajustaban a la definición de MM (incluidos 29 en el ámbito de la salud global). Sólo 60 documentos (15,8%), incluidos 11 estudios sobre la salud global, proporcionaron una descripción detallada de los tres componentes principales de la MM y, por lo tanto, cumplieron las tres primeras

recomendaciones del GRAMMS (recomendaciones 1-2-3). Los documentos restantes no describían la metodología y los métodos, o lo hacían en una frase o párrafo, o describían sólo los componentes CUAL y CUAN.

Figura 1. Diagrama de flujo eSRAP-MM



De los 51 documentos que no se ajustaban a nuestra definición de MM (incluidos seis en el ámbito de la salud global), 20 informaban de la recopilación o el análisis de información cualitativa que no era “datos de investigación” porque no se refería a una metodología y un método CUAL (incluidos dos en el ámbito de la salud global) y 19 documentos informaban de la recopilación o el análisis de información cuantitativa que no era “datos de investigación” porque no se refería a un diseño y un método CUAN (incluidos cuatro en el ámbito de la salud global); un documento informó de una yuxtaposición de los métodos de CUAL y CUAN sin ninguna integración (como dos estudios separados), otro informó de una combinación de métodos de CUAN, otro informó de ninguna metodología o método (descripción del programa), y nueve no fueron elegibles por varias razones (revisión de la literatura, texto completo confuso o no disponible a través de la biblioteca McGill).

Los siguientes resultados se refieren únicamente al análisis de las 60 publicaciones que informaron de estudios de RAP que satisfacen la definición de MM y que proporcionaron una descripción detallada de los tres componentes principales de la MM que satisfacen las tres primeras recomendaciones del GRAMMS (recomendaciones 1-2-3). En esta muestra, 38 estudios (63,3%) siguieron una especificación convergente, 17 (28,3%) siguieron una especificación secuencial, cuatro (6,7%) siguieron una especificación multifacética, y un estudio (1,7%) siguió una especificación multinivel. De los 60 documentos seleccionados, sólo seis (10,0%) cumplieron las tres últimas recomendaciones del GRAMMS (recomendaciones 4-5-6), 18 (30,0%) cumplieron dos, 25 (41,7%) cumplieron una y 11 (18,3%) no cumplieron ninguna. Todas las estrategias específicas y combinaciones de estrategias propuestas se presentan en la tabla 1 y se ilustran en el resto del capítulo con ejemplos de la salud global.

Tabla 1 : eSRAP-MM: Distribución de las publicaciones de MM en el RAP por tipo de cotización y combinaciones de estrategias (N=60)

3 TIPOS DE INTEGRACIÓN 9 TIPOS DE ESTRATEGIAS	COMBINACIONES DE ESTRATEGIAS ESPECÍFICAS DE MM (DEPENDIENDO DEL TIPO DE CITA) Símbolo de combinación (n = número de estudios que ilustran cada combinación)			
	Secuencial	Convergente	Multifase	Multinivel
1. CONEXIÓN DE FASE				
1.1 Fase de CUAL a CUAN	1.1 (n=1)	na	1.1 (n=1)	
1.2 Fase de CUAN a CUAL	1.2 (n=3)	na		
1.3 Caso especial: Sigue un hilo		na	1.1 y 1.2 y 1.3 y 2.2 (n=1)	
2. COMPARACIÓN DE RESULTADOS				
2.1 Resultados de CUAL y CUAN obtenidos por separado	s.o.	2.1 (n=12)	na	
2.2 Resultados de CUAL y CUAN logrados de forma interdependiente	1.2 (n=7) y 2.2	2.2 (n=18)	1.1 (n=2) y 2.2	
2.3 Caso especial: Divergencia	1.1 y 2.3 (n=2) 1.2 y 2.2 y 2.3 (n=1)	2.1 y 2.3 (n=2)		
3. ASIMILACIÓN DE DATOS				
3.1 Datos de CUAL a CUAN	1.1 (n=1) y 3.1	2.2 y 3.1 (n=2) 3.1 (n=1)		3.1 (n=1)
3.2 Datos de CUAN a CUAL		3.2 (n=1)		
3.3 Fusión de datos de CUAL y CUAN	1.1 y 2.2 y 3.3 (n=1) 1.2 y 3.3 (n=1)	2.2 (n=2) y 3.3		

4. ESTRATEGIAS EMERGENTES				
(...)				
Número de combinaciones observadas	4 (n=17)	3 (n=38)	2 (n=4)	1 (n=1)

**Los casos especiales no están incluidos en el recuento de combinaciones plausibles.*

Nota: El símbolo “y” se utiliza para representar combinaciones con más de una estrategia.

Integración 1. Conexión de fase

Estrategia 1.1 Conectar una fase CUAL a una fase CUAN

Fenenga y colaboradores utilizaron 42 entrevistas con familias y 7 entrevistas con informantes clave (Fase 1: datos CUAL), y un cuestionario estructurado completado por una muestra aleatoria representativa de familias de 64 regiones (Fase 2: datos CUAN; n=1903), para analizar la influencia de las relaciones sociales en las percepciones, el comportamiento y los procesos de toma de decisiones de las familias con respecto a su intención de afiliarse al plan nacional de seguro médico de Ghana (Fenenga y otros, 2015). Los resultados de los análisis de los datos de CUAL (teorización anclada) informaron (conexión de fase) el desarrollo del cuestionario para la recolección de datos de CUAN en la segunda fase del estudio (encuesta transversal). Al hacerlo, los investigadores desarrollaron preguntas ancladas en la realidad de las familias (validez del contenido ecológico). Los resultados de los análisis de los datos CUAN (regresión logística) permitieron entonces generalizar estadísticamente los resultados CUAL, es decir, la significación y la probabilidad de las asociaciones entre las variables estudiadas. Además de la conexión de fase, se compararon los resultados de CUAL y CUAN (véase más abajo: “Caso especial de las estrategias 2.1 y 2.2”), ilustrando una combinación de estrategias (conexión de fase y comparación de resultados).

Estrategia 1.2 Conexión de una fase CUAN a una fase CUAL

De Allegri y colaboradores estudiaron los factores asociados con la toma de decisiones de los hombres que viven en las zonas rurales de Burkina Faso acerca de la posibilidad de someterse a pruebas de detección del virus del SIDA (De Allegri y otros, 2015). Una muestra representativa de hombres adultos completó un cuestionario estructurado validado (n=1058). Los resultados de los análisis de los datos CUAN de la Fase 1 del estudio (estadísticas descriptivas y análisis multivariantes) informaron la estrategia de muestreo intencional para la planificación de la Fase 2 del estudio CUAL. En esta fase se realizaron entrevistas semiestructuradas a 38 hombres (19 sometidos a prueba y 19 no sometidos a prueba) reclutados según edad y región. Los resultados de los análisis descriptivos (inductivos) de los datos CUAL permitieron una mejor comprensión e interpretación de los resultados de la Fase 1 CUAN.

Caso especial de las estrategias 1.1 y 1.2: Monitoreo de un “rastreo de migas de pan” entre las fases CUAL y CUAN

El uso de esta estrategia se ilustra en un estudio realizado por Boot y colaboradores (Boot et al. 2016). El objetivo de los investigadores era comprender mejor las diferencias y similitudes entre los factores asociados a la participación en el trabajo de los trabajadores holandeses de 58 a 65 años con al menos uno de los tres problemas de salud crónicos: depresión, enfermedades cardiovasculares y osteoartritis. Los datos de CUAN provienen de un estudio de cohorte de población (Longitudinal Aging Study Amsterdam; n=1002). Los investigadores realizaron primero análisis estadísticos (pruebas t y pruebas de chi-cuadrado) para probar sus hipótesis iniciales y comparar los resultados entre cuatro grupos de trabajadores (depresión vs. enfermedad cardiovascular vs. osteoartritis vs. una combinación de al menos dos de estas condiciones). Esto condujo al desarrollo de una guía de entrevistas para el componente CUAL del estudio que exploró la posibilidad de factores emergentes (n=14). Los resultados del análisis de datos CUAL (temáticos) permitieron a los investigadores descubrir y proponer nuevas variables (predictores emergentes – no previstos inicialmente). Esas variables (rastreo de migas de pan) se utilizaron para generar nuevas hipótesis y realizar análisis estadísticos adicionales.

Integración 2. Comparación de los resultados

Estrategia 2.1. Comparación de los resultados obtenidos por separado

Esta estrategia fue utilizada por Pfadenhauer y Rehfuess en un estudio para evaluar y mejorar el comportamiento de las personas en relación con la calidad del agua, el saneamiento y el lavado de manos en las escuelas de una región de Filipinas (Pfadenhauer y Rehfuess 2015). En el estudio se evaluaron las repercusiones de esos comportamientos en la aparición de la diarrea infantil, teniendo en cuenta la influencia de los factores socioculturales y ambientales. Para ello, los investigadores analizaron (regresión logística) los datos CUAN de una encuesta nacional transversal (muestra representativa de la población; n = 12.469 hogares). Al mismo tiempo, llevaron a cabo seis grupos de discusión con mujeres que viven en zonas urbanas y rurales de la región. Los datos de CUAL se analizaron de manera inductiva (análisis temático) para producir un marco conceptual de los factores asociados a los comportamientos en cuestión.

Los datos de CUAN y CUAL fueron analizados y presentados por separado. En la sección “Discusión” de su trabajo, los investigadores compararon los resultados de CUAN y CUAL (matriz de resultados) e interpretaron esta comparación como metatemas comunes a los resultados de CUAL y CUAN, identificando similitudes (completas/parciales), diferencias (presencia/ausencia) y discrepancias entre los resultados de CUAN y CUAL (véase “Caso especial de las estrategias 2.1 y 2.2” más adelante).

Estrategia 2.2. Comparación de los resultados obtenidos de forma interdependiente

Oguntunde y colaboradores publicaron un ejemplo de esta estrategia en el contexto de un estudio para examinar los factores (facilitadores y barreras) asociados con el uso del sulfato de magnesio para tratar a las mujeres con preeclampsia (hipertensión grave relacionada con el embarazo) en los centros de salud de Nigeria septentrional (Oguntunde y otros, 2015). En una muestra aleatoria de 80 centros, los investigadores llevaron a cabo una encuesta transversal de CUAN a 80 clínicos (uno por centro) sobre el tratamiento de las mujeres con preeclampsia (protocolos de conocimiento y

atención, y recursos humanos y materiales disponibles). Además, llevaron a cabo una investigación descriptiva de CUAL y entrevistaron a una muestra intencionada (por ejemplo, un centro rural/urbano) de 30 gestores (uno por centro). Los datos de CUAN se analizaron con estadísticas descriptivas, mientras que los datos de CUAL se analizaron temáticamente (siendo los temas los factores estudiados). En la sección de resultados, los investigadores compararon los resultados de CUAN y CUAL sistemáticamente para cada factor (grupo temático). Para la mayoría de los factores, los resultados se presentan de tal manera que los resultados CUAL explican y complementan los resultados CUAN. Para otros cinco factores, los resultados presentados son sólo CUAN (n=4) o CUAL (n=1). Según los investigadores, esta comparación les permitió comprender mejor los factores estudiados y proponer intervenciones adecuadas.

Caso especial de las estrategias 2.1 y 2.2: Divergencia de los resultados de CUAL y CUAN

Pfadenhauer y Rehfuss (véase la “estrategia 2.1” supra) dicen que crearon una “matriz de convergencia” (no presentada en su documento) en la que prestaron atención específica a las discrepancias entre los resultados de CUAN y CUAL. En el artículo, estas divergencias parecen reconciliarse en la forma de interpretación de los resultados de CUAN por los resultados de CUAL (matices). De manera similar, Fenenga y colaboradores (véase la “estrategia 1.1” anterior) compararon sus resultados de CUAL y CUAN en una tabla y presentaron en detalle la coherencia entre estos resultados y una discrepancia. Contrariamente a lo que sugerían los resultados de CUAL, la participación de los participantes en los grupos de acción social no estaba estadísticamente asociada con la intención de adherirse al plan nacional de seguro médico de Ghana. Los investigadores propusieron conciliar esta discrepancia sugiriendo que la acción social/comunitaria rara vez se centraba en la salud en este contexto.

Integración 3. Asimilación de datos

Estrategia 3.1. Transformación de datos CUAL en datos CUAN (cuantificación)

Valentijn y colaboradores utilizaron esta estrategia en un estudio de los procesos de colaboración y la eficacia percibida de los proyectos de atención integrada en Holanda (Valentijn et al. 2015). En 42 proyectos, estos dos aspectos fueron evaluados en tres niveles: sistémico, organizativo y profesional. A nivel del sistema, los evaluadores entrevistaron a los coordinadores de los proyectos y a los interesados (datos de CUAL). A nivel organizativo y profesional, las medidas se basaron en cuestionarios estructurados completados respectivamente por los miembros de los comités directivos del proyecto y por los profesionales implicados (datos de CUAN). Los investigadores utilizaron un procedimiento válido y fiable (análisis temático seguido de un análisis de contenido cuantitativo) para transformar los datos CUAL en datos CUAN. Estos últimos (datos cuantificados de CUAL) se utilizaron luego con los demás datos de CUAN para realizar análisis estadísticos (agrupación jerárquica) y comparar los procesos de colaboración y la eficacia percibida de los proyectos de atención integrada según los tres niveles previstos (sistémico, organizativo y profesional).

Estrategia 3.2. Transformación de los datos CUAN en datos CUAL (cualitativos)

Reichwein y colaboradores emprendieron un estudio piloto para planificar un programa de promoción de la salud en la planificación familiar (Reichwein et al. 2015). Concretamente, los investigadores trataron de identificar a los usuarios que más necesitaban métodos anticonceptivos en Uganda y de determinar cuáles eran los métodos más eficaces. Específicamente, los investigadores se propusieron identificar a los usuarios que más necesitaban métodos anticonceptivos en Uganda y Vietnam para orientar el programa en beneficio de ellos. Los investigadores analizaron estadísticamente (agrupación jerárquica) los datos de CUAN recogidos en encuestas nacionales transversales e identificaron tres perfiles de usuarios potenciales: mujeres ugandesas que querían prolongar el tiempo entre

embarazos, hombres ugandeses que querían limitar el número de hijos que tenían y mujeres vietnamitas jóvenes sexualmente inactivas. Para cada uno de estos grupos, se recogieron datos de CUAL (entrevistas individuales y grupos de discusión) y se analizaron temáticamente. A continuación, los investigadores asimilaron los datos CUAN transformados en perfiles narrativos y los datos CUAL (tipos de relatos de casos) para crear dos arquetipos biográficos (dos ideales típicos) en los que centrar la planificación del programa: “Kibuuka”, un agricultor semianalfabetizado de 52 años de edad que vive en la Uganda rural, y “Anh”, una estudiante de 20 años de edad que vive en Hanoi, la capital de Vietnam. Las biografías ideales típicas así desarrolladas personifican las características y creencias de los dos grupos objetivo del programa. La investigación cualitativa se utilizó para asimilar los perfiles estadísticos (cualitativos) y los datos CUAL en forma de arquetipos (tipos ideales).

Estrategia 3.3. Fusionando los datos de CUAL y CUAN

Johansson y colaboradores examinaron el efecto de las pruebas de detección en el uso de drogas en niños con fiebre en los países africanos donde la malaria es endémica (Johansson et al. 2015). En 12 países, reunieron datos de CUAN de los servicios nacionales de salud y de las encuestas sobre salud infantil (n=16.323) y analizaron estos datos para cada país (regresión logística). En seis de esos países, los investigadores utilizaron un diseño de estudio de casos múltiples (cada país es un caso) y reunieron datos de CUAL mediante documentos sobre programas de control del paludismo y entrevistas con siete expertos (en total) de los programas y países involucrados. Para cada país (caso), los resultados de CUAN y CUAL se agruparon en una matriz (los resultados de CUAL se utilizaron para desarrollar cuatro explicaciones plausibles principales – temas – para los resultados de CUAN). Se elaboró una meta-matriz que incluía todos los datos de todos los casos y que constituía de por sí una nueva base de datos que permitía evaluar los efectos estudiados en cada país (cada estudio de caso se presenta en un fichero adicional en el apéndice) además de revelar las similitudes y diferencias entre los países (comparación de casos).

Combinaciones de estrategias y estrategias emergentes

En la tabla 1 se presentan las múltiples combinaciones de estrategias observadas en esta muestra de 60 estudios de MM en el ámbito de la salud global. Todas las combinaciones posibles están representadas (1; 2; 3; 1 y 2; 1 y 3; 2 y 3; 1 y 2 y 3). Alrededor de la mitad de las combinaciones están representadas para las estimaciones secuenciales y convergentes. Debido a la escasez de citas multifacéticas y multiniveles, se observaron pocas combinaciones para estas citas. No se observaron estrategias o combinaciones emergentes (imprevistas).

Análisis reflexivo y lecciones aprendidas

Estos resultados ilustran el marco conceptual propuesto: tres tipos de integración, nueve estrategias específicas y siete combinaciones de estrategias para integrar las fases, resultados y datos de CUAL y CUAN en métodos mixtos. Aunque nuestra conceptualización y el sistema de vigilancia eSRAP nos permiten identificar estrategias emergentes, no hemos detectado ninguna hasta la fecha. La mayoría de los ejemplos que se presentan en este capítulo proceden de estudios de gestión de la movilidad en la salud global y demuestran el dinamismo de los investigadores en este campo. Este capítulo contribuye al conocimiento de la gestión de la movilidad proponiendo una conceptualización operativa (práctica) de los tipos de integración y estrategias específicas para integrar las fases, resultados y datos de CUAL y CUAN. En cuanto a las innovaciones, muestra la riqueza de combinaciones de estrategias de MM para cada cita (tabla 1).

Sólo se supervisó una muestra de estudios que utilizaban la MM en el PAR publicado en 2015, sin intentar ofrecer una descripción exhaustiva de la investigación sobre la salud global utilizando la MM. Aunque de alcance limitado, estos resultados muestran que el eSRAP ofrece la ventaja a una comunidad creciente de investigadores de la MM de mantenerse al día de las nuevas tendencias de manera prospectiva, compartiendo la carga de la selección de los estudios que utilizan la MM y ofreciendo una descripción detallada de este uso. Específicamente, eSRAP es adaptable a las necesidades individuales de los evaluadores, investigadores y educadores, permitiéndoles mantenerse al día en un tema de su elección sin agotarse. Por ejemplo,

cada miembro de la comunidad eSRAP MM puede beneficiarse en cualquier momento de los resultados de la labor prospectiva de todos los miembros. Para los educadores de MM, eSRAP permite con unos pocos clics antes de cada curso retener los estudios recientes como material de enseñanza para los estudiantes mediante funciones de “filtro”. Por ejemplo, el eSRAP permite encontrar rápidamente artículos recientes que ilustran en detalle cada estrategia.

En este capítulo se ha demostrado que un porcentaje sustancial de los estudios denominados MM corresponden en realidad a métodos CUAL (10%) o CUAN (10%). En 2016, todos los investigadores están de acuerdo en que MM combina los métodos CUAL y CUAN (Plano Clark e Ivankova 2016) y hemos incluido estudios en los que los investigadores “de forma convincente y rigurosa recogen y analizan datos CUAN y CUAL” (Creswell y Plano Clark 2011, p. 5). Este rigor depende de la experiencia de los investigadores, incluyendo su elección de (a) instrumentos de medición (validez y fiabilidad) y diseño (ensayo controlado aleatorio vs. no aleatorio vs. estudio descriptivo) para los datos CUAN y (b) interacciones de los participantes (credibilidad y contextualización) y diseño (estudio descriptivo o interpretativo, estudio de caso, teoría fundamentada, biografía, etnografía, etc.) para los datos CUAL (Creswell y Plano Clark 2011). Al aplicar este criterio a la definición de MM, excluimos uno de los cinco estudios de RAP denominado MM publicados en 2015 ($39/187 = 20,9\%$ en total; $6/35 = 17,1\%$ en salud global).

Estos estudios excluidos pueden interpretarse como una señal del reconocimiento de la MM en la comunidad científica y una ilustración de los extremos del continuo entre los métodos CUAL y CUAN. En primer lugar, el reconocimiento recientemente obtenido de la gestión de la movilidad sugiere que el escaso número de personas revisoras y editoras todavía permite en 2015 publicar estudios denominados “métodos mixtos” que no cumplen los criterios de la definición más frecuentemente citada y aceptada por la comunidad de investigadores, incluidos los miembros de la Asociación Internacional de Investigación sobre Métodos Mixtos (MMIRA) (Creswell y Plano Clark 2011; Johnson, Onewuegbuzie y Turner 2007). En otras palabras, la proporción sustancial de estos estudios sugiere que la MM es atractiva para los investigadores/as y editores/as orientados al paciente (sesgo de publicación positivo).

En segundo lugar, el continuo entre los métodos CUAL y CUAN se centra en MM (Figura 2) (Johnson et al. 2007; Niglas 2010). Nuestros resultados

sugieren que entre los métodos CUAL y MM hay búsquedas CUAL denominadas MM en las que los participantes se describen con números (entre otra información). Se excluyeron 19 estudios por esta razón (presencia de información cuantitativa, pero ausencia de datos de la CNA que cumplan los criterios de rigor mínimo arriba mencionados). Al otro lado del continuo entre los métodos MM y CUAN se encuentran las encuestas CUAL denominadas como MM con unas pocas palabras en un recuadro de comentarios opcional al final de un cuestionario anónimo autoadministrado. Se excluyeron 20 estudios por este motivo (presencia de información cualitativa, pero ausencia de datos CUAL que cumplan los criterios de rigor mínimo).

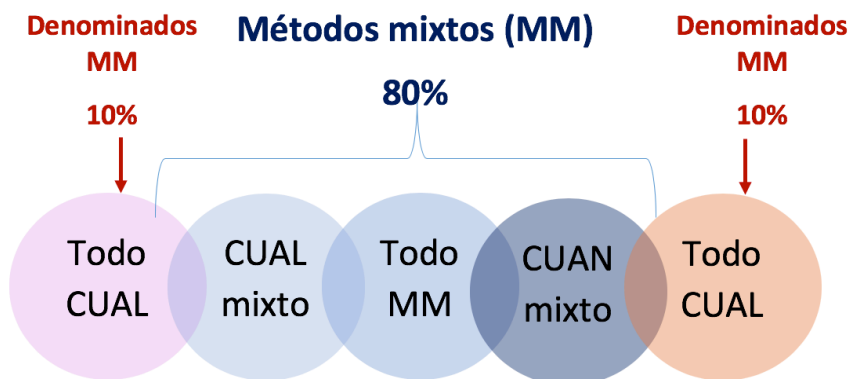


Figura 2. Investigación orientada al paciente en 2015: Estudios cualitativos (CUAL) o cuantitativos (CUAN) denominados como Métodos mixtos (MM) sin cumplir la definición de MM (Fuente: Johnson et al. 2007)

En porcentaje de estudios denominados MM que no cumplieron los criterios de la definición habitual de MM debería disminuir si las instrucciones a los autores y autoras de las revistas (y las tablas de evaluación de los/as revisores) incluyeran recomendaciones para informar sobre las indicaciones del estudio GRAMMS (Good Reporting of A Mixed Methods Study) (O’Cathain, Murphy y Nicholl 2008). Recordamos aquí que las seis recomendaciones del GRAMMS que se deben describir como mínimo:

1. el objetivo y la especificación de MM;
2. el muestreo, la recogida y el análisis de los datos de CUAL y CUAN;

3. la integración de las fases, resultados y datos de CUAL y CUAN;
4. una justificación para el uso de MM;
5. una descripción de los beneficios asociados a la integración de los métodos;
6. una descripción de las limitaciones de esta integración.

Finalmente, ocho años después de la publicación de las seis recomendaciones de la GRAMMS (para informar adecuadamente sobre los estudios que usan MM), nuestros resultados son desalentadores. Más de la mitad de los estudios de MM en la PAR ($76/136 = 55,9\%$ en toda la muestra; $18/29 = 62,1\%$ en salud global) no cumplen las tres primeras recomendaciones de la GRAMMS (recomendaciones 1-2-3), por ejemplo, debido falta de descripción de CUAL, CUAN y MM. Muy pocos documentos implementan las seis recomendaciones del GRAMMS (recomendaciones 1-2-3-4-5-6): seis ($6/136 = 4,4\%$) en toda la muestra, incluyendo una ($1/29 = 3,4\%$) en salud global (Akelo et al., 2015). Esto sugiere las responsabilidades conjuntas de las personas autoras, revisoras, editoras, publicaciones y organizaciones internacionales como la Asociación Internacional para la Investigación de Métodos Mixtos (MMIRA) para promover la aplicación de estas recomendaciones teniendo en cuenta las especificidades de las disciplinas y las revistas. Como todo trabajo científico, los MM deben ser explícitos, transparentes y reproducibles. Concretamente, una mejor descripción de los MM será útil para ayudar a los estudiantes a comprender mejor los métodos y estrategias utilizados para integrar las fases, resultados y datos de CUAL y CUAN.

Conclusión

Este capítulo recuerda que los métodos mixtos suelen requerir la colaboración de personas expertas en métodos CUAL y CUAN (por ejemplo, una persona experta en CUAL y otra en CUAN para los estudiantes de doctorado) que deben conciliar sus interpretaciones de las ciencias cuando éstas difieren y garantizar el rigor metodológico de los aspectos CUAL y CUAN. Además, mencionamos seis recomendaciones mínimas para informar bien sobre un estudio utilizando MM y proponemos tres tipos de integración, nueve estrategias específicas de MM y un mundo completamente nuevo para explorar con siete combinaciones de estrategias para integrar las fases,

resultados y datos de CUAL y CUAN. Esta conceptualización contribuye al conocimiento de la MM y pretende ser práctica. Puede guiar a los estudiantes e investigadores a planificar, realizar y evaluar estudios y evaluaciones de programas utilizando la MM, especialmente en la salud global.

Referencias clave

Méthodes mixtes francophonie :

<http://methodesmixtesfrancophonie.pbworks.com>

Esta nueva asociación internacional sin ánimo de lucro tiene como objetivo promover los métodos mixtos en francés. En 2018, ya cuenta con 121 miembros de 8 países diferentes (Alemania, Bélgica, el Canadá, Francia, Rwanda, el Senegal y Suiza). Ofrece apoyo metodológico, una bibliografía abierta que contiene las publicaciones de los miembros sobre métodos mixtos y una lista de expertos, recursos y capacitación. En particular, ofrece acceso gratuito a la primera obra colectiva de referencia metodológica sobre métodos mixtos en francés (14 capítulos, 266 páginas): Bujold, M., Hong, Q. N., Ridde, V., Bourque, C. J., Dogba, M. J., Vedel, I. y Pluye, P. (2018). Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé. Montréal : 117e Cahiers scientifiques de l'ACFAS.

Onwuegbuzie, A. J. et Combs, J. P. (2010). Emergent data analysis techniques in mixed methods research: A synthesis. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (p. 397-430). Thousand Oaks, CA: Sage Publications.

Es una imagen general de los análisis de métodos mixtos a lo largo de 13 dimensiones que incluyen, por ejemplo, la orientación de los análisis (casos, variables, procesos) y los vínculos entre los análisis y otros aspectos metodológicos (filosofía, objetivos y formas de los datos). Permite a los evaluadores, estudiantes e investigadores que utilizan métodos mixtos situar las estrategias analíticas propuestas e ilustradas en este capítulo sobre la salud global dentro de un marco más general.

Système de veille collaborative des tendances en recherche (eSRAP) : www.esrap.ca

Este sitio web es producido por l'Unité de soutien SRAP du Québec que apoya la investigación orientada al paciente. El eSRAP es una plataforma de supervisión en colaboración diseñada para ayudar a las comunidades interesadas en un tema (incluidos investigadores, clínicos, evaluadores, gestores y socios pacientes/ciudadanos) a mantenerse al día de las últimas publicaciones científicas sobre ese tema. El eSRAP facilita la selección de las publicaciones pertinentes y el acceso a las publicaciones seleccionadas. El eSRAP tiene por objeto optimizar el intercambio de esfuerzos y conocimientos de vigilancia y, en última instancia, democratizar la vigilancia, por ejemplo, en los métodos mixtos de este capítulo. Una versión de autoservicio estará disponible en 2019, y podrá apoyar a cualquier comunidad interesada en un tema de salud global.

Referencias

- Akelo, V., McLellan-Lemal, E., Toledo, L., Girde, S., Borkowf, C. B., Ward, L., ... Thomas, T. K. (2015). Determinants and experiences of repeat pregnancy among HIV-positive Kenyan women: a mixed-methods analysis. *PloS one*, 10(6), e0131163.
- Boot, C. R., Kruif, A. T. C., Shaw, W. S., Beek, A. J., Deeg, D. J. et Abma, T. (2016). Factors important for work participation among older workers with depression, cardiovascular disease, and osteoarthritis: a mixed method study. *Journal of occupational rehabilitation*, 26(2), 160-172.
- Creswell, J. W. et Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2^e éd.). Thousand Oaks, CA: Sage Publications.
- De Allegri, M., Agier, I., Tiendrebeogo, J., Louis, V. R., Yé, M., Mueller, O. et Sarker, M. (2015). Factors affecting the uptake of HIV testing among men: a mixed-methods study in rural Burkina Faso. *PloS one*, 10(7), e0130216.
- Fenenga, C. J., Nketiah-Amponsah, E., Ogink, A., Arhinful, D. K., Poortinga, W. et Hutter, I. (2015). Social capital and active membership in the Ghana National Health Insurance Scheme-a mixed method study. *International journal for equity in health*, 14(1), 1.
- Johansson, E. W., Gething, P. W., Hildenwall, H., Mappin, B., Petzold, M., Peterson, S. S. et Selling, K. E. (2015). Effect of diagnostic testing on

- medicines used by febrile children less than five years in 12 malaria-endemic African countries: a mixed-methods study. *Malaria journal*, 14(1), 1.
- Johnson, R. B., Onewuegbuzie, A. J. et Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1(2), 112-133.
- Niglas, K. (2010). The multidimensional model of research methodology: an integrated set of continua. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2^e éd., p. 215-236). Thousand Oaks, CA: Sage Publications.
- O'Cathain, A. (2010). Assessing the quality of mixed methods research: towards a comprehensive framework. Dans A. Tashakkori et C. Teddlie (dir.), *Handbook of mixed methods in social and behavioral research* (2^e éd., p. 531-555). Thousand Oaks, CA: Sage Publications.
- O'Cathain, A., Murphy, E. et Nicholl, J. (2008). The quality of mixed methods studies in health services research. *Journal of Health Services Research and Policy*, 13(2), 92-98.
- Oguntunde, O., Charyeva, Z., Cannon, M., Sambisa, W., Orobato, N., Kabo, I. A., ... et Sadauki, H. (2015). Factors influencing the use of magnesium sulphate in pre-eclampsia/eclampsia management in health facilities in Northern Nigeria: a mixed methods study. *BMC Pregnancy et Childbirth*, 15(1), 1.
- Pfadenhauer, L. M. et Rehfuss, E. (2015). Towards effective and socio-culturally appropriate sanitation and hygiene interventions in the Philippines: a mixed method approach. *International journal of environmental research and public health*, 12(2), 1902-1927.
- Clark, V. L. P. et Ivankova, N. V. (2016). *Mixed methods research: a guide to the field* (vol.). Thousand Oaks, CA: Sage Publications.
- Pluye, P. (2012). Méthodes mixtes en santé mondiale. Dans J. C. Suárez-Herrera et M. J. Blain (dir.), *Regards croisés sur la socio-anthropologie réflexive et la recherche en santé mondiale* (p. 129-135). Montréal : 114^e Cahiers scientifiques de l'ACFAS.
- Reichwein, B., Wolmarans, L., Nantayi, L., Nassali, F., Kakinda, A., Musumba, D., ... et Baatsen, P. (2015). SegWeigh: a mixed-method approach to

segmenting potential contraceptive user groups and meeting Family Planning 2020 goals. *International Journal of Gynecology et Obstetrics*, 130, E8-E14.

Tang, D., Pluye, P. et Bouthillier, F. (2015). eSRAP : système de surveillance de tendances en recherche permettant la collaboration des membres de communautés en recherche axée sur le patient (RAP) pour évaluer et partager les résultats et les projets de recherche de manière adaptée. Dans M. Bujold, Q. N. Hong, V. Ridde, C. J. Bourque, M. J. Dogba, I. Vedel et P. Pluye (dir.), *Oser les défis des méthodes mixtes en sciences sociales et sciences de la santé* (p. 40). Montréal : 117^e Cahiers scientifiques de l'ACFAS.

Valentijn, P. P., Ruwaard, D., Vrijhoef, H. J., de Bont, A., Arends, R. Y. et Bruijnzeels, M. A. (2015). Collaboration processes and perceived effectiveness of integrated care projects in primary care: a longitudinal mixed-methods study. *BMC health services research*, 15(1), 1.

Agradecimientos

Pierre Pluye tiene una beca de investigación del *Fonds de recherche du Québec en santé* (FRQS). El desarrollo del sistema eSRAP cuenta con el apoyo de la Unité Soutien SRAP du Québec (componente de desarrollo metodológico).

Los autores desean expresar su agradecimiento a los miembros fundadores de la comunidad eSRAP-MM que pusieron a prueba el prototipo del eSRAP, incluidos todos los estudiantes y profesores nuevos y antiguos de los cursos WMDF-672 “Applied Mixed Methods in Health Research”, WMDF-600 “Mixed Studies Reviews” y WMDF 608 “Advanced Mixed Methods” que se han ofrecido desde 2008 en el Departamento de Medicina Familiar de la Universidad McGill.

Resumen / Résumé / Abstract

Los métodos mixtos (MM) son cada vez más populares. Al utilizar los métodos mixtos, los investigadores integran métodos cualitativos (QUAL) y cuantitativos (QUAN) con respecto a las cosmovisiones, metodologías y diseños, preguntas de investigación, técnicas de recopilación y análisis de

datos y resultados de estudios. Se han propuesto varias estrategias de integración, pero su conceptualización suele estar impulsada por el diseño, o ser fragmentada, o no haber sido probada empíricamente. Esto es un reto para la planificación y realización de estudios que utilizan métodos mixtos y para la formación de estudiantes de postgrado. Basado en la literatura metodológica, este capítulo presenta un marco conceptual que incluye tipos de integración, estrategias prácticas y posibles combinaciones de estrategias para integrar las fases, resultados y datos de QUAL y QUAN. En el siguiente capítulo, este marco se prueba utilizando una revisión de los estudios que utilizan métodos mixtos de 2015 con una descripción detallada del método. Esto apoya el marco, ilustra múltiples tipos de combinaciones y proporciona ejemplos de salud mundial. Estos dos capítulos complementarios contribuyen a avanzar en el conocimiento metodológico de los métodos mixtos a través de (a) un llamado a informar mejor sobre los estudios que utilizan métodos mixtos en artículos científicos, y (b) una conceptualización probada que comprende tres tipos de integración y nueve estrategias específicas, que explican las posibilidades actuales y futuras de combinar estrategias para integrar las fases, los resultados y los datos de QUAL y QUAN.

Les méthodes mixtes (MM) sont de plus en plus populaires. En utilisant les MM, les chercheuses et chercheurs intègrent des méthodes qualitatives (QUAL) et quantitatives (QUAN), par exemple au niveau des questions et devis de recherche, et des collectes et analyses des données. Plusieurs stratégies d'intégration ont été proposées, mais leur conceptualisation est généralement limitée, fragmentée, ou non testée scientifiquement. Cela constitue un défi pour planifier et effectuer des études utilisant les MM, et pour enseigner les MM. Fondé sur la littérature méthodologique, le chapitre précédent a présenté un cadre conceptuel incluant tous les types d'intégration connus, des stratégies pratiques et des combinaisons possibles de stratégies pour intégrer les phases, les résultats et les données QUAL et QUAN. Dans ce chapitre, ce cadre est testé avec une revue des études utilisant les MM. Cette revue illustre plusieurs types de combinaisons avec des études en santé mondiale. Ensemble, ces deux chapitres contribuent à faire avancer les connaissances méthodologiques par (a) un appel à une meilleure description des MM dans les publications scientifiques, et (b) une

conceptualisation testée proposant trois types d'intégration et neuf stratégies spécifiques, et toutes les combinaisons possibles des stratégies pour intégrer les phases, résultats et données QUAL et QUAN.

Mixed methods (MM) are increasingly popular. In performing MM, researchers integrate qualitative (QUAL) and quantitative (QUAN) methods with respect to worldviews, methodologies and designs, research questions, data collection and analysis techniques, and study results. Several integration strategies have been proposed, but their conceptualization is usually design-driven, or fragmented, or not empirically tested. This is challenging for planning and conducting MM studies, and for training graduate students. Based on the methodological literature, the previous chapter presented a conceptual framework including types of integration, practical strategies, and possible combinations of strategies for integrating QUAL and QUAN phases, results and data. In the present chapter, this framework is tested using a review of 2015 MM studies with a method-detailed description. This supports the framework, illustrates multiple types of combinations, and provides global health examples. These two companion chapters contribute to advance methodological knowledge on MM via (a) a call for better reporting MM studies in scientific articles, and (b) a tested conceptualisation comprising three types of integration and nine specific strategies, which explain current and future possibilities for combining strategies to integrate QUAL and QUAN phases, results, and data.

Pierre Pluye es profesor titular del Departamento de Medicina Familiar de la Universidad McGill, investigador principal del Fondo de Investigación de Quebec Salud (FRQS) y director del componente de Desarrollo Metodológico de la Unidad SOUTIEN SRAP de Quebec. Es miembro asociado de la Escuela de Ciencias de la Información de la Universidad McGill. Es especialista en medicina familiar y salud pública. Es doctor en promoción de la salud (Universidad de Montreal). Ha recibido premios a su carrera de los Institutos Canadienses de Investigación de la Salud (CIHR) y del FRQS. En 2017 fue elegido miembro de la Academia Canadiense de Ciencias de la Salud y recibió el premio "Investigador del Año" del Colegio de Médicos de

Familia del Canadá. Tiene experiencia en métodos mixtos. Su investigación estudia los efectos de la información de los recursos electrónicos en los profesionales, los pacientes y el público.

Enrique García Bengoechea es investigador de la Facultad de Educación y Ciencias de la Salud de la Universidad de Limerick (Irlanda). Sus principales intereses de investigación son la promoción de la actividad física y la salud en el contexto comunitario y el desarrollo y la socialización de los jóvenes en el deporte. Anteriormente, Enrique trabajó como investigador en el Centro de Vida Activa de Alberta, enseñó en las universidades de McGill y Western Sydney, y fue asesor de investigación de métodos mixtos en la Plataforma de Desarrollo Metodológico de la Unidad de Apoyo del PASR de Quebec. También fue investigador visitante en el Instituto de Deporte, Ejercicio y Vida Activa (ISEAL) de la Universidad de Victoria en Melbourne. Enrique es actualmente miembro del equipo de investigación del estudio BuiltEnvironmentand Active Transport to School (BEATS), con sede en la Universidad de Otago. También está afiliado al Grupo de Investigación de Educación Física y Promoción de la Actividad Física de la Universidad de Zaragoza.

David Li Tang es Técnico de Desarrollo de Servicios del Consejo Nacional de Investigaciones de Canadá y Profesor Adjunto del Departamento de Medicina Familiar de la Universidad McGill. Su área de trabajo se centra en la innovación. David y colaboradores diseñan y desarrollan, entre otras cosas, infraestructuras y metodologías para estimular la innovación al más amplio nivel colectivo, involucrando a agencias gubernamentales, universidades, industrias, asociaciones profesionales e investigadores individuales. Tienen repercusiones en los sistemas de innovación y atención de la salud del Canadá. La investigación multidisciplinaria de David abarca las ciencias sociales, la ingeniería y las ciencias de la información.

Vera Granikov es especialista en información integrada en el Departamento de Medicina Familiar de la Universidad McGill. Proporciona servicios de información a medida a los investigadores y clínicos de atención primaria y coordina el sistema de vigilancia colaborativa financiado por la Unité SOUTIEN SRAP du Québec. Vera está cursando actualmente estudios de doctorado en la Escuela de Ciencias de la Información de la Universidad McGill y tiene un premio de doctorado del Fonds de recherche du Québec – Société et culture (FRQSC).

Cita

Pierre Pluye, Enrique García Bengoechea, David Li Tang y Vera Granikov (2020). La práctica de la integración en métodos mixtos. Las múltiples combinaciones de estrategias de integración. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 221-246. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

PARTE IV

EVALUACIÓN DE LA EFECTIVIDAD Y LA EFICIENCIA

9. Métodos cuasi-experimentales

El efecto de la edad mínima legal sobre el consumo de alcohol entre los y las jóvenes en los Estados Unidos

TARIK BENMARHIA Y DANIEL FULLER

Métodos cuasi-experimentales

Definición del método

El término “cuasiexperimental” se refiere a “un experimento con una intervención determinada, resultados de salud medidos y al menos dos unidades experimentales (grupo de intervención y control), pero sin el uso de la aleatoriedad para crear las condiciones de inferencia causal” (Cook, Campbell y otros, 1979).

Fortalezas del método

Los métodos cuasiexperimentales son un diseño de estudio alternativo a los ensayos controlados aleatorios, atractivo cuando estos últimos son difíciles o imposibles de realizar. De hecho, estos métodos también permiten utilizar el paradigma contrafactual de causalidad sobre la base de situaciones observadas en las que no hay asignación deliberada y aleatoria a los grupos de intervención y control.

Desafíos del método

El método cuasi-experimental requiere que se respeten hipótesis de trabajo adicionales. A veces estas suposiciones son imposibles de verificar empíricamente. Los efectos que se estiman pueden variar con respecto a las poblaciones involucradas y, como en los ensayos aleatorios controlados, es importante distinguir la validez interna y externa.

La evaluación de los resultados de las políticas públicas en la salud de las poblaciones es un ámbito que se ha desarrollado especialmente en los últimos años. La evaluación del impacto difiere de otros tipos de evaluaciones, como la evaluación de la implantación (Sabatier y Mazmanian 1980; Belaid y Ridde 2012) o la evaluación económica (Brousselle y Lessard 2011; Drummond, Sculpher y otros 2015, Brousselle, Benmarhnia y otros 2016). La evaluación de los resultados de las políticas públicas en la salud de las poblaciones es esencial por varias razones (Champagne, Contandriopoulos y otros, 2009; Macintyre, 2010). Ayuda a: i) asegurar que esas políticas contribuyan a mejorar el estado de salud de la población, que es uno de los mandatos de la salud pública, ii) asegurar que los objetivos de un programa de salud pública se cumplan realmente, iii) probar nuevas intervenciones y decidir si se mantienen, modifican o suspenden las existentes, y iv) asegurar que los recursos públicos se utilicen eficazmente.

En pocas palabras, el objetivo principal de este tipo de evaluación es atribuir cambios a una intervención concreta con respecto a un indicador de salud específico. En otras palabras, se trata de atribuir un efecto causal a una intervención sobre un indicador de salud en un contexto particular.

El concepto de inferencia causal

La noción de inferencia causal es fundamental para muchas disciplinas hoy en día, entre ellas la econometría, la psicología social y la epidemiología. Uno de los modelos más utilizados para identificar los efectos causales se basa en la noción contrafáctica (Susser 2001; Pearl 2009; Naimi y Kaufman 2015). Por ejemplo, para evaluar si los cambios en los resultados de salud son directamente atribuibles a una intervención, se compararían los cambios en los resultados de salud en un grupo observado (denominado “factual”) con un grupo hipotético similar o intercambiable (denominado “contrafactual”) que se distingue por una sola característica; no recibir la intervención. Así, el efecto causal se mediría por el contraste entre el grupo de hechos y el grupo contrafáctico. El problema fundamental de la causalidad (Holland 1986; Kaufman 2007) es que, por definición, este grupo contrafáctico no es directamente observable. Así pues, se ha propuesto la aleatorización como solución a este problema, manteniendo al mismo tiempo la intercambiabilidad y la sustitución entre los grupos de intervención y los de no intervención, y asegurando que no se violen las hipótesis que permiten la inferencia causal, de modo que la intervención sólo pueda afectar a los resultados de salud del grupo donde se efectúa, por ejemplo. En particular, la aleatorización permite superar la cuestión crucial de tener en cuenta factores de confusión que pueden o no medirse (Bonell, Hargreaves et al. 2011). Así pues, se han introducido y utilizado ensayos controlados aleatorios (ECA) en salud y en ciencias sociales: un grupo se asigna al azar a la intervención (o al tratamiento) y un grupo no. Esta asignación aleatoria garantiza que no haya diferencias entre los dos grupos antes de la intervención, es decir, que no haya ningún sesgo de confusión, medido o no medido. Sin embargo, este tipo de experimentación, en la que se elige deliberadamente un grupo para intervenir y no el otro (el grupo de control), no es factible en muchas situaciones, ni por el costo ni por razones éticas (Royall 1991; Hawe, Shiell y otros 2004; Moore y Moore 2011; Petticrew, Chalabi y otros 2011).

Los “experimentos naturales” son un excelente diseño de estudio alternativo a los ECA porque permiten utilizar el paradigma contrafáctico

para estimar la causalidad sobre la base de las situaciones observadas y sin una asignación deliberada y aleatoria de los grupos de intervención y control (Bor 2016). Para ello se han desarrollado métodos cuasiexperimentales (MCE) para diseñar grupos contrafácticos basados únicamente en datos observados. El término “cuasiexperimental” se refiere a un “experimento con una intervención determinada, resultados de salud medidos y dos unidades experimentales (grupos de intervención y control), pero sin utilizar la aleatoriedad para crear las condiciones para la inferencia causal” (Cook, Campbell et al. 1979: 14; Shadish y Cook 2009). De esta manera, el uso de los MCE puede imitar los resultados que se producirían en un ECA, mientras que sólo se utilizan las situaciones existentes. Se han desarrollado varios MEC para evaluar los efectos de la política pública y están empezando a utilizarse en la salud pública.

Alcance del capítulo

El propósito de este capítulo es presentar dos de los principales métodos cuasiexperimentales (MCE) y su aplicación en un estudio para evaluar el efecto de la edad mínima legal para el consumo de alcohol en los y las jóvenes. Estos métodos cuasiexperimentales están concebidos para tener en cuenta los factores de confusión no medidos que pueden tener un impacto significativo en la validez de las previsiones. Los métodos que sólo permiten tener en cuenta los factores de confusión medidos, como los métodos de comparación que utilizan *propensity score*, por ejemplo, se describen a veces como MCE (Shadish, Luellen et al. 2006; Austin 2011). En este capítulo, sin embargo, no los discutiremos. En su lugar, nos centraremos en dos tipos de MCE: el método de “diferencia en diferencias” (DD), y un tipo particular de variable instrumental, el enfoque de “regresión de discontinuidad” (RD).

Las variables instrumentales para evaluar el impacto de las políticas públicas se utilizan comúnmente en la econometría (Newhouse y McClellan 1998; Hernán y Robins 2006), donde se utiliza una terminología diferente. En pocas palabras, una variable instrumental es una variable que estará altamente correlacionada con la probabilidad de beneficiarse de una intervención, pero no estará relacionada con el resultado sanitario de interés. Así pues, esta variable instrumental puede utilizarse para abordar los posibles sesgos de confusión (medidos o no medidos) en la relación entre

la probabilidad de beneficiarse de la intervención y los posibles cambios en el resultado sanitario de interés (véase Angrist y Pischke 2008 y 2014). En particular, las variables geográficas (límites, altitud, precipitaciones, etc.) o genéticas pueden utilizarse como variables instrumentales. No discutiremos el caso general de las variables instrumentales en este capítulo. Del mismo modo, no discutiremos los análisis de series temporales continuas como un método de evaluación de los efectos de las políticas públicas (véase Fuller, Sahlqvist y otros, 2012; Benmarhnia, Zunzunegui y otros, 2014; Kontopantelis, Doran y otros, 2015).

El objetivo general de este capítulo es describir dos MCE, aplicarlos empíricamente a un ejemplo basado en la simulación y proporcionar todos los elementos, incluidas las hipótesis preliminares que se han de probar, los diversos pasos que se han de dar y la sintaxis para realizar los análisis con los programas R y Stata. Esto permitirá a los/as lectores/as reproducir este tipo de análisis en su contexto de evaluación.

Todos estos recursos informáticos, así como la sintaxis para aplicarlos a otras cuestiones de investigación, pueden encontrarse en línea en https://github.com/walkabilly/methode_evaluation ¡Aprovechadlo!

Estableciendo el contexto para la evaluación: Intervenciones para reducir el consumo de alcohol entre los y las jóvenes

En este capítulo, utilizaremos el cambio de la edad mínima legal para el consumo de alcohol, como ejemplo de política pública para reducir el consumo de alcohol entre los y las jóvenes.

El consumo de alcohol contribuye de manera significativa a la carga de morbilidad y mortalidad entre los/as adolescentes y adultos jóvenes en muchos países (Rehm, Gmel y otros, 2003). Hay muchas intervenciones destinadas a prevenir el consumo de alcohol entre los y las jóvenes. Durante muchos decenios, varios países han establecido una edad mínima legal para beber (Carpenter y Dobkin 2011). A pesar de este marco jurídico y de muchas

otras intervenciones de salud pública destinadas a reducir el consumo de alcohol entre los y las jóvenes, el alcohol sigue siendo un factor importante de mortalidad. Así pues, varios países han decidido aumentar la edad legal de consumo de alcohol con el fin de reducir el consumo de alcohol y prevenir los accidentes atribuibles al consumo. En los últimos años se han publicado algunos estudios que evalúan este tipo de política pública con MCE de tipo DD o RD (Donald y Lang 2007; Carpenter y Dobkin 2009; Yörük y Yörük 2012). Presentaremos cómo aplicar dos enfoques sobre una base de datos simulados, para estimar los efectos de este tipo de política pública sobre el consumo de alcohol entre los y las jóvenes.

Presentación y aplicación de los métodos propuestos

Método de “diferencia en diferencias”

Visión general del método DD

El método de diferencias en diferencias (DD) utiliza uno o más grupos de control para aproximar la trayectoria de un grupo contrafáctico que no recibió una intervención que sería intercambiable con el grupo que recibió la intervención. Por ejemplo, el grupo o grupos de control podrían ser un país o región que nunca ha recibido la intervención o que la recibió en un momento distinto del grupo de intervención. El principal supuesto que debe tenerse en cuenta en este tipo de enfoque es que, en ausencia de la intervención, las tendencias del grupo de control son un buen indicador de las tendencias del grupo que recibirá la intervención (Angrist y Pischke 2008; Dimick y Ryan 2014). En otras palabras, si no se hubiera producido la intervención, las tendencias entre los dos grupos con respecto al indicador de interés se mantendrían constantes.

Consideremos esquemáticamente dos Estados en el mismo país, con cuatro períodos de observación (2009 [t0], 2010 [t1], 2011 [t2], 2012 [t3]), un Estado que realiza la intervención en 2010 (t1) y el otro Estado que no hace nada. Suponiendo y/o verificando que los dos estados son intercambiables, se estimaría primero la primera diferencia, que sería, por ejemplo, la diferencia de prevalencia de consumo de alcohol entre los dos estados en

todos los períodos de tiempo. Luego se estima la misma diferencia controlando el efecto del tiempo y el estado. La hipótesis nula (es decir, que la intervención no tiene ningún efecto sobre el consumo de alcohol de los y las jóvenes) implicaría que estas dos diferencias se mantienen iguales a lo largo del tiempo. Por lo tanto, si los dos estados son intercambiables y se cumple la hipótesis de la DS, entonces cualquier diferencia entre estas dos diferencias puede ser interpretada como el efecto de la intervención sobre el consumo de alcohol de los y las jóvenes en el estado de intervención.

Descripción de los datos

El objetivo de este estudio es estimar el impacto en el consumo de alcohol de una política pública organizada en 2011 y habiendo aumentado la edad legal para el consumo de alcohol de 18 a 19 años en 10 estados del país.

A fin de presentar la aplicación del método de diferencias en diferencias, creamos una base de datos simulada que contiene datos similares a los de un censo nacional. Asumimos que 20.000 encuestados de entre 16 y 25 años de edad fueron muestreados en 20 estados. La probabilidad de selección entre los encuestados es igual para cada estado. Asumimos una probabilidad igual de selección para evitar la necesidad de ponderación. Sin embargo, es totalmente posible incluir pesos en la aplicación de este método.

El censo contiene las siguientes variables:

- *stateID*: el identificador único de los estados.
- *individualID*: el identificador único de los encuestados.
- *Conocimiento*: una variable individual de conocimiento sobre el alcohol. Un aumento del valor indica un mayor conocimiento sobre el impacto negativo del alcohol.
- *timeEffect*: indica el año con valores entre 2009 y 2012 inclusive.
- *Ley*: variable binaria que indica si el Estado ha aprobado o no una legislación que aumenta la edad de consumo de alcohol de 18 a 19 años (es decir, política pública).
- *Edad*: una variable que indica la edad del encuestado.
- *consParSem*: número de bebidas por semana por el demandado.

Por lo tanto, tenemos los datos necesarios para hacer un análisis de diferencias en las diferencias. La base de datos simulada que se presenta aquí es típica para un análisis de DD. Incluye datos de varios años (dos años como mínimo), de diferentes estados, una política aplicada de manera diferente en diferentes años entre los estados. Estas condiciones son necesarias para construir el contrafactual y realizar el análisis de SD.

Construyendo un contrafactual

El contrafáctico está representado por una condición inobservable en la que no se aplica la política pública. Si el contrafáctico es válido, asumimos que no hay cambios en el consumo de alcohol en los estados donde no se aplica la política, y por lo tanto no hay efecto de la política pública. Hemos incluido ajustes por tiempo, estado, edad y conocimiento de los efectos negativos del alcohol en la salud para crear el contrafactual más plausible. Para validar el supuesto principal del método DD, también tenemos que suponer que la pendiente del consumo semanal a lo largo del tiempo sigue la misma trayectoria en los estados de intervención y control (véase la figura 1).

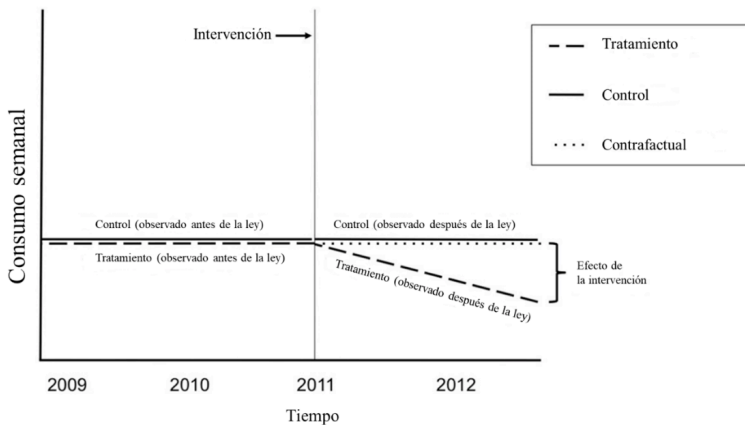


Figura 1. Representación gráfica del método de diferencias en diferencias (DD)

Estimación del impacto de la política pública

Antes de proceder a la estimación del efecto de la política pública, se requieren procedimientos estándar de depuración de los datos. Nuestros datos simulados no contienen ningún dato perdido o atípico. Dada la distribución de nuestra variable de interés (número de bebidas por semana), utilizaremos un modelo binomial negativo (para más detalles, ver Hilbe 2011).

Para estimar el efecto de la ley sobre el consumo de alcohol usando el método DD, procederemos en cuatro etapas.

Etapas 1: El primer paso es incluir la variable de intervención (ley) que indica si un estado ha aprobado o no la ley que aumenta la edad de beber de 18 a 19 años. El modelo 1 (cuadro 1) presenta el resultado de incluir la variable ley en el modelo. El efecto de la ley es una reducción de aproximadamente una unidad de consumo de alcohol por persona entre los y las jóvenes de 16 a 25 años de edad (Beta: -1,11, 95% CI -1,08 a -1,13). La variable ley es una comparación burda entre los grupos de tratamiento y control, ya que no controla dos factores importantes, el efecto del tiempo y las diferencias de estado.

Etapas 2: El modelo 2 (Tabla 1) añade un efecto fijo para la variable temporal. Para ello, incluimos una variable categórica para cada año. La inclusión de esta variable controla el efecto secular del consumo de alcohol a lo largo del tiempo, sin restricción lineal en el efecto del tiempo. El efecto estimado de la variable de la ley sobre nuestra variable de interés sigue siendo similar al paso anterior con una reducción de aproximadamente una unidad de consumo de alcohol por persona entre los y las jóvenes de 16 a 25 años (Beta: -1.08, 95% CI -1.08 a -1.03).

Etapas 3: El modelo 3 (Tabla 1) añade un efecto fijo para los estados. Incluimos una variable categórica para cada estado. La inclusión de esta variable controla cualquier característica específica observable e inobservable a nivel estatal que sea invariable en el tiempo. La inclusión de los efectos fijos por tiempo (véase Etapa 2) y por estado permite estimar el efecto intraestatal antes y después de la ley entre los estados con tratamiento, los que organizan la ley, y los estados que controlan, los que no lo hacen. En el Modelo 3, el efecto estimado de la variable ley sobre nuestra variable de interés sigue siendo estadísticamente significativo, pero la magnitud de la asociación se reduce (-0,32, IC del 95%: -0,39 a -0,24).

Etapla 4: En el Modelo 4 (Tabla 1), añadimos las dos variables individuales, nivel de conocimiento y edad. La inclusión de estas variables no cambia nuestra estimación del efecto de la ley sobre el consumo (-0,32; IC del 95%: -0,39 a -0,24).

Tabla 1 : Resultados de los modelos de Diferencia en Diferencias

	Modelo 1 Beta (95% CI)	Modelo 2 Beta (95% CI)	Modelo 3 Beta (95% CI)	Modelo 4 Beta (95% CI)
Intercepto	2.23 (2,21 a 2.24)	2.21 (2,19 a 2.24)	2.12 (2,07 a 2.18)	2.14 (2,02 a 2.25)
Ley				
No	1	1	1	
Si	-1,11 (-1,08 a -1,13)	-1,11 (-1,13 a -1,08)	-0,32 (-0,39 a -0,24)	-0,32 (-0,39 a -0,24)
Años				
2009	1	1	1	
2010		0.04 (0,00 a 0.07)	0.03 (-0,01 a 0.06)	0.03 (-0,01 a 0.06)
2011		0.04 (0,00 a 0.07)	0.04 (0,00 a 0.07)	0.04 (0,00 a 0.07)
2012		-0.02 (-0,06 a 0.01)	-0,04 (-0,07 a 0.00)	-0,04 (-0,07 a 0.00)
Conocimientos				0.00 (-0,01 a 0.00)
Edad				0.00 (-0,01 a 0.00)

* Es común no presentar los resultados del efecto fijo para los Estados. Por lo tanto, no representamos los coeficientes de cada estado. El Estado 20 está excluido del modelo debido a la colinealidad con la ley y el tiempo.

Conclusión

Usando el método DD, pudimos estimar el efecto de una política pública para cambiar la edad legal de consumo de alcohol en los y las jóvenes de 16 a 25 años. Encontramos que esta política pública contribuyó a la reducción del consumo de alcohol en este grupo de edad.

De esta manera, utilizando los contrastes entre los estados en cuanto a la fecha de implementación de esta política pública, pudimos construir situaciones contrafácticas que nos permitieron comparar los cambios observados en el consumo de alcohol entre los y las jóvenes de 16 a 25 años con lo que hubiera ocurrido sin la implementación de esta política.

El método de “regresión de la discontinuidad”

Presentación general del método RD

La regresión de la discontinuidad (RD) puede utilizarse cuando una intervención se basa en un umbral específico de una variable continua para determinar la elegibilidad para esa intervención. Este umbral puede ser la edad (como en nuestro ejemplo), el nivel de ingresos (Andalón 2011) o los marcadores biológicos como las células CD4, las células del sistema inmunológico afectadas por el VIH (Bor, Moscoe et al. 2014) por ejemplo. Todos los individuos de un lado del umbral recibirán la intervención mientras que los del otro lado no. La idea general de este enfoque es considerar como intercambiables a los individuos que están muy cerca de un lado del umbral y al otro lado del mismo, considerando la elección del umbral en su valor preciso como aleatoria. Así pues, el grupo contrafáctico se construye a partir de una situación en la que no habría discontinuidad en el umbral de elegibilidad para la intervención. Así pues, se podrá inferir que una diferencia

en los resultados sanitarios entre los dos grupos será la causa del establecimiento de este umbral. Se requieren varios pasos para realizar un análisis de tipo RD a fin de probar varias hipótesis relativas a la intercambiabilidad entre los dos grupos, y la elección de modelos de regresión. Estas siete Etapas se detallan a continuación.

Descripción de los datos

El objetivo de nuestro estudio es estimar el impacto de la edad legal de consumo de alcohol en los y las jóvenes. Para presentar la aplicación del método de regresión de la discontinuidad, creamos una base de datos simulada. La base de datos simulada contiene datos por cada 1.000 personas de 19 a 23 años de edad. Al igual que en el caso del análisis DD, no se incluyeron pesos en los análisis aquí, pero esto todavía es posible.

Una variable de interés es el consumo de alcohol. Tenemos datos por cada 1.000 personas de 19 a 23 años.

La base de datos contiene las cinco siguientes variables:

- id: el identificador único para los encuestados.
- Edad: una variable que indica la edad del encuestado.
- propAlcDía: el número de bebidas por día consumidas por el encuestado.
- fruta: consumo semanal de frutas y verduras.
- Cigarrillo: consumo semanal de cigarrillos.

Aplicando el método de RD de siete etapas

1. Definir el umbral

Los datos requieren un umbral de probabilidad de consumo en relación con una variable de atribución continua. En nuestro ejemplo, la edad es la variable de atribución. Aquí, la mayoría de edad es de 21 años. Así, antes de los 21 años de edad el consumo de alcohol no es legal y después de los 21 años de edad sí lo es. Esto es una clara ruptura a la edad de 21 años.

2. Comprobar que no hay sesgo o error de medición de la variable de interés en relación con el punto de corte para el consumo de alcohol.

Un sesgo o error de medición para la variable de interés relativa al umbral creará un sesgo sistemático en las estimaciones de RD. Por ejemplo, si se subestima el consumo de alcohol entre los participantes menores de 21 años, ya sea porque los encuestados no dan respuestas reales (o se niegan a responder) porque el comportamiento es ilegal, la estimación de RD es sistemáticamente sesgada (Imbens y Lemieux 2008). En nuestro caso, asumimos que no hay sesgo o error de medición para la variable de interés.

3. Verificar que la variable de atribución es continua y se mide con precisión

En la tercera etapa, visualizamos un histograma, un gráfico “Q-Q”, y realizamos una prueba de Shapiro-Wilk para asegurarnos de que la variable de asignación continua está distribuida uniformemente. Una variable de atribución que no sea continua será problemática en el contexto del método de regresión de la discontinuidad, ya que es posible que existan varios umbrales. Por ejemplo, si existe la variable de atribución y varios umbrales, no es posible identificar un efecto causal ya que es posible un efecto en cada umbral. La variable de edad en nuestra muestra es consistente.

4. Visualizar la variable de atribución por la variable de interés

En la cuarta etapa, visualizamos la variable de atribución en el eje X con las variables de interés en el eje Y. La figura 2 muestra estas relaciones. Presentamos los datos completos antes y después de los 21 años de edad usando opciones de gráficos específicos. El loess de alisado es el gráfico por defecto creado por el comando rRDobust con binning (imitando el método de varianza uniformemente espaciada usando estimadores de espaciado) y alisado con polinomio de 4º orden. Se recomienda presentar un gráfico con los datos de los contenedores (Calonico, Cattaneo et al. 2015).

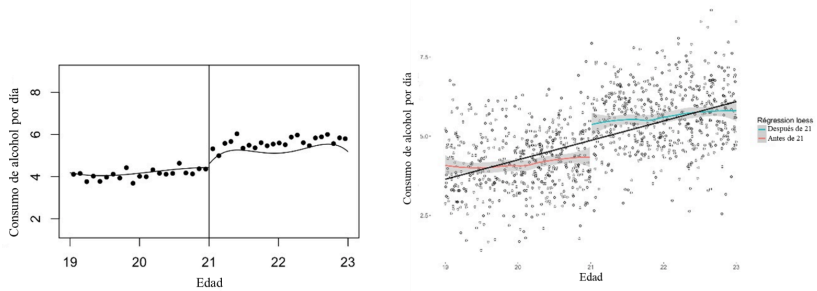


Figura 2. Representación gráfica del método de Regresión con Discontinuidad

5. Comparación de variables de interés y co-variables

En la quinta etapa, queremos verificar que los dos grupos son intercambiables, excepto la variable de interés de la salud. Comparamos la media y la desviación estándar de la variable de interés antes y después de 21 años. También examinamos esto para dos co-variables, el consumo de frutas y verduras y el consumo de cigarrillos (Tabla 2). Dada la premisa de que la política sólo influye en el consumo de alcohol, la política no debe influir en otros comportamientos. La idea de este paso es asegurar que, excepto en el caso del consumo de alcohol, no haya otra discontinuidad en el mismo límite (es decir, a los 21 años) para otras variables de la muestra.

Tabla 2 : Resultados de los modelos de regresión con discontinuidad

	Antes de los 21 años (promedio y DE)	Después de 21 años (promedio y DE)
Alcohol	4.2 (1,0)	5.6 (1,0)
Fruta	7.5 (4,5)	7.6 (4,3)
Cigarrillo	39.0 (22,0)	40.0 (23,0)

DE: *Desviación estándar.*

6. Seleccione el rango de muestra seleccionado

Una premisa importante del método RD es que los grupos a ambos lados del umbral son intercambiables. Esto significa que deben ser similares para cualquier característica que no sea la variable de atribución (la edad en nuestro ejemplo) y la variable de salud de interés. Hay varios métodos propuestos para seleccionar el ancho de banda. El ancho de banda define el rango de la muestra seleccionada para la regresión. Es muy importante ya que define la plausibilidad de que los grupos a ambos lados del umbral sean intercambiables. El ancho de banda es típicamente estimado con una estimación del núcleo. Aquí proponemos dos métodos de selección, a saber: MSE (Mean Square Error Optimal Bandwidth Selector) y CER (Coverage Error Rate Optimal Bandwidth Selector)). (Imbens y Kalyanaraman 2011, Calonico, Cattaneo et al. 2015) y tres métodos diferentes de estimación del núcleo (triangular, Epanechnikov y uniforme). Los métodos de selección de ancho de banda están disponibles con el programa *rdrobust* para R (Calonico, Cattaneo et al. 2015) y Stata (Calonico, Cattaneo et al. 2014).

La selección del ancho de banda puede influir en el tamaño de la muestra y en la muestra para respetar el hecho de que los dos grupos a cada lado del umbral son intercambiables. En nuestro ejemplo, nuestra muestra de 1.000 encuestados no se ve afectada por la elección de ancho de banda, con 484 encuestados de la derecha y 516 de la izquierda incluidos en el análisis para una selección de ancho de banda de IK MSE con una estimación de núcleo triangular o una selección de ancho de banda de CT con una estimación de núcleo uniforme (véase la tabla 3 para los resultados completos).

En nuestro caso, elegiremos el método MSE con estimación de núcleo triangular para los análisis, es decir, el método por defecto bajo R con *rdrobust*. Independientemente de la elección, es necesario realizar análisis de sensibilidad para asegurar que el resultado no se vea afectado por la elección del modelo.

7. Estimación del efecto de la política pública mediante modelos de regresión

Primero, estimamos el efecto de la política con una regresión lineal en la muestra completa. La estimación de la regresión lineal es un aumento de

la bebida de 0,6 por semana (95% CI 0,6 a 0,7). Con la muestra completa, la premisa de que los grupos a ambos lados del corte son intercambiables es menos plausible. Para estimar un efecto plausible, estimamos la regresión con discontinuidad con el método IEM y la estimación del núcleo triangular. La regresión incluye ahora 484 encuestados a la izquierda (antes de los 21 años) y 516 encuestados a la derecha (después de los 21 años) (Tabla 4). La regresión se estima con una regresión local lineal y una corrección local cuadrática. La estimación no se ve influida por el método de selección del ancho de banda, lo que indica que el resultado es robusto.

Tabla 3 : Estimación del efecto del consumo de alcohol con RD

	Bandwidth MSE Beta (95% CI)	Bandwidth CER Beta (95% CI)
Convencional	0.88 (0,2 a 1,5)	0.80 (0,0 a 1,6)
Corregido por los sesgo	0.80 (0,2 a 1,4)	0.75 (-0,0 a 1,5)
Robusto	0.80 (0,0 a 1,6)	0.75 (-0,1 a 1,6)

Los resultados muestran que el estar en el grupo de edad legal para beber aumenta en aproximadamente una unidad de consumo de alcohol por semana (95% CI 0.2 a 1.5). Los análisis de sensibilidad (CER corregidos por sesgos y robusto) muestran que la estimación puede ser sensible a la elección del modelo. El/la investigador/a debe ser cauteloso/a al interpretar y comunicar los resultados.

Conclusión

Usando el método de RD, pudimos estimar el efecto de la edad mínima legal para beber en el consumo de alcohol entre los 19-23 años. Encontramos que el consumo de alcohol aumenta en una unidad de consumo por semana cuando se pasa de los 20 a los 21 años de edad, lo que indica que esta política pública tiene un efecto sobre el consumo de alcohol entre los y las jóvenes de 19 a 23 años.

Utilizando los contrastes entre los y las jóvenes de 19-20 y 21-23 años, pudimos construir un contrafactual que nos permitió comparar los cambios observados en el consumo de alcohol entre los y las jóvenes de 19-23 años con lo que habría ocurrido sin la aplicación de esta política pública.

Discusión

En este capítulo hemos presentado dos métodos cuasiexperimentales en el contexto de la evaluación de los efectos de las políticas públicas sobre la salud. Hemos demostrado que estos enfoques pueden ser buenas alternativas cuando los ensayos controlados aleatorios no son factibles. Presentamos dos enfoques principales: el método de diferencia en la diferencia y el enfoque de regresión de la discontinuidad. Estos dos enfoques permiten explotar diferentes situaciones de experimentos naturales. En el caso del método de diferencias en las diferencias, se utilizó por primera vez la fecha de aplicación de la intervención para comparar los cambios en los indicadores de salud con el grupo de control. En el caso del enfoque de regresión con discontinuidad, en cambio, se explotó una característica específica de la intervención, a saber, la edad legal para beber. Ambos enfoques tienen por objeto responder a la misma pregunta, pero la principal

diferencia radica en el efecto que se estima y en la población a la que se pueden generalizar estos resultados. En el primer caso, se cuantifica un efecto medio del tratamiento que toma la población en su conjunto (es importante señalar que también es posible cuantificar el efecto en subgrupos de la población). En el segundo caso, se cuantifica un promedio local total del efecto del tratamiento en la población incluida en el análisis (es decir, las personas de 19 a 23 años), de modo que los resultados no pueden extrapolarse a otros grupos de edad en particular.

Se han descrito todos los pasos necesarios para llevar a cabo este tipo de análisis y las hipótesis que se van a probar. Estos enfoques, especialmente desarrollados en el campo de la econometría, se han utilizado ampliamente para evaluar la eficacia de las intervenciones sanitarias en los países de ingresos bajos y medios, en particular en lo que respecta a los programas de transferencia de efectivo (condicional o no) (Andalón 2011) o a los programas de acceso a la atención (Dimick y Ryan 2014). Es importante señalar que es posible llevar a cabo este tipo de evaluación utilizando los datos ya reunidos en los países de ingresos bajos y medios, lo que permite realizar una gran cantidad de investigaciones en el futuro.

En este capítulo hemos abordado un solo tipo de evaluación, a saber, la evaluación de los efectos o impactos de una política pública, movilizand la noción de inferencia causal. Inevitablemente, este tipo de evaluación es complementario de otros tipos, como las evaluaciones de la aplicación, que movilizan diferentes metodologías. Por último, hemos abordado aquí el impacto de las políticas públicas en la salud de la población estudiada en su conjunto. Sin embargo, es posible, incluso sistemático, que los efectos de una política pública no se distribuyan homogéneamente entre varios subgrupos de una población. Este aspecto, a través de la noción de equidad, se aborda por lo tanto en el siguiente capítulo.

Referencias clave

Angrist, J. D. et Pischke, J.-S. (2014). *Mastering metrics : The Path from Cause to Effect*. Lieu d'édition : Princeton University Press.

Este libro es una excelente introducción a los métodos cuasi-experimentales y otros enfoques comúnmente utilizados en la econometría y la

epidemiología. Se trata de una versión más accesible que la de Mostly harmless econometrics: An empiricist's companion (Princeton University Press, 2008) de los mismos autores, que es más detallada pero de más difícil acceso.

Bor, J. (2016). Capitalizing on natural experiments to improve our understanding of population health. *American Journal of Public Health*, 106(8), 1388.

Este artículo, muy corto y de fácil acceso, es un muy buen resumen del concepto de los experimentos naturales y de cómo su explotación puede ser una excelente alternativa a los ensayos controlados aleatorios.

Deaton, A. et Cartwright, N. (2017). Understanding and misunderstanding randomized controlled trials. *Social Science et Medicine*, volume(numéro), p.-p.

En este artículo se resume gran parte de la labor realizada en los últimos años con el fin de aclarar lo que son los ensayos controlados aleatorios y, especialmente, lo que no son. Este artículo, y los numerosos comentarios que lo acompañan, deben leerse para comprender los límites y ventajas de este tipo de enfoque, que lamentablemente se malinterpreta y moviliza con demasiada frecuencia.

Referencias

Andalón, M. (2011). Oportunidades to reduce overweight and obesity in Mexico? *Health economics*, 20(S1), 1-18.

Angrist, J. D. et Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton : Princeton University Press.

Angrist, J. D. et Pischke, J.-S. (2014). *Mastering metrics : The path from cause to effect*. Princeton : Princeton University Press.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate behavioral research*, 46(3), 399-424.

Belaid, L. et Ridde, V. (2012). An implementation evaluation of a policy aiming to improve financial access to maternal health care in Djibo district, Burkina Faso. *BMC pregnancy and childbirth*, 12(1), 1.

- Benmarhnia, T., Zunzunegui, M.-V., Llácer, A. et Béland, F. (2014). Impact of the economic crisis on the health of older persons in Spain: research clues based on an analysis of mortality. *SESPAS report 2014. Gaceta Sanitaria*, 28, 137-141.
- Bonell, C. P., Hargreaves, J., Cousens, S., Ross, D., Hayes, R., Petticrew, M. et Kirkwood, B. (2011). Alternatives to randomisation in the evaluation of public health interventions: design challenges and solutions. *Journal of Epidemiology and Community Health*, 65(7): 582-587.
- Bor, J. (2016). Capitalizing on natural experiments to improve our understanding of population health. *American Journal of Public Health*, 106(8), 1388.
- Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.-L. et Bärnighausen, T. (2014). Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology*, 25(5), 729-737.
- Brousselle, A., Benmarhnia, T. et Benhadj, L. (2016). What are the benefits and risks of using return on investment to defend public health programs?. *Preventive Medicine Reports*, 3, 135-138.
- Brousselle, A. et Lessard, C. (2011). Economic evaluation to inform health care decision-making: promise, pitfalls and a proposal for an alternative path. *Social science et medicine*, 72(6), 832-839.
- Calonico, S., Cattaneo, M. D. et Farrell, M. H. (2015). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association*, 113(522), 767-779.
- Calonico, S., Cattaneo, M. D. et Titiunik, R. (2014). Robust data-driven inference in the regression-discontinuity design. *Stata Journal*, 14(4), 909-946.
- Calonico, S., Cattaneo, M. D. et Titiunik, R. (2015). Optimal data-driven regression discontinuity plots. *Journal of the American Statistical Association*, 110(512), 1753-1769.
- Calonico, S., Cattaneo, M. D. et Titiunik, R. (2015). rdrobust: An r package for robust nonparametric inference in regression-discontinuity designs. *R Journal*, 7(1), 38-51.

- Carpenter, C. et Dobkin, C. (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1), 164-182.
- Carpenter, C. et Dobkin, C. (2011). The minimum legal drinking age and public health. *The Journal of Economic Perspectives*, 25(2), 133-156.
- Champagne, F., Contandriopoulos, A.-P., Brousselle, A., Hartz, Z. et Denis, J.-L. (2009). L'évaluation dans le domaine de la santé : concepts et méthodes. In *L'évaluation : Concepts et méthodes*. Montréal : Presses de l'Université de Montréal.
- Cook, T. D., Campbell, D. T. et Day, A. (1979). *Quasi-experimentation : Design et analysis issues for field settings*. Boston : Houghton Mifflin.
- Dimick, J. B. et Ryan, A. M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama*, 312(22), 2401-2402.
- Donald, S. G. et Lang, K. (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, 89(2), 221-233.
- Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L. et Torrance, G. W. (2015). *Methods for the economic evaluation of health care programmes*. Lieu d'édition : Oxford University Press.
- Fuller, D., Sahlqvist, S., Cummins, S. et Ogilvie, D. (2012). The impact of public transportation strikes on use of a bicycle share program in London: interrupted time series design. *Preventive medicine*, 54(1), 74-76.
- Hawe, P., Shiell, A. et Riley, T. (2004). Complex interventions: how "out of control" can a randomised controlled trial be?. *Penelope*, 328, 1561-1563.
- Hernán, M. A. et Robins, J. M. (2006). Instruments for causal inference: an epidemiologist's dream?. *Epidemiology*, 17(4), 360-372.
- Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge : Cambridge University Press.
- Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396), 945-960.
- Imbens, G. et Kalyanaraman, K. (2011). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3), 933-959.

- Imbens, G. W. et Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.
- Kaufman, J. S. (2007). Making causal inferences about macrosocial factors as a basis for public health policies. *Macrosocial Determinants of Population Health*, chapter 17, 355-373.
- Kontopantelis, E., Doran, T., Springate, D. A., Buchan, I. et Reeves, D. (2015). Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *British Medical Journal*, 350, h2750.
- Macintyre, S. (2010). Good intentions and received wisdom are not good enough: the need for controlled trials in public health. *Journal of Epidemiology and Community Health*, 65(7), 564-567.
- Moore, L. et Moore, G. F. (2011). Public health evaluation: which designs work, for whom and under what circumstances?. *Journal of Epidemiology and Community Health*, 65(7), 596-597.
- Naimi, A. I. et Kaufman, J. S. (2015). Counterfactual theory in social epidemiology: reconciling analysis and action for the social determinants of health. *Current Epidemiology Reports*, 2(1), 52-60.
- Newhouse, J. P. et McClellan, M. (1998). Econometrics in outcomes research: the use of instrumental variables. *Annual Review of Public Health*, 19(1), 17-34.
- Pearl, J. (2009). *Causality*. Cambridge : Cambridge University Press.
- Petticrew, M., Chalabi, Z. et Jones, D. R. (2011). To RCT or not to RCT: deciding when 'more evidence is needed' for public health policy and practice. *Journal of Epidemiology and Community Health*, 66(5), 391-396.
- Rehm, J., Gmel, G., Sempos, C. T. et Trevisan, M. (2003). Alcohol-related morbidity and mortality. *Alcohol Research et Health*, 27(1), 39-51.
- Royall, R. M. (1991). Ethics and statistics in randomized clinical trials. *Statistical Science*, 6(1), 52-62.
- Sabatier, P. et Mazmanian, D. (1980). The implementation of public policy: A framework of analysis. *Policy Studies Journal*, 8(4), 538-560.
- Shadish, W. R. et Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, 60, 607-629.

- Shadish, W. R., Luellen, J. K. et Clark, M. (2006). Propensity Scores and Quasi-Experiments: A Testimony to the Practical Side of Lee Sechrest. Dans R. R. Bootzin et P. E. McKnight (dir.), *Strengthening research methodology: Psychological measurement and evaluation* (p. 143-157). Washington, DC : American Psychological Association.
- Susser, M. (2001). Glossary: causality in public health science. *Journal of Epidemiology and Community Health*, 55(6), 376-378.
- Yörük, C. E. et Yörük, B. K. (2012). The impact of drinking on psychological well-being: Evidence from minimum drinking age laws in the United States. *Social Science et Medicine*, 75(10), 1844-1854.

Resumen / Résumé / Abstract

Se han desarrollado métodos cuasi-experimentales para evaluar los “experimentos naturales”, un diseño alternativo a los ensayos controlados aleatorios que se basa en la noción de lo contrafactual para estimar los efectos causales. El objetivo de este capítulo es describir métodos cuasi-experimentales, aplicarlos empíricamente a un ejemplo y proporcionar todos los elementos de la evaluación, permitiendo al lector reproducir estos análisis (en R y Stata) en su contexto dado. El objetivo principal de este tipo de evaluación de políticas es atribuir un efecto causal de una intervención a un indicador de salud. Aplicamos estos métodos para estimar el impacto de una política que aumenta el consumo mínimo legal de alcohol en el consumo de alcohol entre los jóvenes. En esta evaluación se utilizan dos tipos de métodos cuasiexperimentales: el método de la Diferencia de Diferencias (DD) y el enfoque de regresión discontinua (RD). Demostramos que estos enfoques pueden ser buenas soluciones para evaluar las políticas cuando los ensayos controlados aleatorios son imposibles. Estas herramientas de evaluación pueden utilizarse para garantizar que las políticas contribuyan a mejorar la salud de la población, para informar y decidir si se mantienen, modifican o interrumpen las intervenciones existentes y para garantizar que los recursos públicos se utilicen de manera eficaz.

Les méthodes quasi-expérimentales (MQE) sont des alternatives aux essais contrôlés randomisés, qui permettent, en se basant sur le modèle

contre-factuel, d'évaluer les effets d'interventions en exploitant des expérimentations naturelles. L'objectif de ce chapitre est d'introduire deux MQE, à savoir la méthode de Différence dans les Différences (DD) ainsi que l'approche de Régression avec Discontinuité (RD). Leur application afin d'évaluer l'effet de l'âge légal minimum sur la consommation d'alcool chez les jeunes est ensuite présentée. Sont présentées toutes les étapes nécessaires et les codes statistiques (pour R et Stata) permettant de reproduire ce type d'analyse. Ces approches constituent de bonnes solutions lorsque des essais contrôlés randomisés sont impossibles ou difficiles à mener, permettant de s'assurer que des objectifs d'un programme de santé publique sont bel et bien atteints, de décider quant au maintien, à la modification ou à l'arrêt d'interventions existantes et de s'assurer que les ressources publiques sont utilisées de manière efficace.

Quasi-experimental methods have been developed to evaluate 'natural experiments,' an alternative design to randomized controlled trials that relies on the notion of the counterfactual to estimate causal effects. The objective of this chapter is to describe quasi-experimental methods, apply them empirically to an example, and provide all elements of the evaluation, allowing the reader to reproduce these analyses (in R and Stata) in their given context. The main objective of this type of policy evaluation is to attribute a causal effect of an intervention on a health indicator. We apply these methods to estimate the impact of a policy raising minimum legal drinking on alcohol consumption among young people. Two types of quasi-experimental methods are used in this evaluation: the Difference in Differences (DD) method and the discontinuous regression (RD) approach. We show that these approaches can be good solutions to evaluate policies when randomized controlled trials are impossible. These evaluation tools can be used to ensure that policies contribute to the improvement of population health, to inform and decide whether to maintain, modify or discontinue existing interventions and to ensure public resources are used effectively.

Tarik Benmarhnia es Profesor Adjunto de Epidemiología de la Universidad de California en San Diego en el Instituto Scripps de Oceanografía de la Facultad de Medicina. Realizó su doctorado en

Epidemiología en la Universidad de Montreal y en la Université Paris Sud. También es investigador asociado en Chile y Francia y miembro del consejo editorial de *Environmental Health Perspectives*. Sus investigaciones se centran en los efectos de los fenómenos extremos en la salud, con especial atención a las poblaciones vulnerables y las consecuencias para las políticas públicas. También está elaborando enfoques metodológicos para evaluar el efecto de las políticas ambientales en la salud, como las medidas de adaptación al cambio climático y las medidas de control de la contaminación atmosférica.

Daniel Fuller ocupa una cátedra de investigación sobre actividad física de la población en la Universidad Memorial de St. Su investigación se centra en las tecnologías portátiles para estudiar la actividad física, las intervenciones de transporte y la desigualdad en los espacios urbanos. Su trabajo metodológico se centra en los métodos de experimentación natural y el aprendizaje de las máquinas. Dan tiene una Maestría en Kinesiología de la Universidad de Saskatchewan, un Doctorado en Salud Pública de la Universidad de Montreal y realizó investigaciones postdoctorales en el Departamento de Salud Comunitaria y Epidemiología de la Universidad de Saskatchewan. Daniel es codirector del equipo de Factores de Vecindad del Consorcio de Investigación de Salud Ambiental Urbana Canadiense (CANUE) e investigador principal del equipo INTERACT.

Cita

Tarik Benmarhnia y Daniel Fuller (2020). Métodos cuasi-experimentales. El efecto de la edad mínima legal sobre el consumo de alcohol entre los y las jóvenes en los Estados Unidos. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 249-275. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

10. Ensayos aleatorizados por conglomerados

Ejemplo dirigido a la salud materna e infantil

ALEXANDRE DUMONT

Ensayos aleatorizados por conglomerados

Definición del método

Los ensayos aleatorizados por conglomerados evalúan el impacto de las intervenciones sanitarias aplicadas a nivel de un grupo de personas, como las pacientes que dan a luz en un mismo hospital o las personas que viven en la misma aldea.

Fortalezas del método

- Evita el sesgo de la contaminación mediante la intervención de individuos (u objetivos) de dos grupos diferentes
- Permite un alto nivel de pruebas de los efectos de la intervención
- Los ensayos controlados por conglomerados son con frecuencia estudios experimentales pragmáticos que evalúan intervenciones de la vida real.

Desafíos del método

- La intervención debe ser lo más estandarizada posible
- Desafíos estadísticos relacionados con el hecho de que la unidad de aleatorización y análisis no es la misma
- Desafíos en la selección de grupos representativos del sistema de salud en el que se lleva a cabo la investigación.

La aplicación de las políticas sanitarias requiere la experimentación periódica de nuevas intervenciones adaptadas al contexto y destinadas a facilitar la aplicación de buenas prácticas que tengan un fuerte impacto en la salud de las poblaciones. En la lucha contra la mortalidad materna, la mayoría de las políticas de salud perinatal del mundo se basan en el



La maternidad en la RDC. Crédito de "Action Santé des Femmes".

parto asistido por personal calificado y el tratamiento de las complicaciones obstétricas cuando se producen durante el trabajo de parto o el parto. Estos tratamientos son conocidos, pero el desafío para la mayoría de los sistemas de salud en desarrollo es implementarlos de manera oportuna. Se han ensayado muchas iniciativas en países de ingresos bajos y medios para movilizar a las comunidades o mejorar el sistema de atención de la salud y las prácticas médicas con el propósito de que las mujeres puedan acceder más rápidamente a servicios obstétricos de calidad.

Algunas evaluaciones experimentales o cuasi-experimentales han proporcionado pruebas tempranas sobre la eficacia y la aplicación de las intervenciones para reducir la mortalidad materna. Sin embargo, esta información sigue siendo fragmentada y se han realizado pocos estudios de impacto de buena calidad en los países de bajos ingresos en comparación con los países con mayores niveles de recursos. Sin embargo, es necesario generar y difundir más pruebas sobre la eficacia de esas intervenciones, ya se trate de programas nacionales o de iniciativas más localizadas. Esto debería permitir a los encargados de formular políticas y adoptar decisiones y a los responsables de los programas de salud materna elaborar políticas públicas más efectivas.

El nivel de pruebas de los estudios de impacto basados en la población puede compararse con el de los estudios clínicos de los tratamientos con drogas (véase el cuadro 1). Los estudios con mayor nivel de pruebas son los ensayos controlados aleatorios (ECA) en los que se controlan la mayoría de los sesgos de selección de los participantes, asegurando así que los cambios observados son efectivamente atribuibles a la intervención y no a las diferencias entre las personas o las prácticas de atención. Existen dos categorías de ECA: ECA individuales en los que los sujetos se seleccionan de forma aleatoria individualmente para participar en diferentes grupos de intervención (aleatorización individual) y ECA por conglomerados en los que los grupos de sujetos se aleatorizan (aleatorización por conglomerados).

Tabla 1 : Niveles de evidencia científica y grados de recomendación (Fuente: Hautes Autorités de Santé, Francia)

Nivel de evidencia científica proporcionada por la literatura (estudios terapéuticos)	Grado de recomendaciones
Nivel 1 (NP1) Ensayos controlados aleatorios de alta potencia Meta-análisis de ensayos controlados aleatorios Análisis de decisiones basado en estudios bien realizados	Pruebas científicas establecidas A
Nivel 2 (NP2) Ensayos controlados aleatorios de baja potencia. Estudios controlados no aleatorios bien realizados Estudios de cohortes	Presunción científica B
Nivel 3 (NP3) Estudios de control de casos	Bajo nivel de evidencia C

Los ECA individuales son poco frecuentes en los países de ingresos bajos y medios por tres razones principales: i) el costo de esos estudios es elevado, ya que por lo general requieren la contratación de pacientes de múltiples centros durante varios años; ii) la mayoría de las intervenciones que deberían probarse en un entorno de bajos recursos ya tienen una eficacia conocida en los países de altos recursos; y iii) la asignación aleatoria de sujetos puede no ser posible para evaluar las intervenciones en la comunidad por razones logísticas, administrativas o incluso éticas. Otra desventaja de los ECA individuales es que la captación de sujetos suele ser muy selectiva debido a los múltiples criterios de inclusión y exclusión, y las condiciones en que se aplican las intervenciones están demasiado estandarizadas. Los resultados obtenidos por un ECA son entonces difíciles de generalizar a contextos distintos de aquel en el que se realizó el estudio.

El enfoque utilizado en los ECA por conglomerados es mucho más pragmático que el de los ECA individuales (Cuadro 2). La intervención probada no se dirige directamente al individuo, sino a una unidad “social” independiente (familia, pueblo, médico, hospital, etc.) para cambiar el comportamiento o introducir nuevas prácticas en la “vida real”.

Tabla 2 : Ensayos aleatorios individuales y por conglomerados

	ECAs individuales	ECAs por conglomerados
Unidad de Intervención y Aleatoriedad Grupo temático	Sujeto	Grupo de sujetos
Tipo de intervención	Tratamiento	Programa de intervención
Objetivo	Mejorar el estado de salud	Cambiar los comportamientos
Riesgo de contaminación	Alto	Bajo
Contexto del estudio	Altamente estandarizado	El entorno clínico de la vida real
Costo	Alto	Alto

Desafíos de los ensayos controlados aleatorios por conglomerados

Si bien los ensayos aleatorios por conglomerados ofrecen varias ventajas en los países de ingresos bajos y moderados con respecto a los ensayos individuales, presentan algunas dificultades metodológicas porque la unidad de análisis es diferente de la unidad de aleatorización o intervención (figura 1).

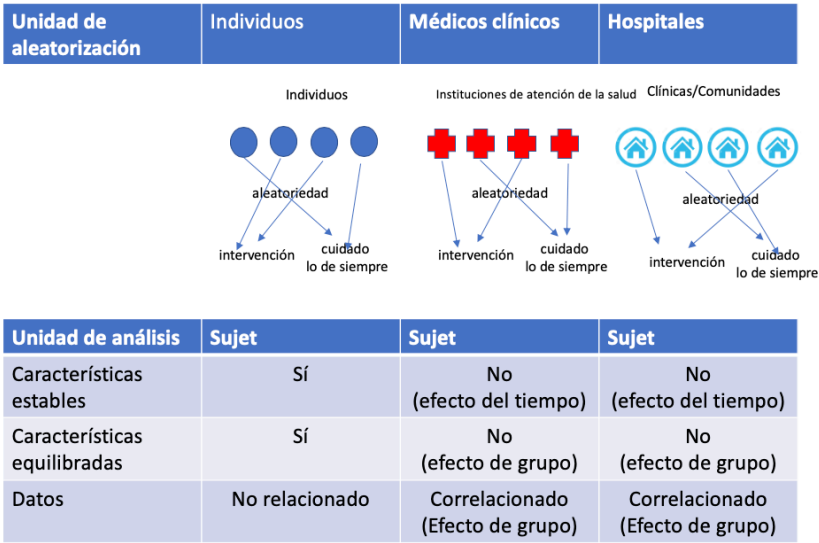


Figura 1. Implicaciones de los ensayos controlados aleatorios por conglomerados

La aleatorización de los conglomerados, nos pone en la posición de recoger información a nivel de individuos cuyas características no son necesariamente estables en el tiempo, ni equilibradas en los grupos del ensayo. Además, estos datos están correlacionados dentro de cada grupo. Esta situación lleva a tres importantes consecuencias a considerar cuando se analizan los resultados de un ensayo de grupo.

En primer lugar, las características de los individuos de los grupos, pero también el desempeño general del sistema de salud, pueden cambiar durante el período de estudio. Esto explica los cambios observados en los indicadores

de salud en los grupos en los que no se aplica ninguna intervención externa (grupo de control), a los que nos referiremos como “efecto del tiempo”. Por ejemplo, la tasa de mortalidad materna ha disminuido casi un 44% en los últimos 25 años en todo el mundo. Muchos países han establecido políticas sanitarias que han contribuido a esta tendencia positiva. Si no se tiene en cuenta la tendencia natural de los indicadores de salud materna en un estudio de los efectos de una nueva intervención, ello equivaldría a atribuir los cambios observados únicamente a esa intervención y no al efecto combinado del tiempo y la intervención. Por ello es esencial comparar los resultados observados en un grupo experimental con los observados en un grupo de control sin intervención externa. Veremos cómo el enfoque de “diferencia en diferencias” en este caso nos permite individualizar el efecto de la intervención del efecto del tiempo.

Otra implicación metodológica de la aleatorización por conglomerados está relacionada con un frecuente desequilibrio en las características de los individuos entre los grupos de comparación. De hecho, la aleatorización de grupos conduce casi sistemáticamente a una situación en la que la aleatorización se produce antes de la inclusión de los sujetos. Las características de los grupos se distribuyen entonces homogéneamente entre los grupos si la aleatorización se realiza adecuadamente. El desequilibrio resultante, denominado “efecto de grupo”, puede explicar las diferencias en los indicadores de salud entre los grupos de comparación, incluso antes de que se inicie la intervención. El riesgo de no tener en cuenta este desequilibrio sería atribuir el efecto de la intervención a la diferencia inicial entre los grupos, aun cuando la intervención fuese inefectiva.

Por último, mediante la asignación aleatoria por conglomerados, se pueden recoger datos correlacionados. Los sujetos de una unidad o grupo “social” no pueden considerarse independientes entre sí. De hecho, los resultados observados en dos sujetos del mismo grupo tenderán a ser más “similares” que si los resultados se asociaran a dos sujetos de dos grupos separados (y por lo tanto independientes). Si no se tiene en cuenta la correlación inducida por el diseño experimental (el efecto de conglomerado) se subestimaría la potencia de las pruebas estadísticas, se reducirían falsamente los intervalos de confianza y se llegaría a conclusiones de estudio erróneas.

En el caso del ensayo QUARITE en Senegal y Malí, se asignaron hospitales al azar para que intervinieran en los profesionales de la salud en

el supuesto de que la mejora de la calidad de la atención conduciría a una disminución de la mortalidad materna. Veremos, a través de este ejemplo, cómo el uso de métodos estadísticos apropiados nos permite analizar correctamente estos datos.

Contexto y programa evaluado: el ensayo QUARITE para reducir la mortalidad materna en el hospital

A pesar de la disminución de la mortalidad materna en todo el mundo, ésta sigue siendo muy elevada en muchos países de ingresos bajos y medios. Una gran proporción de las muertes maternas se deben a un tratamiento tardío o inapropiado de las complicaciones obstétricas. Por lo tanto, existe una necesidad urgente de intervenciones efectivas para hacer frente a esta situación. Esas intervenciones suelen entrañar la movilización de las comunidades para mejorar la utilización oportuna de la atención y el mejoramiento de la continuidad de la atención y las prácticas médicas en los diferentes niveles de atención sanitaria.

Probamos un enfoque que combina el análisis de casos de mortalidad materna en centros de salud (o auditorías de muertes) y la capacitación en el servicio del personal de salud (Dumont y otros, 2013) mediante un ECA por conglomerados (ensayo QUARITE) entre 2007 y 2011. La unidad de aleatorización e intervención fue el hospital de referencia en el que se trataron las complicaciones obstétricas más graves en Malí y Senegal. La intervención se dirigió a los profesionales de la salud que trabajaran en la misma unidad (sala de maternidad). La intervención, de dos años de duración, consistió en la capacitación inicial de líderes de opinión locales en las mejores prácticas y en auditorías de muertes maternas (capacitación de instructores). A continuación, los líderes de opinión capacitaron a sus equipos (médicos, parteras y enfermeras capacitadas) en cada hospital e implementaron las auditorías de muerte materna. Un facilitador externo visitó cada centro cada tres meses para supervisar el cumplimiento de la agenda de entrenamiento y supervisar una sesión de auditoría. El objetivo de la Auditoría de Muerte Materna era investigar cualitativa y exhaustivamente las causas y circunstancias de las muertes maternas en los centros de salud. Como primer paso, estas muertes fueron identificadas y analizadas en las instalaciones. Sin embargo, cuando fue posible, estos estudios también

trataron de identificar los factores ajenos al hospital, al sistema de salud o a la comunidad que contribuyeron a la muerte. La capacitación del personal sanitario se centró en los problemas de gestión identificados en las sesiones de auditoría para ayudar a mejorar la calidad de la atención en los hospitales participantes.

Protocolo de evaluación

Para medir los efectos de la intervención QUARITE sobre la mortalidad materna en los 46 hospitales de referencia que participaron en el ensayo, se registraron todos los nacimientos en estos centros durante cuatro años, entre octubre de 2007 y octubre de 2011. Las muertes maternas se identificaron a partir de múltiples fuentes de información como los registros de partos, quirófanos, hospitales o morgues y los registros clínicos disponibles. El principal criterio de valoración del ensayo fue la tasa de mortalidad materna estimada (TMM), definida como el número de muertes maternas registradas en el hospital dividido por el número de partos durante el mismo período.

Cuatro años de recogida de datos permitieron estimar este indicador antes de que se aplicara la intervención (año 1), durante el período de intervención (años 2 y 3) y después de que se finalizara la intervención (año 4).

Los hospitales se asignaron al azar al final del primer año de recogida, justo antes del comienzo de la intervención, entre un grupo experimental expuesto al programa y un grupo de control sin intervención externa. La figura 2 muestra los cambios en la mortalidad materna en los dos grupos durante el período de estudio. Disminuyó de 10,4 muertes por cada 1.000 partos a 6,8 muertes por cada 1.000 en el grupo experimental entre los períodos “antes” y “después” de la intervención. Esta disminución fue menos marcada en el grupo de control: entre 8,1 y 7,1 por 1000 partos.

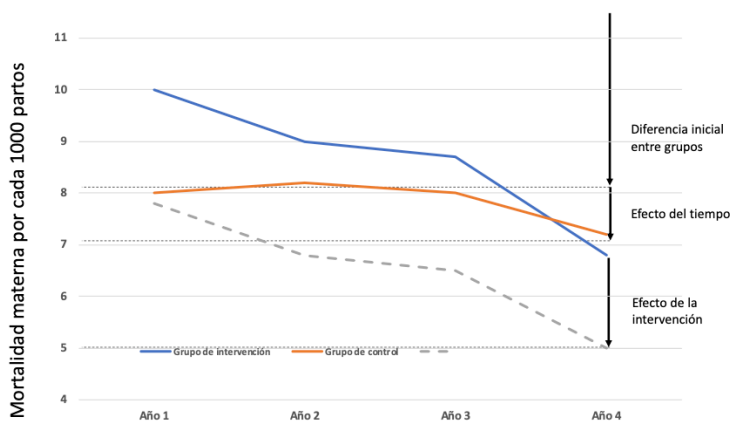


Figura 2. Cambio en la mortalidad materna en ambos grupos – ensayo QUARITE. Nota: La línea de puntos representa el cambio en la mortalidad materna en el grupo de intervención si su nivel de base (antes de la intervención) hubiera sido equivalente al del grupo de control.

Aplicación del método

Teniendo en cuenta el efecto del tiempo: el enfoque de “diferencia de diferencias”

La diferencia de mortalidad materna entre los dos períodos del grupo de control representa la tendencia natural de la mortalidad materna en los hospitales de referencia de Malí y Senegal (figura 2). Esta tendencia es pequeña (-1,0 muertes por cada 1.000 partos) durante el período de estudio, pero no es insignificante. Refleja el efecto de la modificación de las características de los pacientes a lo largo del tiempo, pero también los cambios en el sistema de salud. La cesárea gratuita iniciada en 2005 en Senegal y luego en 2006 en Malí probablemente facilitó el acceso a los servicios de salud y mejoró gradualmente el resultado de los partos entre 2007 y 2011. Además, la supervisión periódica de los equipos de atención de la salud durante el período de prueba para mejorar la calidad de los datos recogidos, puede haber contribuido a mejorar la atención de rutina en los hospitales de ambos grupos y el tratamiento de los pacientes. Por todas estas

razones, los resultados de salud materna mejoraron durante el estudio en el grupo de control, aunque no recibieron ninguna intervención. El método de diferencias en diferencias (DD) permite tener en cuenta este efecto del tiempo en el estudio del impacto.

En el enfoque DD, el efecto de la intervención se mide por la diferencia de los cambios observados entre los dos grupos (véase la tabla 3). Los resultados pueden expresarse en términos de diferencia en los riesgos de mortalidad o en los odds ratios (OR).

En cuanto a la diferencia de riesgo, el efecto de la intervención es la disminución de las tasas de mortalidad materna en el grupo experimental (-3,5 muertes por 1000) menos la disminución de la mortalidad en el grupo de control (-1,0 por 1000). Por lo tanto, en este ejemplo, se estima que el DD es de $(-3,5) - (-1,0) = -2,5$ muertes por cada 1.000 partos (IC 95% = -3,5; -1,5 por cada 1.000 partos). En otras palabras, una intervención en un hospital con 1.000 partos evitaría 2,5 muertes maternas en comparación con otro hospital con el mismo volumen de actividad pero que no se beneficia del programa. Este efecto es estadísticamente significativo porque el intervalo de confianza del 95% del DD no contiene 0.

Cuando los resultados se expresan como odds ratios (OR), el efecto de la intervención es la relación entre los odds ratios de los dos grupos. En nuestro caso, las probabilidades para un grupo determinado en un momento dado es el número de muertes maternas dividido por el número de mujeres vivas antes del alta hospitalaria. Por lo tanto, el odds ratio (OR) entre los dos períodos es el cambio en el riesgo de mortalidad materna entre el año 1 y el año 4. Debido a que la mortalidad disminuyó en ambos grupos entre los dos períodos de estudio, los OR son inferiores a 1 (tabla 2). Sin embargo, la disminución fue mayor en el grupo experimental (OR = 0,66) que en el grupo de control (OR = 0,89). La relación entre los dos OR, es igual a 0,73 (IC del 95% = 0,61; 0,91), mide el efecto de la intervención, que de nuevo es estadísticamente significativo porque el intervalo de confianza del OR no contiene 1. En otras palabras, la intervención redujo la mortalidad materna en un 27% en comparación con los cambios observados en el grupo de control.

Tabla 3 : Riesgos de mortalidad materna por grupo y período de tiempo – Ensayo QUARITE

	Año 1 Muerte p. 1000 (recuento)	Año 4 Muerte p. 1000 (recuento)	Diferencia de riesgo p. 1000 (95 % CI)	Proporción de probabilidades (95 % CI)
Grupo de intervención	10,4 (445/43269)	6,8 (356/52662)	- 3, 5 (- 2,3; - 4,7)	0,64 (0,57; 1,15)
Grupo control	8,1 (337/41655)	7,1 (381/53581)	- 1, 0 (- 0,1; - 2,0)	0,88 (0,76; 1,16)

Los resultados brutos estimados con el enfoque DD, ya sea en términos de diferencia de riesgo o de proporción de probabilidades, muestran que la intervención es efectiva para reducir la mortalidad materna hospitalaria en un sistema de salud como el de Malí o Senegal. Aunque el impacto medido parezca pequeño, estos resultados serían lo suficientemente significativos como para convencer a un responsable de la toma de decisiones de aplicar un programa de este tipo a nivel nacional en todos los hospitales. Sin embargo, esta persona podría cuestionar los resultados del estudio, que no tienen en cuenta el desequilibrio inicial entre los grupos y la correlación de los datos. En los dos párrafos siguientes veremos cómo tener en cuenta el efecto de grupo y el efecto de racimo.

Teniendo en cuenta el efecto del grupo

La figura 2 muestra que el nivel de mortalidad materna en el año 1 fue mayor en los hospitales del grupo de intervención que en el grupo de control. La disminución más rápida de la mortalidad en el brazo experimental redujo entonces la tasa de mortalidad materna a casi el mismo nivel que en el grupo de control en el año 4. Si hubiéramos comparado los resultados del ensayo entre los dos grupos en el año 4, habríamos llegado a la conclusión errónea de que la intervención no tuvo ningún efecto en la mortalidad.

El enfoque DD, que analizó los cambios en ambos grupos simultáneamente, abordó este problema, pero no controló el desequilibrio inicial entre los grupos. Debido a que los hospitales fueron asignados al azar antes de la inclusión de los pacientes, sus características no se distribuyeron uniformemente en los brazos del ensayo.

Tabla 4 : Características de los pacientes por período de tiempo y grupo – Ensayo QUARITE

	Año 1		Año 4	
	Intervención (n=43269)	Control (n=41655)	Intervención (n=52662)	Control (n=53581)
Vive lejos del hospital (fuera de la región)	2 242 (5,2)	509 (1,2)	1 115 (2,1)	854 (1,6)
Evacuado de otra institución	11 644 (26,9)	9 384 (22,5)	15 382 (29,2)	13 097 (24,4)
Edad ≥ 35 años	4 356 (10,1)	4 195 (10,1)	5 325 (10,1)	5 210 (9,7)
Nulíparos	28 435 (65,7)	26 939 (64,7)	35 908 (68,2)	36 397 (67,9)
Antecedentes de cesárea	3 112 (7,2)	2 782 (6,7)	5 196 (9,9)	4 920 (9,2)
Patología antes del embarazo	406 (0,9)	324 (0,8)	516 (1,0)	648 (1,2)
No hay visitas prenatales	4 221 (9,8)	4 535 (10,9)	4 780 (9,1)	5 250 (9,8)
Patología durante el embarazo	3 976 (9,2)	3 401 (8,2)	3 897 (7,4)	4 834 (9,0)
Embarazo múltiple	1768 (4,1)	1555 (3,7)	2 139 (4,1)	2 135 (4,0)

En nuestro caso, la proporción de mujeres que residían lejos del hospital donde dieron a luz era mayor en el brazo experimental que en el grupo de control y, como resultado, las evacuaciones médicas eran más frecuentes entre los hospitales que recibían la intervención (tabla 4). Habida cuenta de los malos resultados de los sistemas de atención de la salud tanto en Malí como en Senegal y de las demoras en la atención médica, el estado de salud de las mujeres del brazo experimental fue en general más deficiente y su riesgo de mortalidad más elevado. El desequilibrio entre los grupos parece disminuir en el año 4, lo que plantea la cuestión del sesgo al que puede conducir en el análisis de la efectividad de la intervención según el enfoque

DD: ¿la mayor reducción de la mortalidad materna en el brazo experimental no se explica simplemente por el cambio en el perfil de los pacientes de este grupo?

Para responder a esta pregunta pueden utilizarse varios métodos estadísticos, teniendo en cuenta los factores de confusión del estudio de impacto. Los modelos de regresión logística son particularmente adecuados para este tipo de análisis, ya que permiten ajustar la medida del efecto (DD u odds ratio) a las características individuales o de grupo, como el lugar de residencia o la forma en que los pacientes son admitidos en el hospital. Veremos en el siguiente párrafo que los modelos de regresión logística generalizada permiten tener en cuenta al mismo tiempo el efecto de conglomerado.

Consideración del efecto de grupo

Dado que las pacientes que dan a luz en un mismo hospital no son completamente independientes entre sí, la información comunicada por un hospital de 1.000 mujeres, por ejemplo, es por lo tanto menor que si tuvieran 1.000 pacientes independientes que dieran a luz en centros diferentes. Esto da lugar a una pérdida de poder en el análisis estadístico que debe compensarse con un aumento del número de sujetos a incluir. Además, en el análisis de los datos de conjunto, se deben utilizar métodos estadísticos apropiados, pues de lo contrario los resultados pueden ser indebidamente significativos. Si no se tiene en cuenta este efecto de agrupación, se disminuirán falsamente los intervalos de confianza, porque las pruebas estadísticas no tendrán en cuenta las variaciones inter e intrahospitalarias de la mortalidad materna. Hay varios métodos que pueden utilizarse para analizar adecuadamente este tipo de datos, incluido el uso de modelos mixtos o marginales. El comando `diff` del programa STATA (versión 13.0), por ejemplo, permite estimar el DD teniendo en cuenta la estructura jerárquica de los datos, organizados en varios niveles anidados (nivel 1: pacientes; nivel 2: hospital) y posibles variables de confusión. El comando `xtlogit` del mismo programa informático permite estimar la relación de probabilidades utilizando el mismo enfoque multivariante y multinivel.

Tabla 5 : Estimaciones crudas y ajustadas de los efectos de la intervención – ensayo QUARITE

Método de medición	Estimación bruta (95% CI)	Estimación ajustada* (95% CI)
Diferencia en los riesgos	-2,5 (- 3,5; - 1,5)	-2,5 (- 4,2; - 0,9)
Ratios de probabilidad	0,73 (0,61; 0,91)	0,85 (0,73; 0,98)
* efecto ajustado sobre el efecto de grupo y sobre las características del paciente y del hospital		

Después de ajustar las variables de confusión y el efecto de conglomerado, la relación OR disminuyó de 0,73 (estimación bruta) a 0,85 (estimación ajustada), como se muestra en la tabla 5. En otras palabras, el desequilibrio de algunas características individuales, como el nivel de accesibilidad a los hospitales, explica en parte la mayor disminución de la mortalidad materna en el brazo experimental en comparación con el brazo de control, pero no del todo. Así pues, puede atribuirse a la intervención una disminución relativa del 15% de la mortalidad materna. Además, después de ajustar el efecto de conglomerado, los intervalos de confianza son mayores que en las estimaciones brutas, independientemente del método de medición utilizado (diferencia de riesgo o coeficiente de probabilidades). Así pues, se ha perdido cierta precisión en las medidas del impacto, pero las diferencias siguen siendo significativas dados los grandes tamaños de las muestras.

Análisis reflexivo y conclusión

En este capítulo hemos visto el valor de realizar un ECA por conglomerados para evaluar el impacto de las intervenciones de salud materna y cómo controlar los tres principales sesgos en la medición de la eficacia: el efecto temporal, el efecto de grupo y el efecto de conglomerado. Hay otras características de los ensayos de grupos que abordaremos en el debate, pero no entraremos en detalles. Estas incluyen las siguientes preguntas: ¿Está el análisis cegado? ¿Se requiere el consentimiento del sujeto? ¿Se pueden generalizar los resultados a otros contextos?

Cuando la intervención que queremos probar toma la forma de un programa de actividades dirigidas a grupos de individuos (como el programa de mejora de la calidad en el ensayo QUARITE que se dirigió a los profesionales de la salud), entonces la inclusión de los sujetos se hace inevitablemente con el conocimiento del brazo en el que serán incluidos. Se dice que el ensayo es “abierto”, a diferencia de la mayoría de los ensayos individuales que son “ciegos”, es decir, sin conocimiento del brazo en el que se asignan los sujetos al azar. Esta situación es potencialmente una fuente de sesgo que puede ser inducida por un comportamiento posiblemente diferente entre los grupos.

La cuestión de la obtención del consentimiento para participar en el estudio también es importante. Si bien el consentimiento del jefe del grupo es esencial, no se requiere necesariamente el consentimiento de los sujetos. Es posible informar a las personas que pueden participar en el estudio sin obtener su consentimiento firmado. En el ensayo de QUARITE, por ejemplo, se obtuvo el consentimiento del director del hospital y del jefe de la sala de maternidad. Sin embargo, no lo hicimos en el caso de las mujeres que dieron a luz en cada uno de los hospitales participantes en el estudio, ya que no intervenimos directamente con ellas y los datos anonimizados recogidos se tomaron de los registros de los hospitales existentes, sin entrevistar a las mujeres.

Otra característica singular del ensayo por conglomerados es que el contexto en el que se aplica la intervención puede variar considerablemente de un conglomerado a otro. Si se conocen de antemano las características de las agrupaciones que afectan al resultado primario, se recomienda estratificar la aleatoriedad con arreglo a esos criterios. En el ensayo QUARITE, asignamos al azar los hospitales según el nivel de atención porque los datos preliminares disponibles indicaban que las tasas de mortalidad diferían entre los hospitales de distrito, regionales y de la capital. El análisis de los resultados por nivel de atención demostró que la intervención fue particularmente efectiva en los hospitales de distrito y poco en los hospitales regionales. Por consiguiente, llegamos a la conclusión de que los beneficios de la intervención deberían aplicarse únicamente a los hospitales de primera línea y que se necesitarían más investigaciones para confirmar si ese programa sería útil o no en los hospitales de nivel superior.

Aunque su aplicación es compleja, los ensayos controlados aleatorios son los estudios de impacto en la salud materna y perinatal que proporcionan el nivel más alto de pruebas si se realizan adecuadamente. Por razones logísticas, organizativas y, a veces, éticas, suele ser más apropiado aleatorizar unidades o grupos “sociales” en lugar de individuos. Los ensayos de agrupación implican desafíos metodológicos que no son insuperables, siempre que se tengan en cuenta en el diseño y el análisis del estudio. También hay que tener en cuenta que, en este tipo de estudio, la unidad de aleatorización es diferente de la unidad de análisis, es decir, la unidad que se tendrá en cuenta para la interpretación de los resultados o el análisis de los impactos. Pero el nivel de pruebas de la efectividad de una intervención no es el único criterio para los responsables de la toma de decisiones. El contexto y

las condiciones de aplicación son también elementos esenciales para diseñar políticas públicas más eficaces. Las investigaciones futuras también deberían ayudar a comprender cómo y por qué esta intervención produce efectos positivos en la salud de las madres y los recién nacidos. Los ensayos híbridos que combinan tanto los ensayos de control de agrupaciones como el análisis de la aplicación son los más adecuados para abordar simultáneamente las cuestiones de eficacia y aplicación.

Referencias clave

Campbell, M. K., Mollison, J., Steen, N., Grimshaw, J. M. et Eccles, M. (2000) Analysis of cluster randomized trials in primary care: a practical approach. *Family Practice*, 17(2), 192–196.

Un artículo que presenta muy claramente el método de los ensayos aleatorios en conglomerados.

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser, W. (2013). Quality of care, risk management, and technology in obstetrics : a cluster-randomized trial of a multifaceted intervention to reduce hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146 – 157.

Para ilustrar este capítulo se utilizaron los resultados de este estudio realizado en Senegal y Malí.

Chaillet, N., Dumont, A., Abrahamowicz, M., Pasquier, J. C., Audibert, F., Monnier, P., ... et Fraser, W. D. (2015). A cluster-randomized trial to reduce cesarean delivery rates in Quebec. *New England Journal of Medicine*, 372(18), 1710–1721.

Los resultados de este estudio canadiense ilustran los retos metodológicos asociados a los ensayos con conglomerados.

Referencias

Dumont, A., Fournier, P., Abrahamowicz, M., Traoré, M., Haddad, S. et Fraser W. (2013). Quality of care, risk management, and technology in obstetrics

: a cluster-randomized trial of a multifaceted intervention to reduce hospital-based maternal mortality in Senegal and Mali (QUARITE). *Lancet*, 382(9887), 146–157

Stanton, C. K., Newton, S., Mullany, L. C., Cofie, P., Agyemang C. T., Adiibokah, E., ... et Gyapong, J. (2013). Effect on postpartum hemorrhage of prophylactic oxytocin (10 IU) by injection by community health officers in Ghana: A community-based, cluster-randomized trial. *PLoS Medicine*, 10(10), e1001524.
doi: 10.1371/journal.pmed.1001524

Villar, J., Ba'aqeel, H., Piaggio, G., Lumbiganon, P., Belizán, J. M., Farnot, U., ... et Garcia, J. (2001). WHO antenatal care trial research group. *Lancet*, 357(9268), 1551–64.

Resumen / Résumé / Abstract

Los ensayos controlados aleatorios grupales permiten evaluar el impacto de las intervenciones de salud con un nivel muy alto de pruebas. Aunque los ensayos aleatorios grupales ofrecen varias ventajas en los países de bajos recursos con respecto a los ensayos individuales, presentan algunos desafíos metodológicos debido al hecho de que la unidad de análisis es diferente de la unidad de aleatorización o de intervención. Estos desafíos, tanto en la implementación de la intervención como en el análisis estadístico, se presentan sobre la base de un experimento realizado en Malí y Senegal (ensayo QUARITE) para reducir la mortalidad materna y neonatal en los hospitales. Veremos cómo el enfoque de “diferencia de diferencias” permite medir el efecto de la intervención teniendo en cuenta el efecto del tiempo, que a menudo tiene un impacto positivo en los indicadores de salud fuera de cualquier intervención externa. Se presentan modelos estadísticos apropiados para ajustar el efecto de la intervención sobre las diferencias entre los grupos experimentales antes del inicio del programa de prueba y sobre el efecto del grupo específico de los ensayos de grupos.

Les essais contrôlés randomisés en grappe permettent d'évaluer l'impact des interventions en santé avec un niveau de preuve très élevé. Si les essais randomisés en grappes offrent plusieurs avantages dans les pays à faibles ressources par rapport aux essais individuels, ils présentent quelques défis

méthodologiques qui tiennent au fait que l'unité d'analyse est différente de l'unité de randomisation ou d'intervention. Ces défis tant sur le plan de la mise en œuvre de l'intervention que de l'analyse statistique sont présentés à partir d'une expérience réalisée au Mali et au Sénégal (essai QUARITE) pour réduire la mortalité maternelle et néonatale hospitalière. Nous verrons comment l'approche « différence des différences » permet de mesurer l'effet l'intervention tout en tenant compte de l'effet du temps qui joue souvent favorablement sur les indicateurs de santé en dehors de toute intervention extérieure. Les modèles statistiques appropriés sont présentés pour pouvoir ajuster l'effet de l'intervention sur les différences qui existent entre les groupes expérimentaux avant le démarrage du programme à tester et sur l'effet de groupe propre aux essais en grappe.

Cluster-randomised trials are used to assess the impact of health interventions with a very high level of evidence expected. While cluster randomized trials offer several advantages in low-resource countries compared to individual trials, they present some methodological issues because the unit of analysis is different from the unit of randomization and intervention. These challenges, both in terms of the intervention implementation and the statistical analysis, are presented based on a trial conducted in Mali and Senegal (QUARITE trial) to reduce maternal and neonatal mortality in hospitals. We will see how the difference-in-differences approach measures the effect of the intervention taking into account the secular trends, which often plays favorably on health indicators outside of any external intervention. The appropriate statistical models are presented in order to adjust the effect of the intervention on baseline case-mix and on the cluster effect which is specific to this type of trial.

Alexandre Dumont es ginecólogo-obstetra y epidemiólogo. Ha llevado a cabo ensayos controlados por conglomerados en materia de salud materna y perinatal en el Canadá, Senegal, Malí y Burkina Faso. Se le concedió el Premio de Cumpleaños de la Comisión Europea por el ensayo QUARITE que redujo la mortalidad materna y neonatal en los hospitales de referencia de Malí y Senegal.

Cita

Alexandre Dumont (2020). Ensayos aleatorizados por conglomerados. Ejemplo dirigido a la salud materna e infantil. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 277-296. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

II. Medición de equidad

Un ejemplo de intervención de gratuidad de la atención obstétrica

TARIK BENMARHANIA Y BRITT MCKINNON

Medición de equidad

Definición del método

Definimos la equidad de los efectos de una intervención en función de la identificación de una heterogeneidad en los beneficios potenciales entre diferentes subgrupos o territorios de una población. Una intervención se evaluará como equitativa si los beneficios son mayores dentro de los subgrupos o territorios que se han definido como los más vulnerables o con las mayores necesidades.

Fortalezas del método

El método propuesto hace posible cuantificar las desigualdades en salud para varias configuraciones (grupos ordenados o no ordenados) en una escala absoluta o relativa.

Desafíos del método

Es importante tener un buen nivel de conocimiento de las poblaciones vulnerables y estudiarlas empírica y teóricamente. Para algunos de los métodos propuestos (por ejemplo, análisis de descomposición, etc.), a veces es necesario tener muestras grandes para identificar una heterogeneidad en los beneficios potenciales de una intervención dentro de dos o más grupos.

La equidad en salud se identifica ampliamente como un objetivo al que las políticas y actividades de salud pública deben responder como una prioridad (Starfield 2007; Marmot, Friel et al. 2008; Blas y Kurup 2010). Sin embargo, existe una gran diversidad en las definiciones del concepto de equidad en salud y tales implicaciones se discuten ampliamente (Whitehead 1992; Braveman y Gruskin 2003; Braveman, Cubbin et al. 2005). A veces, esta noción, a menudo definida a través de su opuesto, que es la inequidad o la desigualdad, se refiere a la descripción de diferencias evitables o remediables entre diferentes grupos de personas (Whitehead 1992; Kawachi, Subramanian et al. 2002), por lo que incluye una dimensión moral y ética (Daniels 2006; Smith 2015). A veces, se refiere a las consecuencias de una intervención de salud pública. De hecho, en esta última dimensión, la noción de inequidad implica una intervención que aumentará las desigualdades de

salud ya existentes. Estas dos nociones a veces se fusionan, pero a menudo las implicaciones y suposiciones para las teorías de justicia social subyacentes son escasamente discutidas. (Smith 2015).

En este capítulo, proponemos una definición de equidad en salud que se refiere a la distribución de los beneficios de las políticas públicas.

Esta definición implica, en primer lugar, la identificación de subgrupos o territorios de una población para la cual el estado de salud se considerará más negativo o dentro del cual el impacto de un determinante de salud será más importante (noción epidemiológica de modificación del efecto) en comparación con otro subgrupo o territorio de la misma población. Llamaremos a estos subgrupos de una población “poblaciones vulnerables”, en comparación con las poblaciones no vulnerables.

Por lo tanto, en este capítulo, definimos la equidad de los efectos de una intervención en función de la identificación de una heterogeneidad en los beneficios potenciales entre diferentes subgrupos o territorios de una población. Una intervención se evaluará como equitativa si los beneficios son mayores dentro de los subgrupos o territorios que se han definido como vulnerables.

La noción de equidad en salud se ha discutido durante varias décadas, pero se implementó relativamente poco. Entendido en el contexto de una distribución desigual de los efectos de las intervenciones de salud pública, su evaluación es minoritaria en comparación con la evaluación de los efectos en toda una población. Para ilustrar esto, podemos tomar dos ejemplos. Las evaluaciones de los efectos de las políticas públicas de control del tabaco se han estudiado ampliamente en la literatura epidemiológica (Hoffman y Tan 2015) su número se cuenta por centenares. Sin embargo, la gran mayoría de estos estudios se refieren principalmente a una población en su conjunto (Frazer, Callinan et al. 2016) o un subgrupo de la población como los/las adolescentes (Gilman, Rende et al. 2009) o algunos/as profesionales (Farrelly, Nonnemaker et al. 2005). Los estudios que evalúan las desigualdades en los efectos de una política pública de control del tabaco entre dos o más subgrupos de una población son mucho más raros (Brown, Platt et al. 2014a; Brown, Platt et al. 2014b) y se refieren, por ejemplo, a las desigualdades por sexo, edad o nivel de educación. El segundo ejemplo se refiere a intervenciones para combatir la contaminación del aire en áreas urbanas (Cartier, Benmarhnia et al. 2015). En una revisión sistemática

(Benmarhnia, Rey et al. 2014), pudimos destacar que de más de 50 estudios publicados, sólo dos estudios evaluaron formalmente la heterogeneidad de los efectos de la intervención dentro de diferentes subgrupos y ocho estudios permitieron una evaluación *a posteriori* entre las estimaciones de diferentes estratos. Estos dos ejemplos destacan, por un lado, el pequeño número de estudios que han investigado esta pregunta de investigación específica, a pesar de la gran disponibilidad de datos que permiten realizar dichos estudios y, por otro lado, la falta de métodos apropiados para evaluar la equidad en salud. La mayoría de los estudios basan sus conclusiones en comparaciones cualitativas entre diferentes estratos de población o mediante una prueba de interacción en un modelo de regresión.

En este capítulo presentamos diferentes métodos para medir las desigualdades en salud y su aplicación en la evaluación de los efectos de las intervenciones utilizando el ejemplo del efecto de la gratuidad en el acceso a la atención obstétrica, como parte de un ejemplo simulado.

Primero presentamos el contexto de evaluación describiendo la intervención considerada, los datos estudiados, y las justificaciones para un posible efecto heterogéneo de esta intervención en el acceso a la atención obstétrica. A continuación, presentamos diferentes métodos para medir las desigualdades en salud y diferentes enfoques para aplicarlos a la evaluación de los efectos de las intervenciones, por lo tanto, a la equidad tal y como se define en este capítulo.

Presentación del contexto de la evaluación de la equidad: el ejemplo de exención del pago de los costos de la atención (atención gratuita) y su impacto en el acceso a la atención obstétrica

El acceso a la atención básica sigue siendo un problema fundamental en muchos países de bajos y medianos ingresos (Althabe, Bergel et al. 2008; Lagarde y Palmer 2008; Dzakpasu, Powell-Jackson et al. 2013; McKinnon, Harper et al. 2014; McKinnon, Harper et al. 2015). En el contexto de los sistemas de salud en gran medida insuficientemente financiados, estos países se enfrentan a un dilema. Por un lado, pagar los costos de atención médica puede presentar una barrera financiera para acceder a la atención médica. Por otro lado, una disminución en los recursos financieros de los sistemas de salud puede contribuir a una disminución en la calidad de los

servicios de salud, lo que también puede constituir barreras financieras para la asistencia cualificada al parto. Algunos estudios han demostrado que establecer el pago de los costos de atención médica contribuye a aumentar los recursos a nivel local y a mejorar la calidad de los servicios (Ellis 1987; Litvack y Bodart 1993). Otros estudios han mostrado efectos negativos, particularmente en términos de equidad (Yoder 1989; Russell y Gilson 1997). Se han llevado a cabo numerosas campañas internacionales para promover la atención médica gratuita (por ejemplo, exención del costo del parto) (Yates 2009; Ridde y Morestin 2011).

En este contexto, proponemos tomar como ejemplo de una intervención la de la implementación de la atención obstétrica gratuita (es decir, la exención del precio de la atención al parto) y evaluar su impacto y su equidad en el acceso a la atención obstétrica. Para evaluar los efectos sobre la equidad, consideraremos dos dimensiones con respecto a una posible heterogeneidad en los beneficios de la intervención. Primero, consideraremos una dimensión vinculada a las desigualdades materiales que mediremos mediante un índice de riqueza utilizado a menudo en las encuestas internacionales de población (Filmer y Pritchett 1999; Howe, Hargreaves et al. 2009). En segundo lugar, consideraremos una dimensión vinculada a las desigualdades sociales al medir el nivel de educación de la madre. Esta elección se justifica por las desigualdades sociales y materiales en términos de acceso a la atención obstétrica en un contexto de países de medianos y bajos ingresos (Say y Raine 2007). Por lo tanto, definiremos la intervención como equitativa si contribuye a reducir las desigualdades materiales y / o sociales de acceso a la atención obstétrica.

El ejemplo que presentamos a continuación proviene de datos simulados de encuestas de población y no corresponde a una situación de intervención real. Existen muchos ejemplos de este tipo de intervención (Lagarde y Palmer 2008). Nos centramos en las desigualdades socioeconómicas, pero la definición de equidad adoptada en este capítulo también puede aplicarse a otros factores demográficos (edad, sexo, género) o territoriales (urbanos versus rurales).

Métodos para medir las desigualdades en salud

Se pueden usar varios indicadores para medir las desigualdades en salud. Sin embargo, no hay consenso sobre una medida en particular, especialmente cuando se trata de medir variaciones en la desigualdad a lo largo del tiempo. Esta sección comienza con una discusión de algunos temas importantes con respecto a la medición de las desigualdades en salud, particularmente las desigualdades sociales. En la siguiente sección, describiremos diferentes indicadores para medir las desigualdades en salud con ejemplos para ilustrarlos.

Desafíos en la medición de las desigualdades en salud

Desigualdades absolutas vs relativas

En primer lugar, es importante distinguir entre las medidas de desigualdad relativa y absoluta, ya que ambas proporcionan información importante, pero fundamentalmente diferente, sobre la presencia y el alcance de las desigualdades. Las medidas absolutas reflejan la magnitud de las diferencias de salud entre dos grupos. Por ejemplo, las tasas de mortalidad cardiovascular en 2014 fueron de 12 por 1000 hombres y 8 por 1000 mujeres, la medida absoluta de desigualdad sería 4 por 1000 x año. Paralelamente, las medidas relativas reflejan una diferencia proporcional en salud entre dos grupos. En el ejemplo anterior, representaría la relación de las dos tasas, es decir, $12/8 = 1.5$. Debido a que estas dos medidas pueden llevar a conclusiones separadas con respecto a la magnitud y los cambios en la desigualdad, es importante relacionar las dos escalas para presentar una imagen completa de la desigualdad (Lynch, Smith et al. 2006 ; Houweling, Kunst et al. 2007; Harper, King et al. 2010).

Clasificar grupos sociales

Algunos grupos sociales “naturalmente” tienen dos grupos (hombre / mujer; urbano / rural), mientras que otros grupos tienen varios niveles (nivel

de educación, etnia). Para las agrupaciones sociales con varios grupos, la elección de los indicadores de desigualdad que se adoptarán dependerá de si los grupos están organizados de acuerdo con una estructura ordenada o no. El nivel de educación o el índice de nivel de riqueza constituyen, por ejemplo, agrupaciones ordenadas, mientras que las características étnicas o geográficas no se ordenan a priori. Las comparaciones por pares se pueden hacer sistemáticamente entre dos grupos para cualquier tipo de variable social. Además, el alcance de las desigualdades sociales para varios grupos (ordenados o no) se puede medir utilizando indicadores más complejos como el índice de desigualdad de la pendiente (*slope index of inequalities*) o el índice de Theil.

Medidas de desigualdades en salud

Indicadores de desigualdad simples (comparación por pares)

Las medidas estándar para comparar el estado de salud entre dos grupos sociales son la diferencia y la razón. Estas dos medidas son complementarias. Si el indicador de salud se expresa como una tasa (por ejemplo, mortalidad por cáncer por 100,000 personas por año), las medidas absolutas y relativas corresponden respectivamente a una diferencia de tasa y una razón de tasa, comparando dos grupos sociales de interés.

Para los grupos sociales ordenados por nivel de educación, por ejemplo, las comparaciones por pares generalmente se hacen comparando grupos en los extremos. Por ejemplo, si consideramos un índice de riqueza estratificado en quintiles, compararemos el quintil más rico con el quintil más pobre. Sin embargo, este tipo de comparación ignora el estado de salud de los otros tres quintiles, que podrían proporcionar información importante sobre las desigualdades dentro de la población. La Figura 1 muestra la proporción de mujeres que recibieron atención obstétrica de un profesional cualificado por quintil de un índice de riqueza para dos poblaciones hipotéticas. Las dos poblaciones tienen los mismos valores de diferencia o razón al comparar los dos quintiles extremos, mientras que el retrato dentro

de los otros tres quintiles es diferente. Es por eso que generalmente se prefieren medidas más complejas de desigualdad, considerando todos los grupos.

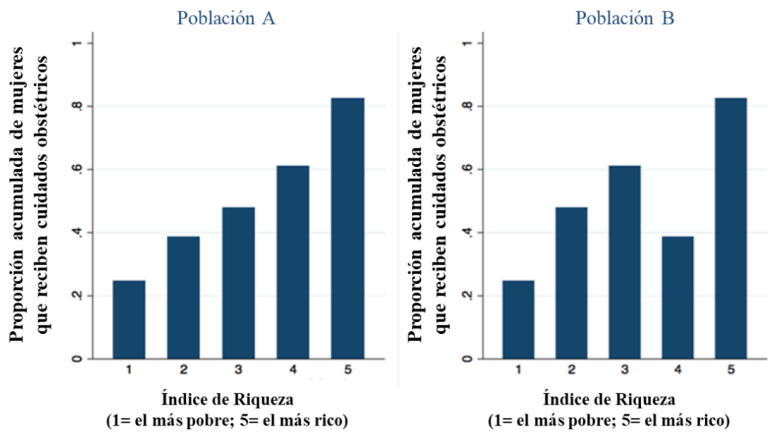


Figura 1. Proporción de mujeres que reciben cuidados obstétricos por parte de un profesional cualificado por quintiles de Índice de Riqueza para dos poblaciones hipotéticas.

El índice de inclinación de la desigualdad y el índice de concentración son dos medidas que a menudo se usan para medir el gradiente de salud entre los grupos sociales ordenados. A diferencia de las medidas simples presentadas anteriormente, estos índices tienen en cuenta el tamaño relativo de cada grupo.

El índice de inclinación de la desigualdad

El índice de inclinación de la desigualdad mide la diferencia absoluta de los valores pronosticados entre los grupos más altos y más bajos en la distribución de la posición social para los grupos ordenados (educación, índice de riqueza, etc.) (Pamuk 1985). Este índice se estima a partir del modelo de regresión apropiado para la medida de salud considerada (por ejemplo, una regresión lineal para una medida continua y una regresión logística para una medida de salud binaria). Es posible calcular este índice

utilizando el índice de riqueza continua, por ejemplo, yendo del más pobre (rango = 0) al más rico (rango = 1). La medida de salud se incluye luego en un modelo de regresión al incluir, como predictor, el rango de cada individuo dentro de la distribución acumulativa de riqueza. Por lo tanto, el coeficiente correspondiente a este rango representará la diferencia de salud estimada entre los individuos más pobres y más ricos de la población.

La Figura 2 es una representación gráfica de este índice, que muestra el valor pronosticado del resultado de salud (aquí, si recibe atención obstétrica por parte de un profesional cualificado) en función de la proporción acumulada de rango por nivel de educación. La prevalencia estimada de recibir atención obstétrica (por un profesional cualificado) entre las mujeres con el nivel educativo más bajo (rango = 0) es del 17,7%. Este valor sería del 84% entre las mujeres con el nivel educativo más alto (rango = 1). Por lo tanto, el valor de este índice en términos porcentuales será de $84 - 17.7 = 66.3\%$, lo que significa que las mujeres en la parte superior de la distribución educativa tienen un 63% más de posibilidades de recibir atención obstétrica de un profesional cualificado en comparación con las mujeres en el nivel más bajo del nivel educativo de esta distribución.

Cuando los individuos se clasifican de los más desfavorecidos a los menos desfavorecidos, los valores positivos de este índice indican que la prevalencia de la medida de salud es más alta entre los individuos menos desfavorecidos, mientras que un valor negativo indica que esta prevalencia es más alta entre los más desfavorecidos. Un valor de 0 significa igualdad dentro del gradiente social. También se puede estimar una medida relativa de este índice calculando la razón entre los valores pronosticados de los dos extremos de la distribución. Según la Figura 2, este índice en la escala relativa sería $84 / 17.7 = 4.7$, lo que significa que la recepción de atención obstétrica por parte de un profesional cualificado es 4.7 veces mayor dentro de del grupo más favorecido en comparación con el grupo menos favorecido.

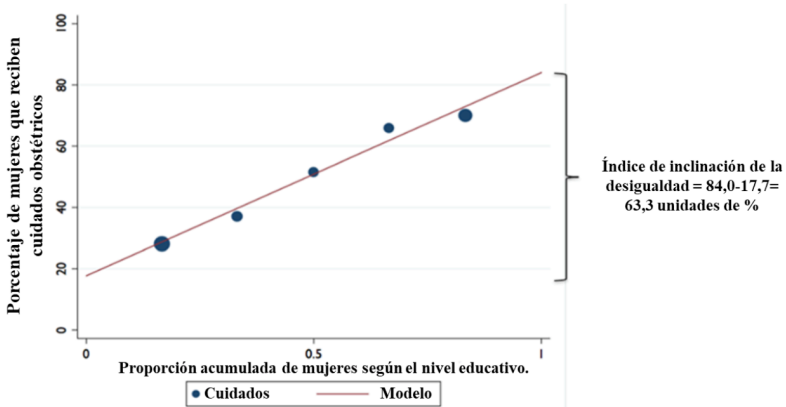


Figura 2. Índice de inclinación de la desigualdad mostrando las desigualdades de los cuidados obstétricos en función del nivel educativo.

Índice de concentración

El índice de concentración es una medida relativa de desigualdad basada en la curva de concentración, una cifra que representa la proporción acumulada de individuos clasificados por rango de salud en el eje X con la proporción acumulada en el eje Y. La figura 3 representa la curva de concentración con desigualdades de riqueza frente a la atención obstétrica utilizando los datos presentados en la Tabla 1. La línea a 45 ° representa la línea de igualdad cuando los recursos se distribuyen equitativamente entre la población. La curva de concentración se encuentra debajo de esta línea de igualdad, lo que indica una desigualdad desfavorable para las mujeres más pobres. Por ejemplo, podemos observar que el 40% más pobre de las mujeres se beneficia de solo alrededor del 25% del total de visitas obstétricas.

Tabla 1 : Cálculo de valores de proporción acumulada para generar la curva de concentración

Índice de Riqueza	Número de mujeres	Proportion de femmes	Proporción acumulada de mujeres	Número de mujeres que reciben Cuidados	Proporción de mujeres que reciben cuidados	Proporción acumulada de mujeres que reciben cuidados
1	1167	0,217	0,217	288	0,105	0,105
2	1017	0,190	0,407	394	0,143	0,248
3	990	0,184	0,592	475	0,172	0,420
4	989	0,184	0,776	604	0,219	0,640
5	1203	0,224	1,0	992	0,360	1,0

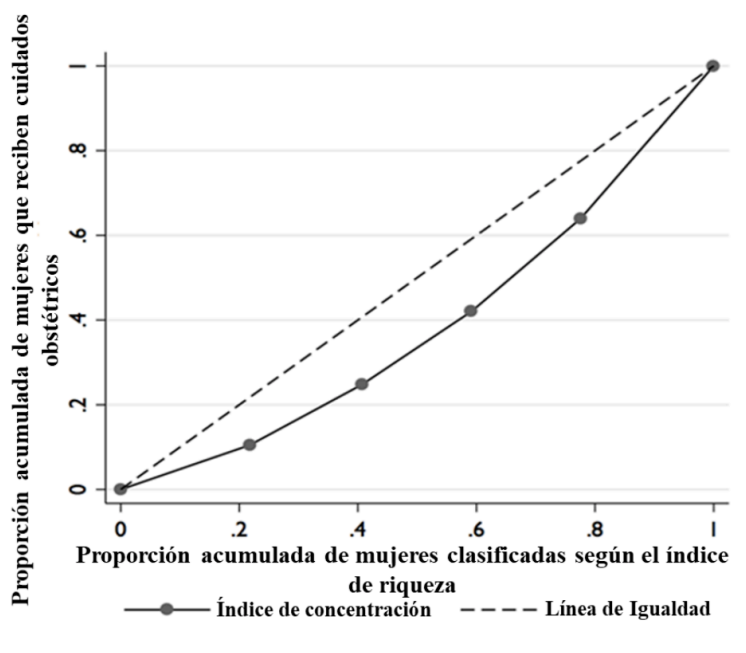


Figura 3. Curva de concentración representando las desigualdades relativas de percepción de cuidados obstétricos según el índice de riqueza

El índice de concentración se define como el doble del área entre la curva de concentración y la línea de igualdad (Wagstaff, Van Doorslaer et al.

1991). Este índice puede estar entre -1 y 1, donde 0 representa la igualdad completa. Los valores positivos significan desigualdades a expensas de los más pobres para una medida de salud positiva (como en este ejemplo). Para los datos agrupados por quintil, por ejemplo, como en la Tabla 2, el índice de concentración se calcula utilizando la siguiente fórmula:

$$C = (p_1L_2 - p_2L_1) + (p_2L_3 - p_3L_2) + \dots + (p_{T-1}L_T - p_TL_{T-1})$$

donde p_t es el porcentaje acumulativo de la muestra clasificada por índice de riqueza para el grupo t , y L_t corresponde al valor de distribución de la medida de salud por quintil (O'Donnell y Wagstaff 2008). La Tabla 2 utiliza los datos de la curva de concentración presentada anteriormente y la última columna presenta el cálculo del índice de concentración. Aquí este índice es 0.224.

Tabla 2 : Cálculo del índice de concentración

Índice de Riqueza (t)	Proporción acumulada de mujeres por quintil de riqueza (pt)	Proporción acumulada de mujeres que reciben cuidados (Lt)	Índice de Concentración
1	0,217	0,105	0,011
2	0,407	0,248	0,025
3	0,592	0,420	0,052
4	0,776	0,640	0,136
5	1,0	1,0	0,000
			0,224

El índice de concentración es una medida de desigualdad relativa, pero es posible obtener una versión absoluta de este índice multiplicando el índice de concentración obtenido por el promedio de la medida de salud. También es posible calcular este índice de concentración a nivel individual utilizando enfoques basados en regresión¹.

Un tema importante en la investigación sobre las desigualdades en salud se refiere a la comprensión de los factores que más contribuyen a las desigualdades en relación con una medida de salud dada. Una ventaja de este índice es que puede usarse en el contexto de análisis de descomposición.

1. Para más información, véase Kakwani, Wagstaff *et al.* (1997) y O'Donnell y Wagstaff (2008).

Este tipo de análisis cuantifica la contribución de diferentes factores a las desigualdades sociales con respecto a una medida de salud. Existen varios ejemplos en la literatura que van desde la comprensión de la contribución de los determinantes de las desigualdades sociales en la obesidad en Canadá (Hajizadeh, Campbell et al. 2014) hasta la de la contribución de las desigualdades sociales en la mortalidad infantil en Irán (Hosseinpour, Van Doorslaer et al. 2006)².

Medidas complejas de desigualdad con grupos desordenados – Índice de Theil

Los dos índices presentados anteriormente son válidos solo cuando se ordenan grupos sociales. Cuando los grupos sociales no están ordenados, es posible usar el índice de Theil (Theil 1967). Las medidas de desigualdad absoluta para categorías desordenadas incluyen el método de varianza (Harper, Lynch et al. 2008) entre grupos y el índice de disparidad (Pearcy y Keppel 2002) recientemente extendido al análisis de series temporales (Benmarhnia, Grenier et al. 2015).

El índice de Theil se estima a partir de la fórmula :

$T = \sum_{i=1}^n p_i r_i \ln(r_i)$ donde p_i es la proporción de la población en el grupo social i y r_i es la razón de la media de la medida de salud en el grupo i entre la media del conjunto de todos los grupos. Por lo tanto, si cada grupo tiene un valor promedio de la medida de salud igual al promedio de todos los grupos, entonces no hay desigualdad relativa y el índice de Theil será igual a 0. Este índice siempre es positivo. Cuanto mayor sea su valor, mayor será el nivel de desigualdad relativa entre los grupos.

Los datos en la Tabla 3 muestran cómo calcular el índice de Theil para medir una desigualdad relativa en el acceso a la atención obstétrica dentro de cinco regiones geográficas distintas.

2. Para más información sobre los análisis de descomposición del índice de concentración, véase Wagstaff, Van Doorslaer et al. (1991), Wagstaff, Van Doorslaer et al. (2003) y O'Donnell y Wagstaff (2008).

Tabla 3 : Cálculo del Índice de Theil del acceso a los cuidados obstétricos en cinco regiones geográficas

Région	Soins obstétricaux (%)	Proportion de population (p_i)	Ratio de % de soins dans la région par rapport à la population totale (r_i)	$\ln(r_i)$	Indice de Theil
A	45,8	0,29	0,90	-0,11	-0,028
B	55,7	0,12	1,09	0,09	0,011
C	48	0,18	0,94	-0,06	-0,010
D	82,5	0,22	1,62	0,48	0,175
E	22,8	0,19	0,45	-0,81	-0,067
Toutes régions	51,0				0,081

En esta sección, hemos presentado un retrato de diferentes medidas de desigualdades en salud. No es una lista exhaustiva de las medidas existentes para medir las desigualdades en salud, ni una discusión en profundidad de los problemas conceptuales. Para este último punto, recomendamos los trabajos de Kawachi y Kennedy (1997), Mackenbach y Kunst (1997) y Regidor (2004). Sin embargo, pudimos presentar varios escenarios que luego se pueden utilizar para evaluar la distribución de los efectos de las intervenciones de salud pública como se describe en la siguiente sección.

Aplicación a los estudios de evaluación de efectos

En esta parte, presentaremos cómo aplicar los métodos de medición de desigualdad presentados anteriormente a la evaluación de los efectos de una intervención de salud pública.

Presentamos dos tipos de enfoques: un enfoque de evaluación simple que permite ilustrar fácilmente los métodos presentados anteriormente y que también nos permitirá introducir pruebas de heterogeneidad y un enfoque basado en diseños cuasi-experimentales. Los detalles de los métodos de evaluación, particularmente con respecto a los enfoques cuasi-experimentales, se presentan en el capítulo anterior.

Enfoques simples antes / después

Un primer enfoque simple para evaluar si los beneficios de una intervención (aquí la implementación de la atención obstétrica gratuita) son equitativos es comparar si las desigualdades de acceso a la atención obstétrica entre los diferentes grupos se modifican antes y después la puesta en marcha de la intervención. Podemos tomar como ejemplo la comparación del primer quintil del índice de riqueza con el quinto quintil de este índice. Por lo tanto, al calcular la diferencia o la relación entre estos dos grupos, por un lado, antes de la implementación de la intervención y, por otro lado, después de su implementación, podemos obtener dos mediciones (en la escala absoluta o en la relativa) de acceso desigual a la atención obstétrica. Entonces es suficiente realizar una prueba de heterogeneidad entre estas dos medidas para evaluar si los beneficios han sido distintos entre los dos grupos y en qué dirección. Hay muchas pruebas de heterogeneidad que se pueden realizar en este contexto. Ofrecemos como ejemplo la prueba Cochran Q (Kaufman y MacLehose 2013). Esta prueba puede llevarse a cabo en mediciones absolutas o relativas. Esta prueba de heterogeneidad se construye de la siguiente manera. Si consideramos, por ejemplo, la escala relativa y la razón de prevalencia (RP) como una medida de asociación, debemos estimar una RP que represente el efecto de la intervención para cada uno de los grupos (razón de prevalencia de acceso a atención obstétrica antes y después de la intervención) RP_1 y RP_2 y para los dos grupos juntos (RPp). Luego calcule la siguiente estimación de Cochran Q:

$$Cochran's Q = \left[\frac{(\beta_1 - \beta_p)^2}{VAR(\beta_1)} + \frac{(\beta_2 - \beta_p)^2}{VAR(\beta_2)} \right]$$

En esta fórmula, β_1 representa el logaritmo natural de RP_1 y β_2 representa el logaritmo natural de RP_2 y VAR representa la varianza. β_p representa el logaritmo natural de RPp que es el efecto para los dos grupos juntos. Una vez que se calcula esta estimación de Cochran Q, es suficiente llevar a cabo una prueba de χ^2 (Kaufman y MacLehose 2013) con un número de grados de libertad igual al número de estratos menos uno.

Por lo tanto, si RP_1 representa el efecto de la intervención en el primer quintil del índice de riqueza (el más pobre) y RP_2 el efecto de la intervención en el quinto quintil (el más rico) y RPp para el dos grupos juntos, la intervención se considerará equitativa si: i) hay un efecto heterogéneo de

la implementación de la exención de los costos de atención médica en el acceso a la atención obstétrica; y ii) este efecto es más importante para el primer quintil que para el quinto.

Hemos presentado un enfoque muy simple para evaluar la posible heterogeneidad en los beneficios de una intervención. Este enfoque puede ampliarse para tener en cuenta los factores de confusión medidos en el contexto de la evaluación de los efectos de una intervención. Para esto, será posible realizar, por ejemplo, análisis de coincidencia (es decir, comparar individuos tratados y no tratados con características observables similares) utilizando un puntaje de propensión (Shadish, Luellen et al. 2006). Una vez que se haya estimado el puntaje de propensión y se haya realizado el emparejamiento (detalles no discutidos en este capítulo), será posible llevar a cabo el mismo enfoque para evaluar la heterogeneidad a través de la prueba Q de Cochran.

Sin embargo, en un entorno no experimental, a menudo es preferible utilizar enfoques cuasiexperimentales que se presentan en detalle en el capítulo anterior, para evaluar los efectos de una intervención. En la siguiente sección, proponemos diferentes enfoques para evaluar la heterogeneidad de los efectos de una intervención utilizando métodos cuasiexperimentales.

Enfoques cuasiexperimentales

El uso de métodos cuasiexperimentales se ha desarrollado ampliamente en los últimos años en salud pública, ya que tienen ventajas definitivas desde el punto de vista de la inferencia causal: permiten tener en cuenta factores de confusión no medidos. Esta consideración a veces puede ser de gran importancia en comparación con los enfoques de tipo antes / después, llegando a modificar la dirección de la asociación que representa el efecto de una exposición (Grabich, Robinson et al. 2015). Existen varios métodos, por ejemplo, el llamado método de diferencia en diferencias (DD) y el enfoque de regresión discontinua, que es un caso especial de análisis de variable instrumental, ambos presentados en el capítulo anterior. Aquí nos centraremos en el enfoque de tipo DD.

Además de la evaluación de los efectos en una población en su conjunto (Benach, Malmusi et al. 2011) o en un subgrupo de la población como las

mujeres, la evaluación de la equidad mediante el uso de métodos cuasi-experimentales sigue siendo poco frecuente. Aquí, proponemos tres enfoques distintos que permiten evaluar la heterogeneidad en la distribución de los efectos de una intervención con el ejemplo del enfoque DD. Para ilustrar, de manera deliberadamente simplificada, cómo se organizaría la evaluación de la exención del pago de los costos de atención en el acceso a la atención obstétrica utilizando un enfoque de tipo DD, podemos considerar la siguiente situación. Entre 2000 y 2010, medimos la prevalencia de acceso a la atención obstétrica para mujeres en dos países diferentes (países A y B) similares cada año. En 2005, el país A decidió organizar la exención del pago de los costos de atención (exención del precio de los partos) mientras que el país B continuó aplicando tarifas por los partos. Suponemos que los países son intercambiables (ver el capítulo anterior para este aspecto) y que el país B constituye un contrafactual válido para representar lo que habría sucedido en el país A sin la implementación de la intervención. El efecto causal de la intervención sobre el acceso a la atención obstétrica se medirá comparando la diferencia entre los dos países después de 2005 con la diferencia entre los dos países antes de 2005. Recordamos que este efecto se mide en los modelos de regresión por un término de interacción entre una variable que indica el período posterior a la intervención y una variable que indica el país con intervención (Dimick y Ryan 2014). Para evaluar la distribución de los efectos entre los diferentes subgrupos, consideramos los quintiles del índice de riqueza.

Análisis estratificados combinados con una prueba de heterogeneidad.

Un primer enfoque sería realizar el análisis de tipo DD por separado para las mujeres en el primer quintil de riqueza y para las mujeres en el quinto quintil. Se obtendrían dos estimaciones DD del efecto de la intervención: una para las mujeres más pobres y otra para las mujeres más ricas. Luego, es suficiente realizar una prueba de heterogeneidad (como se describió anteriormente) para evaluar si el efecto de la intervención sobre el acceso a la atención obstétrica es heterogéneo entre los dos grupos³. Si por un

3. Esta prueba puede realizarse en varios grupos también, siempre que se adapte la prueba Q de Cochran.

lado, este efecto es heterogéneo entre los grupos, y, por otro lado, es más importante dentro del primer quintil de riqueza, entonces será posible concluir que la distribución de los beneficios de esta intervención es equitativa.

Análisis de interacción dual o diferencias en diferencias en diferencias (DDD)

Un segundo enfoque sería evaluar una doble interacción. Como se mencionó anteriormente, la estimación del efecto de una intervención a través de un análisis de tipo DD se realiza a través de un término de interacción. Aquí, evaluaríamos la presencia de una doble interacción del término DD con un indicador del quintil de riqueza. En otras palabras, se trata de evaluar el término DD varía según el nivel de riqueza. Aquí, la equidad será evaluada por la importancia de esta interacción y su dirección. En nuestro ejemplo, dado que el primer quintil representa el nivel más bajo de riqueza, este término de interacción debe ser negativo para que la distribución de los beneficios de esta intervención se considere equitativa. Este enfoque puede ser similar a un análisis de Diferencias en las diferencias en las diferencias (DDD). Se pueden encontrar ejemplos de este enfoque en la literatura de salud pública, con un enfoque individual (Harper, Strumpf et al. 2014) o un enfoque mixto con análisis de series temporales (Benmarhnia, Bailey Z et al. 2016).

Análisis con métodos de descomposición

En la sección anterior, se presentaron varias medidas de desigualdad, que van desde enfoques simples que comparan dos grupos, hasta enfoques que comparan varios grupos (ordenados o no), como el índice de concentración, por ejemplo.

Los métodos de descomposición se desarrollaron en el campo de la econometría y se utilizaron inicialmente para explicar, por ejemplo, las diferencias en los ingresos por sexo (por ejemplo, por edad, sector de actividad, nivel de experiencia). Estos métodos se pueden aplicar a la comparación entre dos grupos donde se puede usar el llamado método

Oaxaca Blinder (Oaxaca 1973; O'Donnell y Wagstaff 2008). Estos métodos también se pueden aplicar a medidas de desigualdad más avanzadas, como el índice de concentración (O'Donnell y Wagstaff 2008) como se discutió anteriormente.

Sin embargo, su aplicación en la evaluación de efectos y mediante métodos cuasi-experimentales también sigue siendo poco frecuente. Aquí simplemente describiremos cómo aplicar estos métodos de descomposición a un análisis de tipo DD.

Los métodos de descomposición consisten en modelar lo que sería la reducción de las desigualdades entre dos o más grupos si las exposiciones se llevaran a niveles equivalentes. Por lo tanto, en nuestro contexto, será suficiente realizar análisis de descomposición entre dos o más grupos de riqueza, por ejemplo, y analizar en particular el término de interacción DD. La interpretación de este término al final del análisis de descomposición indicará si los beneficios de la intervención contribuyen a reducir las diferencias entre los grupos de riqueza y la magnitud y dirección de esta reducción, lo que permitirá evaluar si la distribución de los beneficios de esta intervención puede considerarse como equitativa.

Conclusión

En este capítulo, hemos presentado el concepto de equidad en salud pública, así como varios métodos para medir las desigualdades en salud y diferentes enfoques para evaluar la heterogeneidad de los efectos de las políticas públicas. Es importante tener disponibles varios enfoques metodológicos que correspondan a diferentes situaciones de investigación, en los que, por ejemplo, uno puede estar interesado en la comparación entre dos grupos o en la comprensión de un gradiente social frente a un indicador de salud. Cada uno de los métodos propuestos puede ser parte de diferentes enfoques de salud pública (Benach, Malmusi et al. 2011; Benach, Malmusi et al. 2013) en cuanto a al modo de abordar una intervención de salud pública, ya sea a través de enfoques basados en la población, o de enfoques de universalismo proporcional.

La evaluación de la equidad en salud pública es importante para garantizar, por un lado que las intervenciones no contribuyan a aumentar las desigualdades en salud cuando se trata de enfoques basados en la población,

y, por otro lado, que cumplan los objetivos establecidos en el contexto de las intervenciones destinadas a reducir las desigualdades en salud, en particular permitiendo que los beneficios sean mayores dentro de los subgrupos o territorios que se identifican como los más vulnerables o cuyos necesidades sean más importantes.

Naturalmente, esta evaluación es solo un paso, entre otros, para comprender la distribución de los efectos de las políticas públicas y las desigualdades en salud en general. Antes de cualquier evaluación de equidad, es necesario tener una buena comprensión del contexto de la intervención para formular claramente hipótesis desde el punto de vista de la heterogeneidad de los efectos de las políticas públicas. Posteriormente, es importante también estudiar los mecanismos que entran en juego en la variabilidad de los beneficios de las intervenciones de salud pública, para mantener (?), modificar las políticas existentes y formular experimentos destinados a reducir las desigualdades. de salud.

Referencias clave

O'Donnell, O. A. et A. Wagstaff (2008). *Analyzing health equity using household survey data: a guide to techniques and their implementation*, World Bank Publications.

Este libro de libre acceso (<https://openknowledge.worldbank.org/handle/10986/6896>) es una excelente manera de descubrir los diferentes enfoques para medir las desigualdades en materia de salud, así como los métodos para comprender los determinantes de esas desigualdades y sus cambios a lo largo del tiempo, en particular en relación con una intervención específica.

Harper, S., N. B. King, S. C. Meersman, M. E. Reichman, N. Breen and J. Lynch (2010). Implicit value judgments in the measurement of health inequalities. *Milbank Quarterly* 88(1): 4-29.

El presente artículo es una referencia obligada en la literatura sobre las desigualdades en materia de salud en la medida en que, a través de varios ejemplos, pone de relieve la importancia de la elección del método utilizado para medir las desigualdades en las conclusiones. También detalla los valores implícitos que pueden guiar la elección de ciertos métodos sobre otros.

Benach, J., D. Malmusi, Y. Yasui, J. M. Martínez and C. Muntaner (2011). "Beyond Rose's strategies: a typology of scenarios of policy impact on population health and health inequalities." *International Journal of Health Services* 41(1): 1-9.

En este artículo se propone resumir los diferentes enfoques para estructurar las intervenciones de salud pública, con una interesante perspectiva histórica y una evaluación sistemática de los impactos de cada uno de estos enfoques en las desigualdades en materia de salud.

Referencias

Althabe, F., Bergel, E., Cafferata, M. L., Gibbons, L., Ciapponi, A., Alemán, A., ... et Palacios, A. R. (2008). Strategies for improving the quality of health care in maternal and child health in low-and middle-income countries: an overview of systematic reviews. *Paediatric and perinatal epidemiology*, 22(Suppl. 1), 42-60.

Benach, J., Malmusi, D., Yasui, Y. et Martínez, J. M. (2013). A new typology of policies to tackle health inequalities and scenarios of impact based on Rose's population approach. *Journal of epidemiology and community health*, 67(3), 286-291.

Benach, J., Malmusi, D., Yasui, Y., Martínez, J. M. et Muntaner, C. (2011). Beyond Rose's strategies: a typology of scenarios of policy impact on population health and health inequalities. *International Journal of Health Services*, 41(1), 1-9.

Benmarhnia, T., Bailey, Z., Auger, N., Kaiser, D., King, N. et K. JS. (2016). A difference-in-differences approach to assess the effect of a heat action plan on mortality and equity in Montreal, Quebec. *Environmental Health Perspectives*, 124(11), 1694-1699.

Benmarhnia, T., Grenier, P., Brand, A., Fournier, M., Deguen, S. et Smargiassi, A. (2015). Quantifying vulnerability to extreme heat in time series analyses: a novel approach applied to neighborhood social disparities under climate change. *International Journal of Environmental Research and Public Health*, 12(9), 11869-11879.

- Benmarhnia, T., Rey, L., Cartier, Y., Clary, C. M., Deguen, S. et Brousselle, A. (2014). Addressing equity in interventions to reduce air pollution in urban areas: a systematic review. *International journal of public health*, 59(6), 933-944.
- Blas, E. et Kurup, A. S. (2010). *Equity, social determinants and public health programmes*. Geneva: World Health Organization.
- Braveman, P. et Gruskin, S. (2003). Defining equity in health. *Journal of epidemiology and community health*, 57(4), 254-258.
- Braveman, P. A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K. S., Metzler, M. et Posner, S. (2005). Socioeconomic status in health research: one size does not fit all. *Jama*, 294(22), 2879-2888.
- Brown, T., Platt, S. et Amos, A. (2014). Equity impact of interventions and policies to reduce smoking in youth: systematic review. *Tobacco control*, 23(e2), e98-e105.
- Brown, T., Platt, S. et Amos, A. (2014). Equity impact of population-level interventions and policies to reduce smoking in adults: a systematic review. *Drug and alcohol dependence*, 138, 7-16.
- Cartier, Y., Benmarhnia, T. et Brousselle, A. (2015). Tool for assessing health and equity impacts of interventions modifying air quality in urban environments. *Evaluation and program planning*, 53, 1-9.
- Daniels, N. (2006). Equity and population health: toward a broader bioethics agenda. *Hastings Center Report*, 36(4), 22-35.
- Dimick, J. B. et Ryan, A. M. (2014). Methods for evaluating changes in health care policy: the difference-in-differences approach. *Jama*, 312(22), 2401-2402.
- Dzakpasu, S., Powell-Jackson, T. et Campbell, O. M. (2013). Impact of user fees on maternal health service utilization and related health outcomes: a systematic review. *Health policy and planning*, 29(2), 137-150.
- Ellis, R. P. (1987). The revenue generating potential of user fees in Kenyan government health facilities. *Social science et medicine*, 25(9), 995-1002.
- Farrelly, M. C., Nonnemaker, J. M., Chou, R., Hyland, A., Peterson, K. et Bauer, U. E. (2005). Changes in hospitality workers' exposure to secondhand smoke following the implementation of New York's smoke-free law. *Tobacco Control*, 14(4), 236-241.

- Filmer, D. et Pritchett, L. (1999). The effect of household wealth on educational attainment: evidence from 35 countries. *Population and development review*, 25(1), 85-120.
- Frazer, K., Callinan, J. E., McHugh, J., van Baarsel, S., Clarke, A., Doherty, K. et Kelleher, C. (2016). Legislative smoking bans for reducing harms from secondhand smoke exposure, smoking prevalence and tobacco consumption. *Cochrane Database of Systematic Reviews*, (2), p.-p.
- Gilman, S. E., Rende, R., Boergers, J., Abrams, D. B., Buka, S. L., Clark, M. A., ... et Lipsitt, L. P. (2009). Parental smoking and adolescent smoking initiation: an intergenerational perspective on tobacco control. *Pediatrics*, 123(2), e274-e281.
- Grabich, S. C., Robinson, W. R., Engel, S. M., Konrad, C. E., Richardson, D. B. et Horney, J. A. (2015). County-level hurricane exposure and birth rates: application of difference-in-differences analysis for confounding control. *Emerging themes in epidemiology*, 12(1), 1.
- Hajizadeh, M., Campbell, M. K. et Sarma, S. (2014). Socioeconomic inequalities in adult obesity risk in Canada: trends and decomposition analyses. *The European Journal of Health Economics*, 15(2), 203-221.
- Harper, S., King, N. B., Meersman, S. C., Reichman, M. E., Breen, N. et Lynch, J. (2010). Implicit value judgments in the measurement of health inequalities. *The Milbank Quarterly*, 88(1), 4-29.
- Harper, S., Lynch, J., Meersman, S. C., Breen, N., Davis, W. W. et Reichman, M. E. (2008). An overview of methods for monitoring social disparities in cancer with an example using trends in lung cancer incidence by area-socioeconomic position and race-ethnicity, 1992-2004. *American journal of epidemiology*, 167(8), 889-899.
- Harper, S., Strumpf, E. C., Burris, S., Smith, G. D. et Lynch, J. (2014). The effect of mandatory seat belt laws on seat belt use by socioeconomic position. *Journal of policy analysis and management*, 33(1), 141-161.
- Hoffman, S. J. et Tan, C. (2015). Overview of systematic reviews on the health-related effects of government tobacco control policies. *BMC public health*, 15(1), 744-754.

- Hosseinpoor, A. R., Van Doorslaer, E., Speybroeck, N., Naghavi, M., Mohammad, K., Majdzadeh, R., ... et Vega, J. (2006). Decomposing socioeconomic inequality in infant mortality in Iran. *International journal of epidemiology*, 35(5), 1211-1219.
- Houweling, T. A., Kunst, A. E., Huisman, M. et Mackenbach, J. P. (2007). Using relative and absolute measures for monitoring health inequalities: experiences from cross-national analyses on maternal and child health. *International journal for equity in health*, 6(1), 1.
- Howe, L. D., Hargreaves, J. R., Gabrysch, S. et Huttly, S. R. (2009). Is the wealth index a proxy for consumption expenditure? A systematic review. *Journal of epidemiology and community health*, 63(11), 871-877.
- Kakwani, N., Wagstaff, A. et Van Doorslaer, E. (1997). Socioeconomic inequalities in health: measurement, computation, and statistical inference. *Journal of econometrics*, 77(1), 87-103.
- Kaufman, J. S. et MacLehose, R. F. (2013). Which of these things is not like the others?. *Cancer*, 119(24), 4216-4222.
- Kawachi, I. et Kennedy, B. P. (1997). The relationship of income inequality to mortality: does the choice of indicator matter?. *Social science et medicine*, 45(7), 1121-1127.
- Kawachi, I., Subramanian, S. et Almeida-Filho, N. (2002). A glossary for health inequalities. *Journal of epidemiology and community health*, 56(9), 647-652.
- Lagarde, M. et Palmer, N. (2008). The impact of user fees on health service utilization in low-and middle-income countries: how strong is the evidence?. *Bulletin of the World Health Organization*, 86(11), 839-848C.
- Litvack, J. I. et Bodart, C. (1993). User fees plus quality equals improved access to health care: results of a field experiment in Cameroon. *Social science et medicine*, 37(3), 369-383.
- Lynch, J., Smith, G. D., Harper, S. et Bainbridge, K. (2006). Explaining the social gradient in coronary heart disease: comparing relative and absolute risk approaches. *Journal of epidemiology and community health*, 60(5), 436-441.

- Mackenbach, J. P. et Kunst, A. E. (1997). Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Social science et medicine*, 44(6), 757-771.
- Marmot, M., Friel, S., Bell, R., Houweling, T. A., Taylor, S. et Commission on Social Determinants of Health. (2008). Closing the gap in a generation: health equity through action on the social determinants of health. *The Lancet*, 372(9650), 1661-1669.
- McKinnon, B., Harper, S. et Kaufman, J. S. (2015). Who benefits from removing user fees for facility-based delivery services? Evidence on socioeconomic differences from Ghana, Senegal and Sierra Leone. *Social Science et Medicine*, 135, 117-123.
- McKinnon, B., Harper, S., Kaufman, J. S. et Bergevin, Y. (2014). Removing user fees for facility-based delivery services: a difference-in-differences evaluation from ten sub-Saharan African countries. *Health policy and planning*, 30(4), 432-441.
- O'Donnell, O. A. et Wagstaff, A. (2008). *Analyzing health equity using household survey data: a guide to techniques and their implementation*. Washington D.C.: World Bank Publications.
- Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International economic review*, 693-709.
- Pamuk, E. R. (1985). Social class inequality in mortality from 1921 to 1972 in England and Wales. *Population studies*, 39(1), 17-31.
- Pearcy, J. N. et Keppel, K. G. (2002). A summary measure of health disparity. *Public health reports*, 117(3), 273-280.
- Regidor, E. (2004). Measures of health inequalities: part 2. *Journal of epidemiology and community health*, 58(11), 900-903.
- Ridde, V. et Morestin, F. (2011). A scoping review of the literature on the abolition of user fees in health care services in Africa. *Health policy and planning*, 26(1), 1-11.
- Russell, S. et Gilson, L. (1997). User fee policies to promote health service access for the poor: a wolf in sheep's clothing?. *International Journal of Health Services*, 27(2), 359-379.

- Say, L. et Raine, R. (2007). A systematic review of inequalities in the use of maternal health care in developing countries: examining the scale of the problem and the importance of context. *Bulletin of the World Health Organization*, 85(10), 812-819.
- Shadish, W. R., Luellen, J. K. et Clark, M. (2006). Propensity scores and quasi-experiments: A testimony to the practical side of lee sechrest.
- Smith, M. J. (2015). Health equity in public health: Clarifying our commitment. *Public Health Ethics*, 8(2), 173-184.
- Starfield, B. (2007). Pathways of influence on equity in health. *Social Science et Medicine*, 64(7), 1355-1362.
- Theil, H. (1967). *Economics and information theory*.
- Wagstaff, A., Van Doorslaer, E. et Paci, P. (1991). On the measurement of horizontal inequity in the delivery of health care. *Journal of health economics*, 10(2), 169-205.
- Wagstaff, A., Van Doorslaer, E. et Watanabe, N. (2003). On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam. *Journal of econometrics*, 112(1), 207-223.
- Whitehead, M. (1992). The concepts and principles of equity and health. *International journal of health services*, 22(3), 429-445.
- Yates, R. (2009). Universal health care and the removal of user fees. *The Lancet*, 373(9680), 2078-2081.
- Yoder, R. A. (1989). Are people willing and able to pay for health services? *Social science et medicine*, 29(1), 35-42.

Resumen / Résumé / Abstract

La equidad en salud es ampliamente identificada como un objetivo al que las políticas y actividades de salud pública deben responder como una prioridad, pero cuya definición varía considerablemente. En este capítulo proponemos una definición de equidad en salud que se refiere a la distribución de los beneficios de las políticas públicas. En este capítulo se presentan diferentes métodos para medir las desigualdades en salud, su aplicación a la evaluación de los efectos de las intervenciones a través del ejemplo del efecto de la exención de las tasas de atención sanitaria sobre

el acceso a la atención obstétrica en un ejemplo simulado. Estas medidas incluyen el índice de concentración y el índice de Theil. Finalmente, proponemos diferentes enfoques para aplicarlos a la evaluación de los efectos de las intervenciones y, por lo tanto, a la equidad, incluyendo un enfoque simple y una base sobre el uso de métodos cuasi-experimentales. La evaluación de la equidad en salud pública es importante para garantizar que, por una parte, las intervenciones no contribuyan a aumentar las desigualdades en salud y, por otra parte, que cumplan los objetivos establecidos en las intervenciones para reducir las desigualdades en salud.

L'équité en santé est largement identifiée comme un objectif auquel les politiques et activités de santé publique doivent répondre en priorité, mais dont la définition varie considérablement. Dans ce chapitre, nous proposons une définition de l'équité en santé qui fait référence à la distribution des bénéfices des politiques publiques. Dans ce chapitre nous présentons différentes méthodes de mesures des inégalités de santé, leur application à l'évaluation des effets des interventions à travers l'exemple de l'effet de l'exemption du paiement des frais de soins sur l'accès aux soins obstétricaux dans le cadre d'un exemple simulé. Ces mesures incluent notamment l'indice de concentration et l'indice de Theil. Enfin, nous proposons différentes approches pour les appliquer à l'évaluation des effets des interventions donc à l'équité incluant une approche simple et une base sur l'utilisation de méthodes quasi-expérimentales. L'évaluation de l'équité en santé publique est importante pour s'assurer que, d'une part, des interventions ne contribuent pas à augmenter les inégalités de santé et qu'elles répondent aux objectifs énoncés dans le cadre d'interventions visant à réduire les inégalités de santé.

Health equity is considered a priority for public health interventions and policies, but the concept of an equitable intervention varies in definition. We define an equitable intervention as one that has heterogeneity in its effects, and provides greater benefits for subgroups or territories of a population that have been defined as vulnerable. In this chapter, we assess the impact of an intervention establishing free obstetric care (childbirth costs excluded) on increasing equity and access to care. We notably present the concentration index and the Theil Index. Simulated data from population surveys is used

to discuss important issues regarding measurement of inequalities and different indicators that can be used in evaluating the equitability of an intervention. Lastly, two evaluation strategies are discussed: a simple evaluation approach to illustrate the application of inequality measurement methods and an approach based on quasi-experimental designs. These tools to assess public health equity is important to ensure that interventions meet objectives to reduce inequalities and prevent them from contributing to increasing health disparities.

Tarik Benmarhnia es Profesor Adjunto de Epidemiología de la Universidad de California en San Diego en el Instituto Scripps de Oceanografía de la Facultad de Medicina. Completó su doctorado en Epidemiología en la Universidad de Montreal y en la Universidad Paris Sud. También es investigador asociado en Chile y Francia y miembro del consejo editorial de *Environmental Health Perspectives*. Sus investigaciones se centran en los efectos de los fenómenos extremos en la salud, con especial atención a las poblaciones vulnerables y las consecuencias para las políticas públicas. También desarrolla enfoques metodológicos para evaluar el efecto de las políticas ambientales en la salud, como las medidas de adaptación al cambio climático y las medidas de control de la contaminación del aire.

Britt McKinnon es investigador del Centro de Salud Mundial del Hospital para Niños Enfermos y Profesor Adjunto de la Escuela de Salud Pública Dalla Lana de la Universidad de Toronto. Recibió su doctorado en Epidemiología de la Universidad McGill e hizo una beca de postdoctorado en el Instituto McGill de Salud y Política Social. Sus investigaciones en salud global se centran en la salud materna e infantil, la epidemiología social y la evaluación de políticas y programas de salud.

Cita

Tarik Benmarhnia y Britt McKinnon (2020). Medición de equidad. Un ejemplo de intervención de gratuidad de atención obstétrica. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 297-323. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

12. Análisis de coste-efectividad

Una intervención para descentralizar la atención del VIH/SIDA en Shiselweni (Swazilandia)

GUILLAUME JOUQUET

Análisis de coste-efectividad

Definición del método

Estudio retrospectivo de la efectividad en función de los costes desde el punto de vista del proveedor de servicios, con el cálculo del coeficiente de coste-efectividad incremental (ICER)

Fortalezas del método

- compara dos alternativas integrando los aspectos cualitativos (efectividad) y económicos (coste) de los programas de salud
- se basa en datos reales de un programa a gran escala

Desafíos del método

- el necesario acceso a los datos financieros de los proveedores de servicios puede ser un desafío
- medir el aumento de la eficiencia relacionado con el enfoque alternativo aislando los posibles factores de confusión.

Siempre que me piden que realice una evaluación económica, mi primera pregunta es: ¿para quién y para qué se va a utilizar? ¿Es para justificar los gastos ante terceros, para ayudar a tomar una decisión sobre la asignación de recursos, para simular gastos futuros o para fines de promoción? La respuesta a esta pregunta ayudará a determinar qué tipo de evaluación económica será la más apropiada. Hay tres tipos principales:

- Evaluación de los costes totales, lo que supone calcular el costo de todos los componentes o insumos del programa (por ejemplo, recursos humanos y suministros médicos).
- La evaluación de la productividad, que se basa en las “cantidades producidas”. El costo total se dividirá por los resultados del programa (por ejemplo, el número de pacientes tratados).

- Estudio de coste-efectividad, que introduce un concepto cualitativo dividiendo el costo total por una medida del resultado del programa (por ejemplo, el número de pacientes curados).

La evaluación puede realizarse desde diferentes perspectivas:

- el punto de vista del proveedor de servicios (por ejemplo, una ONG y/o el Ministerio de Salud que sufraga el costo del programa)
- el punto de vista del beneficiario (por ejemplo, el paciente que debe pagar los gastos de transporte para llegar a un centro de atención médica).

El enfoque más completo es considerar ambas perspectivas siempre que sea posible.

Los dos métodos principales que podemos utilizar, ya sea de manera exclusiva o complementaria, para evaluar los costos son el método del costo total y el método de los ingredientes.

- El método del costo total es un método descendente (Conteh 2004): partimos de todos los gastos registrados en las cuentas, que categorizaremos como costos directos e indirectos. Los costos directos se asignarán directamente a los centros de costos o actividades, mientras que los costos indirectos se asignarán de acuerdo con un criterio de asignación.
- El método de los ingredientes es un método ascendente (WHO-CHOICE 2016): para cada insumo utilizado en una actividad determinada se determina la cantidad necesaria y el costo unitario. El costo de la actividad se obtiene entonces sumando los productos del costo unitario y la cantidad requerida de cada insumo.

La elección del tipo de evaluación, de las opiniones y de los métodos dependerá no sólo de los objetivos del estudio, sino también de una doble restricción (Ridde y Dagenais 2012, p172):

- el tiempo y los recursos asignados a la evaluación
- la dificultad de acceder a datos fiables y utilizables.

En este caso, el cliente de la evaluación era Médicos Sin Fronteras (MSF), una organización no gubernamental (ONG) médica internacional. Esta ONG ha establecido un programa de tratamiento del síndrome de inmunodeficiencia adquirida (SIDA) en colaboración con el Ministerio de Salud Pública (MSP) de Swazilandia. Este programa tiene la particularidad de estar descentralizado e integrado a nivel de los Centros de Salud Primaria (CPS). Se trata de un enfoque innovador en Swazilandia, donde el tratamiento del SIDA está centralizado en los centros de salud secundarios. Los enfoques descentralizados del tratamiento del SIDA en los países en desarrollo han sido objeto de varios estudios que demuestran su efectividad (Bedelu y otros, 2007; Cohen y otros, 2009; Loubiere y otros, 2009; Bemelmans y otros, 2010; Chan y otros, 2010). El propósito de esta evaluación es añadir un componente económico. Para comparar una intervención médica con otra que tenga el mismo objetivo, el tipo de evaluación económica más apropiado es un estudio de rentabilidad con cálculo de la relación coste-efectividad incremental (ICER) (Drummond 2005, pág. 39). Esta proporción indicará el costo incremental de ganar un año de vida adicional en el programa más efectivo en relación con el programa menos efectivo. En la cuarta parte de este capítulo se explica cómo se calcula e interpreta esta relación.

Un enfoque médico descentralizado facilita el acceso de los pacientes a la atención, pero dificulta la labor del proveedor de servicios (Mulamba et al. 2010). Por lo tanto, para comprender esta evaluación en su totalidad, debe realizarse desde la doble perspectiva del paciente y del proveedor de servicios. Sin embargo, debido a las limitaciones de recursos y tiempo de nuestro cliente, hemos preferido limitarnos al punto de vista de los proveedores de servicios, en este caso MSF y MSP.

El cálculo del ICER implica estimar el coste y la efectividad de cada enfoque. Para estimar los costos económicos y financieros, optamos por el enfoque de los ingredientes, que tiene ciertas ventajas, pero también ciertas limitaciones, que analizaremos con más detalle en la quinta parte de este capítulo. Para la medición de la efectividad, elegimos el mantenimiento de la terapia antirretroviral (ARV) a los 12 meses como el efecto esperado del enfoque. La justificación de esta elección se detalla en la cuarta parte de este capítulo (medición de la efectividad).

Las principales dificultades para una evaluación como ésta son:

- acceso a los datos financieros de los proveedores de servicios

- medir el aumento de la eficiencia relacionado con el enfoque alternativo aislando los posibles factores de confusión.

Contexto y programa evaluados

Swazilandia

Swazilandia, con una población de alrededor de 1 millón de habitantes en 17.000 km², es un pequeño país montañoso sin litoral entre Sudáfrica y Mozambique. Es una de las tres monarquías que quedan en el continente africano y la única en la que el rey todavía tiene el poder ejecutivo. En el decenio de 1980 y principios del de 1990, el reino experimentó un desarrollo económico que benefició de manera desigual a la población. De hecho, Swazilandia es el séptimo país más desigual del mundo y se estima que el 81% de la población vive con menos de 2 dólares al día. No obstante, este desarrollo económico permitió que la situación sanitaria de la población swazi mejorara considerablemente hasta principios del decenio de 1990 (PNUD 2009). Desafortunadamente, la llegada del SIDA ha cambiado completamente la situación. El país tiene la mayor prevalencia mundial de VIH (virus de la inmunodeficiencia humana) entre los adultos de 15 a 49 años de edad: el 31% de ellos están infectados y corren el riesgo de morir en los próximos años si no reciben tratamiento antirretroviral (ARV). Swazilandia también tiene una de las tasas de incidencia de tuberculosis más altas (OMS 2009). La epidemia ha contribuido de manera significativa a reducir a la mitad la esperanza de vida en los dos últimos decenios, pasando de 60 años en el decenio de 1990 a 31 años en 2007 (ONUSIDA 2010).

Región de Shiselweni

En 2007, tras una evaluación de cuatro meses en todo el país, MSF decidió poner en marcha un programa de prevención y tratamiento del SIDA y la tuberculosis en la región de Shiselweni en asociación con MSP. En noviembre de 2007 se firmó un memorando de entendimiento. La región de Shiselweni fue elegida por MSF porque se considera la más pobre y remota. Se caracteriza, entre otras cosas, por una infraestructura sanitaria

inadecuada y unos resultados de salud pública deficientes. Es una región montañosa y rural dividida en tres zonas sanitarias: Nhlanguano, Hlatikulu y Malangeni.

Cada zona de salud tiene una estructura central de atención secundaria de la salud (CSS) y entre siete y ocho centros de salud de atención primaria (CSP) en la periferia (22 en total en la región). Los principales problemas estructurales identificados con respecto a la atención del VIH/TB son:

- Escasez crónica de personal
- La centralización y la falta de integración de los servicios para el VIH y la tuberculosis
- Capacidad insuficiente para proporcionar atención al VIH/TB en los CSP.

A esto se suma el paisaje montañoso, las poblaciones dispersas y los altos costos de transporte. Todo esto hizo que el cuidado del VIH/TB fuera inaccesible para la mayoría de los suazis que vivían en las zonas rurales de la región.

El programa de MSF

En este contexto de recursos limitados y difícil acceso a la atención, el objetivo del programa establecido por MSF es “apoyar al Ministerio de Salud en la ampliación de los servicios para llenar el vacío en la atención de la salud de las personas que viven con el VIH (PVVIH) en la región de Shiselweni”.

Los resultados esperados a corto, medio y largo plazo son:

- para detectar a las PVVIH que necesitan tratamiento
- reducir las barreras geográficas para el acceso a la atención
- mejorar la retención de los pacientes en la terapia antirretroviral
- aumentan la vida de los pacientes y disminuyen el riesgo de transmisión del VIH.

Para lograr estos resultados, la estrategia de MSF ha sido promover la descentralización e integración de la atención sanitaria del VIH a nivel de las CSP cuya organización ha tenido que ser revisada. Un primer desafío fue la falta de recursos humanos para la salud para hacer frente a la carga de

trabajo adicional de cuidar a las PVVIH. Por lo tanto, MSF asignó personal sanitario además del personal de MSF ya presente. Una segunda dificultad fue la ausencia de médicos a nivel de CSP. Por lo tanto, MSF asignó un médico móvil y organizó la transferencia a las enfermeras de algunas de las tareas que normalmente realizan los médicos. Para que el personal de enfermería pudiera hacer frente a esta nueva carga de trabajo, MSF organizó la contratación y la formación de pacientes expertos a los que se les asignaron determinadas tareas que normalmente realizan los enfermeros, como el apoyo psicológico y la supervisión del cumplimiento del tratamiento del paciente. Una tercera dificultad estaba relacionada con los seguimientos de laboratorio necesarios para los pacientes que recibían ARV. Por lo tanto, se crearon mini-laboratorios en cada CSP, equipados con dispositivos fáciles de usar para realizar las pruebas necesarias. MSF también reclutó y capacitó a personal comunitario no calificado (“flebotomistas”) para realizar estas pruebas en los mini-laboratorios. Se estableció un equipo móvil de supervisión que incluía, además del médico mencionado anteriormente, un director de enfermería, un director de laboratorio y un director experto en clientes. Por último, se ha organizado la logística para que todo funcione, es decir, el transporte del equipo móvil de supervisión, pero también el transporte de los insumos médicos y de laboratorio, así como de las muestras de sangre para ciertas pruebas que todavía se realizan en el laboratorio central.

Plan, programa y protocolo de evaluación

Objetivo

En 2012, cinco años después del inicio del programa Shiselweni, MSF quiso analizar esta experiencia para extraer lecciones de ella y compartirlas dentro del movimiento de MSF, así como con las autoridades nacionales y la comunidad internacional. La evaluación económica aquí descrita forma parte de una evaluación más amplia realizada por cuatro evaluadores independientes y encargada por MSF con el doble propósito de documentar las experiencias y la promoción (Turashvili y otros 2014).

Preguntas de evaluación

La cuestión de la evaluación económica era si el enfoque descentralizado del tratamiento del SIDA aplicado por MSF en la región de Shiselweni era eficaz en función de los costos en comparación con el enfoque centralizado utilizado en otros lugares de Swazilandia. La justificación de esta cuestión era que se esperaba que la descentralización de los servicios al nivel de CSP resultara tanto en mayores costos para el proveedor de servicios como en menores costos y mejores resultados sanitarios para los beneficiarios.

Para responder a esta pregunta, habría sido ideal llevar a cabo:

- una evaluación de la efectividad en función de los costos
- desde el punto de vista del proveedor de servicios y de los beneficiarios
- comparar una población de prueba y control que sea similar en todos los aspectos, excepto en el enfoque de atención del VIH adoptado (centralizado o descentralizado)
- el cálculo de la relación costo-efectividad incremental (ICER) de un enfoque sobre el otro

Teniendo en cuenta las limitaciones de recursos, tiempo y disponibilidad de información, tuvimos que revisar a la baja este esquema ideal que finalmente se limitó a:

- una evaluación de la efectividad en función de los costos basada en un estudio retrospectivo de los datos reunidos sistemáticamente por la MSP y la MSF en la zona de salud de Nhlangano (limitación geográfica)
- desde el punto de vista del proveedor de servicios (únicamente)
- comparando los pacientes con ARV monitoreados en los CSP con los pacientes con ARV monitoreados en la zona de salud CSS en 2012
- calculando la relación costo-efectividad incremental (ICER) de los pacientes de CSP a los de HSC.

El estudio ha sido autorizado por el Departamento de Salud. Dado que el análisis se basó en datos clínicos de rutina, no se solicitó la aprobación del comité de ética ni el consentimiento de los pacientes. Toda la información

relacionada con los pacientes se trató de forma anónima utilizando números de identificación codificados. No se utilizó información que pudiera revelar la identidad de los pacientes.

Presupuesto y planificación

Esta evaluación económica se llevó a cabo entre octubre de 2012 y julio de 2013, en tres fases principales: definición del mandato, recopilación y análisis de datos in situ y redacción del informe de evaluación.

Los principales costos del estudio se relacionaron con los recursos humanos (un economista independiente para 40 días-persona y el epidemiólogo del programa de MSF en Shiselweni para 20 días-persona) y los gastos de viaje para las dos visitas in situ del economista (el epidemiólogo ya estaba en el lugar).

Aplicación del método

Una vez que la pregunta y la cita han sido definidas en asociación con el patrocinador de la evaluación, se pueden identificar cuatro pasos principales en el proceso que seguimos:

- Recogida de datos
- Estimación del costo
- Medición de la efectividad
- Análisis e interpretación de los resultados

Recogida de datos

La recogida de datos debería permitir calcular la relación costo-efectividad incremental (ICER) entre los dos enfoques.

El ICER se calcula de la siguiente manera: $ICER = CCSP - CCSS/ECSP - ECSS$. “C” es el coste anual de un paciente con terapia antirretroviral seguido en un centro de salud primario (CPS) o en un centro de salud secundario (CSS), “E” es la efectividad medida por la proporción de pacientes mantenidos con terapia antirretroviral a los 12 meses en el CPS o en el CSS.

Por consiguiente, tuvimos que recoger información que nos permitiera, para cada enfoque, estimar los costes y medir la efectividad.

Para ello, tuvimos que cotejar varias fuentes de datos cuantitativos (contabilidad, recursos humanos, vigilancia médica, bases de datos de vigilancia de suministros y vehículos) y cualitativos (informes de estudios, entrevistas) de MSF o MSP. La comparación de estas diferentes fuentes cuantitativas y cualitativas es lo que nos ha permitido tener una visión general satisfactoria. Por ello, la visita in situ fue inevitable: me permitió hacerme una idea del tipo de información disponible (o no) y en qué forma, y complementar con entrevistas los datos que faltaban, eran incoherentes o difíciles de entender desde el punto de vista cuantitativo.

Estimación del coste

El coste se estimó utilizando el método de los ingredientes, que tiene dos elementos clave: 1) la identificación de las cantidades de los ingredientes utilizados y 2) el establecimiento del costo unitario o el precio de cada ingrediente. Por consiguiente, el costo por paciente se estimó para cada uno de los dos enfoques de la siguiente manera:

$C = Q_1P_1 + \dots + Q_nP_n$, donde :

- C = costo por paciente
- $Q_1 \dots Q_n$ = cantidades de recursos
- $P_1 \dots P_n$ = precio (o costos unitarios) de cada uno de los recursos cuantificados.

En otras palabras, el coste por paciente se obtiene sumando el costo de cada uno de los ingredientes utilizados para proporcionar los cuidados necesarios. El costo de cada ingrediente se calcula multiplicando su precio unitario por la cantidad utilizada.

El costo por paciente debe expresarse en una unidad de tiempo común para poder comparar un paciente con otro en los CSS y CSP. Aquí hemos optado por estimar los costos en términos de paciente/año (un paciente que, por ejemplo, hubiera muerto tres meses después de iniciar el tratamiento no tiene ni el mismo costo ni la misma efectividad que un paciente que sigue vivo y en tratamiento un año después de iniciarlo).

Para calcular el ICER, no es necesario estimar los costos, que son idénticos de un enfoque a otro. Típicamente en nuestro caso, no había razón para que el costo de los ARVs en sí mismos difiriera entre un paciente que estaba siendo seguido en un CSS y un paciente que estaba siendo seguido en un CSP. Sin embargo, tomamos este costo en consideración en nuestro análisis para mostrar que representa la mayoría del costo total para un paciente en ARVs. Los otros costos que consideramos son:

- El costo de las consultas del personal sanitario
- El costo de la vigilancia psicológica y de la adherencia por parte de pacientes expertos
- El costo del equipo de supervisión
- El costo de las cuatro principales pruebas de laboratorio necesarias para los pacientes en ARV
- El costo del transporte del equipo de supervisión, los insumos médicos y la recolección de muestras de sangre para algunas pruebas todavía se realiza de forma centralizada.

También hemos optado por omitir ciertos costos que, aunque pueden variar de un enfoque a otro, son sin embargo insignificantes. Entre ellas se encuentran la depreciación de los edificios y el costo de los medicamentos para las infecciones oportunistas.

El costo de los ARV por paciente/año se obtuvo sobre la base del precio anual de los tratamientos y la proporción de pacientes bajo los diferentes regímenes de ARV posibles.

Para calcular el costo de las consultas, necesitábamos saber el número de consultas por paciente/año y el costo unitario de una consulta. El número de consultas por paciente/año se obtuvo de la base de datos de seguimiento de pacientes de ARV individuales de Swazilandia (APMR). Por el costo unitario de una consulta, sabíamos que sería diferente en un CSS con un servicio de ARV dedicado y en un CSP donde el cuidado del VIH está integrado con otros cuidados.

En el caso del centro de ARV del CSS, el cálculo fue sencillo, ya que este servicio prácticamente sólo recibe a pacientes con ARV (un pequeño porcentaje se refiere a pacientes con pretratamiento de ARV). Había que calcular el costo anual del personal que trabajaba en el servicio, había que

añadir un porcentaje para cubrir los gastos de funcionamiento, y el total debía dividirse por el número total de consultas realizadas por el servicio durante el año.

En el caso de las CSP, las consultas realizadas por la misma enfermera pueden registrarse en tres registros diferentes: consultas generales (81% del total), consultas de ARV (14% del total) y consultas prenatales y postnatales (5% del total). El análisis mensual de los datos informatizados de estos tres registros por parte de los CSP reveló algunas inconsistencias, algunas de las cuales podrían corregirse fácilmente. Por ejemplo, a veces faltaban datos de un mes, lo que era fácilmente visible y se explicaba después de la investigación por la ausencia de la persona que hacía las entradas. Por lo tanto, corregimos aproximando el mes que faltaba por el promedio del mes anterior y el mes posterior. Dos de los siete CSP analizados tenían tales inconsistencias que preferimos descartarlos del análisis. Nuestras investigaciones confirmaron que los codificadores de los dos CSP en cuestión no funcionaban satisfactoriamente. Por lo tanto, calculamos el costo promedio de una consulta sobre la base de los tres registros anuales de los cinco CPS seleccionados, es decir, un total de 91.183 consultas. El costo promedio de una consulta fue entonces estimado usando el mismo método que para el servicio de ARV del CSS. A continuación, consideramos que el tiempo dedicado por el personal sanitario a cada tipo de consulta se distribuía normalmente y que el costo medio de todas las consultas era un buen indicador del costo medio de las consultas de ARV.

Para estimar el costo por paciente/año de los pacientes expertos a cargo del apoyo psicológico y el monitoreo de la adherencia, la tarea fue más fácil, ya que este personal está dedicado al 100% a los pacientes que reciben ARV. Por lo tanto, bastaba con dividir su costo anual por el número de pacientes/años en cada tipo de centro de salud.

Consideramos que el costo de la supervisión del servicio de ARV en el CSS es insignificante, ya que es casi inexistente. Por otra parte, calculamos el costo del personal del equipo móvil dedicado a la atención del VIH en las CSP y compuesto por un médico generalista, un supervisor de enfermería, un supervisor experto de pacientes y un gerente de laboratorio. Dado que el área de supervisión de este equipo cubre más CSP que los de nuestra muestra, utilizamos una regla de tres para estimar el costo específicamente relacionado con ellos, que luego dividimos por el número de pacientes/años de nuestra muestra.

El costo unitario de cada prueba difiere necesariamente entre un paciente al que se le hace un seguimiento en un CSP y un paciente al que se le hace un seguimiento en un HSC, porque las tecnologías y las categorías de personal no son las mismas en un minilaboratorio o en el laboratorio central. Para estimar el costo de estas pruebas, nos basamos en la entrevista con el responsable de los laboratorios de MSF. Juntos compilamos una lista y las cantidades de reactivos y consumibles necesarios para cada prueba en cada tipo de laboratorio (CSP/CSS). Complementamos esta entrevista con datos del departamento de compras de MSF para establecer el precio unitario de cada reactivo y consumible. El costo del personal de laboratorio se obtuvo dividiendo el costo anual de ese personal por el número de pruebas anuales realizadas en cada entidad. Para estimar el número de pruebas realizadas por paciente y por año, se utilizó el protocolo actual, ya que los datos del NMPA no resultaron fiables.

El último tipo de coste que tuvimos en cuenta es el coste del transporte proporcionado por MSF para los equipos de supervisión, los insumos médicos y algunas muestras de sangre. Hemos calculado, gracias a la contabilidad de MSF y al seguimiento de los vehículos, un coste por kilómetro que incluye el conductor, la depreciación del vehículo, el combustible, el mantenimiento y las reparaciones. El archivo de seguimiento de vehículos de MSF también nos permitió estimar el número de kilómetros recorridos para servir a los CSP de nuestra muestra.

Todos los datos de costos se registraron en Excel y se analizaron mediante tablas dinámicas.

Medición de la efectividad

Para medir la efectividad de ambos enfoques, elegimos como indicador el mantenimiento de los pacientes en terapia antirretroviral a los 12 meses. La OMS (OMS 2010) considera que este es un buen indicador de la calidad general de la atención del VIH porque :

- refleja tanto la calidad de la atención médica y el buen seguimiento de los pacientes, como la accesibilidad de los servicios.
- los resultados negativos (muertes, pérdidas durante el seguimiento) se producen principalmente en el primer año después del inicio del tratamiento (OMS 2012).

La retención del tratamiento a los 12 meses se calculó utilizando el registro informático del programa ARV de Swazilandia (APMR). Se trata de una base de datos electrónica que contiene información básica y el seguimiento de todas las visitas de los pacientes que reciben ARV. Para todos los pacientes incluidos en la población del estudio, los resultados a los 12 meses del inicio del tratamiento se codificaron de la siguiente manera:

- Se mantiene en la terapia ARV
- Fallecido
- Transferido
- El tratamiento se detuvo
- Perdido

Los pacientes que fueron transferidos (desde o hacia otros programas) y los pacientes cuyo tratamiento fue detenido por decisión médica (fracaso del tratamiento no relacionado con el tipo de intervención) fueron excluidos de nuestro análisis. Se consideraba que los pacientes permanecían en la terapia siempre y cuando no hubieran muerto o se hubieran perdido en el seguimiento. Se consideraba que los pacientes se perdían en el seguimiento si no se registraba ninguna visita en el NOPRC en los 90 días siguientes a la última visita programada.

Por lo tanto, nuestro análisis se limitó a los pacientes que habían comenzado el tratamiento ARV en 2011 en la misma zona que para el estudio de costos (servicio ARV en el CSS de Nhlangano y cinco de los siete CSP en la zona de salud). Los pacientes que fueron transferidos al área de estudio en 2011 (102 personas) fueron excluidos del análisis, así como los que fueron transferidos del CSP a un CSS (13). También se hicieron algunas exclusiones debido a errores en los datos. El tamaño final de la muestra fue de 1.140 pacientes.

Se realizó un análisis descriptivo de las características de las poblaciones estudiadas. Se hicieron comparaciones estadísticas de los posibles factores de confusión (como el sexo, el recuento de CD4 y el grupo de edad) entre las poblaciones de pacientes de la SSC y la CSP utilizando la prueba de Chi² de Pearson para las variables categóricas y las pruebas de Student o Kruskal-Wallis para las variables cuantitativas (dependiendo de si su distribución es normal o no). Se realizó una regresión logística incondicional (utilizando un resultado binario: “mantenido en tratamiento a los 12 meses” versus “no

mantenido”) para calcular los Odds ratios (coeficientes de probabilidad) e intervalos de confianza del 95%. El análisis estadístico se realizó usando Stata/SE (StataCorp, Texas, USA, versión 12.1.).

Principales resultados

Nuestro estudio muestra que el cuidado de la salud con ARV en los CSP es tanto más efectivo (88% mantenido en terapia después de 12 meses de tratamiento en los CSP contra 76% en los HSC) como más caro (\$311 por paciente-año en los CSP contra \$233 en los CSS). El ICER es de 652 dólares, lo que representa el costo adicional para el proveedor de servicios para salvar un año más de vida en el enfoque de CSP en comparación con el enfoque de HSC. Por lo tanto, el enfoque CSP domina el enfoque CSS, ya que es más eficaz, pero ¿sigue siendo rentable?

La respuesta a esta pregunta es muy subjetiva, ya que depende del precio que uno está dispuesto y/o es capaz de pagar para salvar un año más de vida. La Organización Mundial de la Salud (OMS) ha propugnado desde hace mucho tiempo la comparación del ICER con el producto nacional bruto (PNB) per cápita para juzgar la efectividad en función de los costos de un programa (OMS, 2002). Sin embargo, desde 2015, esta recomendación, muy controvertida, se ha abandonado en favor de un análisis más detallado del contexto del programa (Marsella 2015).

En este estudio, MSF y el Ministerio de Salud consideraron que el costo de salvar un año de vida adicional era lo suficientemente bajo como para calificar el enfoque de CSP como rentable en comparación con el enfoque de CSS.

Análisis reflexivo y conclusión

El principal valor de esta evaluación económica retrospectiva es que se basa en datos reales de un programa a gran escala. Los resultados de esta evaluación económica, así como los resultados de los otros tres componentes de la evaluación del programa de Shiselweni, se discutieron en las siguientes secciones:

- un informe de estudio que fue ampliamente distribuido dentro de MSF
- una presentación oral en la 17ª Conferencia Internacional sobre el SIDA y las ETS en África (ICASA) en diciembre de 2013
- una presentación general en una conferencia organizada por MSF en Swazilandia en mayo de 2014 a la que fueron invitados representantes de MSP y de las ONG internacionales que participan en la lucha contra el SIDA en Swazilandia.

La principal desventaja de nuestro enfoque es que la evaluación no estaba prevista al inicio del programa. Esto significaba que teníamos que ocuparnos de los datos que se recogían habitualmente en el programa, datos que se recogían originalmente con otro fin y que, por lo tanto, no eran necesariamente pertinentes para los fines de nuestra evaluación económica, o que no satisfacían necesariamente nuestras necesidades en la forma en que se recogían.

Tuvimos que revisar el alcance del estudio a la baja de las previsiones originales. El objetivo inicial del patrocinador era medir la efectividad en función de los costos de un enfoque descentralizado frente a uno centralizado. Lo ideal habría sido comparar dos zonas sanitarias que hubieran adoptado cada una uno de los dos enfoques. Esto no era técnicamente posible. Por lo tanto, limitamos la evaluación económica a una comparación entre pacientes con ARVs seguidos en un CSS o un CSP. Por razones de viabilidad y limitaciones de tiempo, también tuvimos que limitar la evaluación a la perspectiva del proveedor del servicio, mientras que la perspectiva del beneficiario también habría sido interesante. Lo ideal hubiera sido entrevistar a un mínimo de 200 pacientes seleccionados al azar en todo el territorio para obtener datos fiables y estadísticamente significativos. Sin embargo, creemos que nuestra evaluación sigue siendo pertinente, ya que pudimos demostrar que el enfoque del CSP es eficaz en función de los costos a pesar de estas dos limitaciones, que en nuestra opinión minimizan la medición de su efectividad.

Para estimar el costo, seguimos el método de los ingredientes. La elección de los ingredientes que hemos seleccionado se basa en nuestra experiencia en la estimación de los costos utilizando el método de costo total en otros programas de lucha contra el SIDA en entornos similares. Son estas evaluaciones previas las que nos han permitido distinguir entre los factores de costo importantes que deben ser cuidadosamente estimados y

los menos importantes que pueden ser pasados por alto. Confiar únicamente en el método de los ingredientes sin este conocimiento previo puede llevar a un sesgo de estimación, a descuidar costos importantes pero no muy visibles (la punta del iceberg), o a perder un tiempo valioso estimando costos insignificantes (el árbol que esconde el bosque).

Tenga en cuenta que el acceso a la información financiera no siempre es fácil. No tuvimos dificultad en obtener esta información de MSF, que fue el patrocinador del estudio y que, por lo tanto, nos abrió todos sus libros de cuentas con total transparencia. El MSP, por otra parte, no nos dio acceso a sus datos contables. Como el principal costo generado por el MSP es el del personal sanitario, hicimos nuestras estimaciones basadas en la escala salarial de la administración pública de Swazilandia, que es de información pública.

Otra dificultad para estimar los costos es la abundancia de información y a veces las fuentes de información conflictivas. Los datos deben ser triangulados a través de entrevistas para separar el trigo de la paja. El uso de Excel y de las tablas dinámicas también nos permite adoptar un enfoque holístico de la información del director, acercándonos a los detalles cuando es necesario.

En cuanto a la medición de la efectividad, el indicador que hemos elegido es un indicador estándar para el seguimiento de los pacientes que reciben ARV. Por lo tanto, la APMR, una herramienta de vigilancia utilizada por el programa, está perfectamente adaptada a su cálculo. La dificultad aquí radica en otro aspecto: ¿cómo sabemos si la mayor efectividad de un enfoque sobre otro depende del enfoque en cuestión o de otros factores de confusión? Como hemos visto, hemos tenido en cuenta los principales factores de confusión identificados en los programas de lucha contra el SIDA, a saber, el sexo, la edad y el recuento de CD4 de los pacientes, y hemos ajustado nuestros resultados en consecuencia. Otro factor de confusión potencial fue el hecho de que el apoyo de MSF fue más intenso a nivel de CSP que a nivel de CSS. Investigamos más a fondo para determinar si este personal sanitario adicional en los CSP podría haber influido en su mayor efectividad, en cuyo caso habríamos medido no la efectividad del enfoque de los CSP, sino la del personal adicional de MSF. Nuestra conclusión es que los CPS tienen un exceso de personal y que una disminución del personal no llevaría a una disminución de la efectividad. Esto nos llevó a realizar análisis

de sensibilidad con la hipótesis de que los gastos de personal de los CPS serían menores, lo que mejoraría aún más la efectividad en función de los costos de este enfoque.

En conclusión, una de las principales enseñanzas extraídas de este estudio es que es preferible planificar una evaluación económica al comienzo de un programa a fin de reunir prospectivamente los datos adecuados para su aplicación. Sobre la base de las enseñanzas extraídas de la realización de esta evaluación, hemos adaptado nuestra metodología para la evaluación económica de una nueva iniciativa piloto del mismo programa de lucha contra el SIDA en la región de Shiselweni. La nueva iniciativa lanzada por MSF se llama “Acceso temprano a los ARV para todos”. El objetivo es poner a los pacientes en tratamiento ARV tan pronto como den positivo en el test de VIH. El protocolo actual en Swazilandia, como en la mayoría de los países del África meridional, consiste en esperar a que las defensas inmunitarias medidas por las células CD4 hayan caído por debajo de un determinado umbral antes de someter a los pacientes seropositivos a tratamiento. Una de las ventajas de la iniciación temprana de los medicamentos antirretrovirales es que es menos probable que se pase por alto a un paciente que los recibe que a un paciente seropositivo que no ha iniciado el tratamiento. Otra ventaja es que un paciente que recibe medicamentos antirretrovirales tiene menos probabilidades de infectar a otros que un paciente seropositivo sin tratamiento.

Para medir la efectividad en función de los costos de esta nueva iniciativa, utilizamos la misma metodología que se describe en este capítulo, pero esta vez, como pudimos diseñar el método de evaluación antes de que comenzara la iniciativa que se estaba evaluando, pudimos 1) ser más cuidadosos en la definición y selección de las poblaciones de intervención y control, y 2) hacer arreglos en la fase inicial para facilitar la recogida de datos y la medición de la efectividad.

Referencias clave

Conteh, L. (2004). Cost and unit cost calculations using step-down accounting. *Health Policy and Planning*, 19(2), 12735.
<https://doi.org/10.1093/heapol/czh015>.

Este artículo describe en detalle la metodología de costeo total aplicada a los programas de salud.

Drummond, M. F. (dir.). (2007). *Methods for the Economic Evaluation of Health Care Programmes* (3^e éd.). Oxford : Oxford University Press.

Este libro proporciona una introducción a la teoría y la práctica de la evaluación económica. Describe los métodos para medir los costos, los resultados y la recogida de datos durante los estudios clínicos y cómo manejar la incertidumbre en la evaluación económica.

Marseille, E., Larson, B., Kazi, D.S., Kahn, J. G. et Rosen, S. (2015). Thresholds for the cost-effectiveness of interventions: alternative approaches. *Bulletin of the World Health Organization*, 93(2), 11824. <https://doi.org/10.2471/BLT.14.138206>

En este artículo se explican los principales inconvenientes del método propugnado desde hace más de 10 años por el proyecto CHOICE para evaluar la efectividad en función de los costos de un programa (comparación de la RCI con el PNB per cápita) y se recomiendan enfoques alternativos.

WHO-CHOICE. (s. d). Quantities and unit prices (cost inputs)
<http://www.who.int/choice/cost-effectiveness/inputs/en/>

El sitio web del proyecto CHOICE presenta los métodos e instrumentos desarrollados por la OMS para las evaluaciones económicas, incluido el método de los ingredientes.

Referencias

Bedelu, M., Ford, N., Hilderbrand, K. et Reuter, H. (2007). Implementing antiretroviral therapy in rural communities: the Lusikisiki model of decentralized HIV/AIDS care. *The Journal of Infectious Diseases*, 196(Suppl. 3), S464-S468.

<https://doi.org/10.1086/521114>

Bemelmans, M., Van Den Akker, T., Ford, N., Philips, M., Zachariah, R., Harries, A., ... et Massaquoi, M. (2010). Providing universal access to antiretroviral therapy in Thyolo, Malawi through task shifting and decentralization of HIV/AIDS care: Task shifting to support universal access to ART. *Tropical Medicine et International Health*, 15(12), 1413-1420.

<https://doi.org/10.1111/j.1365-3156.2010.02649.x>

- Chan, A. K., Mateyu, G., Jahn, A., Schouten, E., Arora, P., Mlotha, W., Kambanji, M. et van Lettow, M. (2010). Outcome assessment of decentralization of antiretroviral therapy provision in a rural district of Malawi using an integrated primary care model: assessment of decentralization of antiretroviral therapy provision. *Tropical Medicine et International Health*, 15, 90-97.
<https://doi.org/10.1111/j.1365-3156.2010.02503.x>.
- Cohen, R., Lynch, S., Bygrave, H., Eggers, E., Vlahakis, N., Hilderbrand, K., ... et Makakole, L. (2009). Antiretroviral treatment outcomes from a nurse-driven, community-supported HIV/AIDS treatment programme in rural Lesotho: Observational cohort assessment at two Years. *Journal of the International AIDS Society*, 12(1), 23.
<https://doi.org/10.1186/1758-2652-12-23>.
- Loubiere, S., Boyer, S., Protopopescu, C., Bonono, C. R., Abega, S., Spire, B. et Moatti, J.P. (2009). Decentralization of HIV care in Cameroon: Increased access to antiretroviral treatment and associated persistent barriers. *Health Policy*, 92(23), 165-173.
<https://doi.org/10.1016/j.healthpol.2009.03.006>
- Mulamba, D., Fullem, A., Hirschhorn, L., Allers, C., Oser, R. et Rau, B. (2010). *Decentralizing antiretroviral treatment services at primary health care facilities: A guide to expanding access and care in resource-limited, generalized HIV epidemic settings*. PUBL?
https://aidsfree.usaid.gov/sites/default/files/decentralizing_art_services_primary_hc_facilities.pdf
- Ridde, V. et Dagenais, C. (dir.). (2012). *Approches et pratiques en évaluation de programmes : nouvelle édition revue et augmentée*. Montréal : Les Presses de l'Université de Montréal.
- Turashvili, M., Becher, H., Kerschberger, B., Jouquet, G., Browne, M., Haye, J. L., ... et Lassoovski, M. (2014). Decentralisation of HIV/TB care in Shiselweni region of Swaziland: Making a difference. *Titre de la revue, volume(numéro)*, p.-p.
<http://fieldresearch.msf.org/msf/handle/10144/325599>
- Joint United Nations Programme on HIV/AIDS. (2010). *Swaziland Country Progress Report*.
http://data.unaids.org/pub/report/2010/swaziland_2010_country_progress_report_en.pdf.

- United Nations Development Programme. (2009). *Overcoming barriers: Human mobility and development. Human development report*. Houndmills: Palgrave Macmillan.
- World Health Organization. (2002). *The world health report 2002: Reducing risks, promoting healthy life*. Geneva : World Health Organization.
<http://www.who.int/whr/2002/en/>
- World Health Organization. (2009). *Global tuberculosis control: epidemiology, strategy, financing*. Geneva : World Health Organization.
- Organisation mondiale de la Santé. (2010). *Guide sur les indicateurs pour le suivi et le bilan de la riposte du secteur de la santé au VIH/sida*. Geneva : World Health Organization.
http://www.who.int/hiv/data/ua10_indicator_guide_fr.pdf
- World Health Organization. (2012). *Retention in HIV programmes: Defining the challenges and identifying solutions*. Geneva : World Health Organization.
http://www.who.int/hiv/pub/meetingreports/retention_programmes/en/

Resumen / Résumé / Abstract

La evaluación económica aquí descrita se completó en 2013 y forma parte de una evaluación más amplia realizada por cuatro evaluadores independientes encargados por Médicos Sin Fronteras (MSF) con el doble propósito de documentar las experiencias y la promoción. El programa evaluado comenzó en 2007 en la región de Shiselweni (Swazilandia) en asociación con MSF y el Ministerio de Salud Pública. Este programa de tratamiento del SIDA tenía la particularidad de estar descentralizado e integrado a nivel de los centros de atención primaria de la salud. Este fue un enfoque innovador en Swazilandia, donde el tratamiento del SIDA se centralizó entonces en el nivel de centro de salud secundario. Los enfoques descentralizados del tratamiento del SIDA en los países en desarrollo han sido objeto de varios estudios que demuestran su eficacia. El propósito de esta evaluación era añadir un componente económico. Por lo tanto, realizamos un estudio retrospectivo de la eficacia en función de los costos desde el punto de vista del proveedor de servicios con el cálculo del índice de eficacia en función

de los costos incrementales (ICER). Las principales dificultades para una evaluación como ésta radican 1) en el acceso a los datos financieros de los proveedores de servicios 2) en la medición de la ganancia de eficiencia relacionada con el enfoque alternativo mediante el aislamiento de posibles factores de confusión.

L'évaluation économique décrite ici a été achevée en 2013 et fait partie d'une évaluation plus globale menée par quatre évaluateurs et évaluatrices indépendants à la demande de Médecins sans Frontière (MSF) dans un double but de documentation des expériences acquises et de plaidoyer. Le programme évalué a débuté en 2007 dans la région de Shiselweni au Swaziland en partenariat avec MSF et le ministère de la Santé publique. Ce programme de traitement du SIDA avait la particularité d'être décentralisé et intégré au niveau des centres de santé primaire. Il s'agissait d'une démarche innovante au Swaziland où le traitement du SIDA était alors centralisé au niveau des centres de santé secondaire. Les approches décentralisées du traitement du SIDA dans les pays en développement ont fait l'objet de plusieurs études montrant leur efficacité. Le but de la présente évaluation était d'y ajouter un volet économique. Nous avons donc mené une étude rétrospective de coût-efficacité du point de vue du prestataire de service avec calcul du ratio de coût efficacité incrémental (ICER). Les principales difficultés pour une évaluation telle que celle-ci résident 1) dans l'accès aux données financières des prestataires de service 2) dans la mesure du gain d'efficacité liée à l'approche alternative en isolant les éventuels facteurs de confusion.

The economic evaluation described here was finalized in 2013 and is part of a more comprehensive evaluation conducted by four independent evaluators and sponsored by Médecins sans Frontières (MSF) for the purpose of documenting past experiences and advocacy. The evaluated program started in 2007 in the Shiselweni region of Swaziland in partnership with MSF and the Ministry of Public Health. This AIDS treatment program had the specificity of being decentralized and integrated at the level of primary health centers. This was an innovative approach in Swaziland where AIDS treatment was centralized at the secondary health center level. Decentralized approaches to AIDS treatment in developing countries have been the subject of several studies showing their effectiveness. The purpose

of this evaluation was to add an economic component. We have therefore conducted a retrospective cost effectiveness study from the service provider perspective with calculation of the incremental cost-effectiveness ratio (ICER). The main challenges for an evaluation such as this one reside 1) in the access to the financial data of the service providers 2) to the measure of the effectiveness gain compared to the alternative approach and isolating possible confounding factors.

Guillaume Jouquet es un consultor independiente en economía de la salud y gestión de proyectos. Comenzó su carrera en el ámbito de la ayuda humanitaria en 2005 con Médicos sin Fronteras como coordinador adjunto de un programa sobre el SIDA en Sudáfrica. Desde entonces, ha acompañado y evaluado varias docenas de proyectos sanitarios humanitarios, principalmente en el África subsahariana.

Cita

Guillaume Jouquet (2020). Análisis de coste-efectividad. Una intervención para descentralizar la atención del VIH/SIDA en Shiselweni (Swazilandia). En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 325-346. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

13. Análisis espacial

*Un caso de intervención comunitaria para controlar el mosquito *Aedes aegypti* en Burkina Faso*

EMMANUEL BONNET, SAMIRATOU OUÉDRAOGO Y DIANE SARÉ

Análisis espacial

Definición del método

Método de análisis para medir el grado de similitud de un lugar con sus vecinos

Fortalezas del método

Los resultados de la cartografía pueden explicarse fácilmente a los decisores y al público en general

Desafíos del método

- No tiene en cuenta otros factores que pueden influir en los resultados de la intervención (por ejemplo, las características ambientales, el cambio de las precipitaciones y/o las condiciones atmosféricas, etc.).
- Es necesario integrar en el protocolo de evaluación la adquisición de datos geolocalizados de una muestra geográfica.
- Se requieren conocimientos básicos de análisis espacial.

El objetivo de los métodos de análisis espacial es determinar las características de la distribución espacial de las unidades geográficas (por ejemplo, los hogares) o sus valores (por ejemplo, el número de casos de dengue en el hogar). Estas características pueden analizarse globalmente, es decir, a gran escala, como una ciudad entera, para lo cual los métodos mostrarán la nube de puntos de las unidades geográficas, las relaciones espaciales entre los valores y los individuos (dependencia espacial), y las características geométricas (la organización de los valores en el territorio, como un gradiente norte-sur, por ejemplo) de la nube de puntos o los valores. Cuando los valores estudiados son cuantitativos, es posible analizar la distribución geográfica de los valores asociados a los individuos, examinando en particular si existe una relación con la proximidad entre estos individuos. El análisis adopta la forma de un índice que expresa si existe o no una correlación entre el valor de los individuos y su proximidad. Por

ejemplo, permite caracterizar si los valores similares tienden a ser cercanos en el territorio o, por el contrario, están totalmente dispersos. Si se agrupan los valores, hablamos de una autocorrelación espacial positiva. Si se dispersan, hablamos de autocorrelación espacial aleatoria. Por lo tanto, este método tiene como objetivo caracterizar la dependencia espacial de los valores.

Muchas situaciones en la atención sanitaria no son homogéneas (también llamadas estacionales). Nos referimos particularmente en la concentración de casos de enfermedad en localidades muy específicas, unas pocas calles de un barrio, por ejemplo. Imaginemos un barrio cuyo borde está cerca de una presa. Los casos de paludismo existen en todas partes en esta zona, pero están más presentes y concentrados en la calle que bordea la presa. La distribución espacial del paludismo sería en este caso heterogénea a nivel del barrio. Sin embargo, las medidas globales de autocorrelación espacial, a escala de todo el vecindario, pueden enmascarar este fenómeno de concentración localizada cuando la autocorrelación varía de un lugar a otro. El índice de autocorrelación espacial obtenido podría significar que la distribución espacial de los casos de paludismo es homogénea en el vecindario, ya que el cálculo, aplicado a todo el territorio, podría minimizar la concentración identificada en la calle cercana a la presa. En esta situación se dice que hay asociaciones espaciales de valores locales significativos que no son detectables por la medida clásica de autocorrelación espacial global definida por Moran, por ejemplo.

Uno de los objetivos del análisis espacial es detectar formas o lugares singulares en los que se producen altas concentraciones locales de valores, también llamados agregados (de individuos o valores). Así pues, se dispone de métodos para realizar análisis locales muy precisos, lo que resulta útil para orientar determinadas acciones, por ejemplo. Luc Anselin (Anselin, 1995) ha desarrollado un índice para compensar el efecto de la ocultación de las autocorrelaciones locales por los índices globales: el Indicador local de asociación espacial (LISA) (Local indicator of spatial association).

El LISA mide si hay lugares cuyos valores locales están significativamente relacionados con la proximidad de esos lugares y si hay agrupaciones de valores muy altos o muy bajos entre esos lugares. Por lo tanto, estos indicadores miden la concentración de los valores numéricos medidos en un conjunto de puntos. El cálculo se basa en la autocorrelación global de Moran, como la media de los productos de los valores normalizados

de los pares de puntos, ponderados por un peso espacial que se representa por la distancia. El análisis LISA se realiza en cada hogar, representado por un punto, y en las relaciones de cada hogar con sus vecinos. Consiste primero en calcular un índice para cada punto (hogar) de su vecindario. Los indicadores estadísticos expresan el grado de confianza y la importancia que se puede dar a la información adquirida. En el caso de unidades espaciales irregulares para las que sólo se tiene en cuenta la matriz de cercanía, el peso es igual a 1 cuando las unidades son contiguas y 0 para las demás. Así, alrededor del punto observado, el LISA local determina la extensión de una región (*cluster*) formada por puntos que tienen valores cercanos a los suyos. La suma de las LISAs locales es proporcional al indicador global. El indicador local se calcula de manera similar al índice de Moran. La interpretación de las LISA obtenidas se basa en supuestos estadísticos de normalidad. Por consiguiente, es necesario comprobar la importancia estadística de las estructuras de los clusters en la región. Se calcula un valor p para cada punto: cuanto más bajo es el valor p, más significativa es estadísticamente la existencia de cualquier cluster que forme con sus vecinos (Caloz y Collet 2011). Esto resulta en cinco casos. El primero es cuando el valor p no permite rechazar la hipótesis nula: el punto no se considera significativo para ser agregado a otro punto. En los otros cuatro casos, los índices LISA se utilizan para una tipología de cuatro cluster, según el valor del individuo y el valor del barrio. Estos *cluster* están representados en la tabla 1, con el color que se utiliza habitualmente para su cartografía (Oliveau 2010).

		Valor de la unidad estadística	
		Débil	Fuerte
Valor medio del vecindario	Fuerte	Asociación espacial negativa	Asociación espacial positiva
	Débil	Asociación espacial positiva	Asociación espacial negativa

Tabla 1 : Índices LISA

Por consiguiente, este tipo de método es adecuado, pero se utiliza raramente, para evaluar los efectos de una intervención en el espacio, a escalas geográficas muy finas, como la ubicación de los hogares, por ejemplo. Los valores analizados a nivel de hogar (coordenadas geográficas) pueden

representar, por ejemplo, a varias personas afectadas por una enfermedad (datos cuantitativos). El LISA se calculará y mapeará antes y después de la intervención para evaluar los efectos.

Análisis espacial de los efectos de una intervención comunitaria de control del dengue en Uagadugú

Tras varias fases de aparición del dengue en Uagadugú (Burkina Faso), se decidió elaborar y aplicar una intervención de base comunitaria para controlar el mosquito *Aedes aegypti*, principal vector del dengue en las zonas urbanas.

El ámbito del estudio y los participantes

El estudio se llevó a cabo en dos barrios comparables de Uagadugú (Tampouy y Juvenat), seleccionados entre cinco zonas identificadas de la ciudad. En efecto, en los dos barrios hay perfiles socioeconómicos muy contrastados: hogares acomodados que viven en casas modernas de hormigón que se abastecen totalmente de las redes de agua y electricidad, pero también hogares con un nivel de vida muy bajo en casas construidas con arcilla, a veces sin servicios básicos (Observatoire des populations de Ouagadougou 2013). Estos barrios se caracterizan también por zonas en las que no se lleva a cabo regularmente la gestión de residuos y en las que se acumulan desechos en determinadas calles, lo que favorece la proliferación de mosquitos.

Tampouy, que se encuentra en la parte noroccidental de la ciudad, fue seleccionada al azar para recibir la intervención, mientras que Juvenat, en la parte oriental de la ciudad, fue elegida como distrito de “control”, para comparar situaciones. En cada barrio delimitamos una zona con un radio de un kilómetro alrededor del centro de salud primaria



Foto: Charla educativa en un pequeño mercado del distrito de Tampouy (Crédito: Diane Saré)

(considerado como un punto de referencia). En Tampouy, la intervención se llevó a cabo dentro de este perímetro. Se utilizaron las coordenadas geográficas de los hogares para dibujar 287 hogares en Tampouy y 289 hogares en Juvenat dentro de este perímetro (dibujar sin reemplazo) a fin de reunir los datos entomológicos (larvas y pupas de mosquitos) necesarios para evaluar el impacto de la intervención.

La teoría de la intervención

1. Movilización y organización: este paso permitió establecer contacto con los líderes comunitarios y las asociaciones formales e informales que trabajan en la zona de intervención de Tampouy para informarles de la situación del dengue en su barrio. Este paso promovió la concienciación de personas influyentes sobre la situación de esta enfermedad y su participación en la lucha contra ella.
2. Planificación operativa: consistió en un desarrollo participativo de la intervención con la intervención de personas influyentes y asociaciones dispuestas a participar.
3. Acción comunitaria: consistía en la realización de las actividades de intervención por parte de los miembros de la comunidad (facilitadores comunitarios y grupo de teatro) en beneficio de la comunidad. Esas actividades tenían por objeto alentar la participación efectiva de los hogares y la población.
4. Monitoreo/evaluación dirigida por la comunidad: esto permitió una mejor participación de la comunidad.

La interacción de estos diferentes componentes tuvo como efectos directos, el aumentar el conocimiento sobre el dengue y las enfermedades febriles transmitidas por vectores, fortalecer la capacidad de la comunidad en la zona de intervención y reducir los índices entomológicos. El presente capítulo no tiene por objeto describir estos resultados.

Los efectos distales previstos son el mantenimiento de la lucha contra los mosquitos, lo que dará lugar a una reducción de la población de vectores del dengue y el paludismo y, por consiguiente, a una disminución de los episodios y hospitalizaciones relacionados con el dengue y el paludismo.

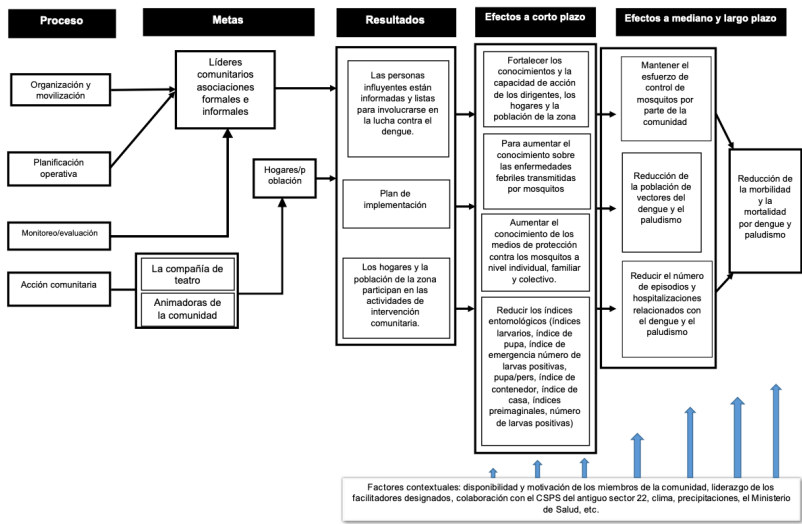


Figura 1. Modelo lógico de la intervención comunitaria para luchar contra el dengue en Tampouy (ex sector 22 de la ciudad de Uagadugú)

Valoración

En la tabla 2, resumimos la estimación de la evaluación de esta intervención en el contexto específico de este capítulo.

Tabla 2 : Resumen del diseño de evaluación de la intervención comunitaria para combatir el dengue

Objetivo de la evaluación	Preguntas de evaluación	Planificación
Evaluar el impacto de la intervención en la densidad y distribución geográfica de las larvas y pupas de <i>Aedes aegypti</i> en el distrito de Tampouy.	¿Disminuyó el número de larvas y pupas en Tampouy después de la intervención en comparación con “Juvenat”, que no recibió intervención? Tras la intervención, ¿persisten en Tampouy zonas de densidad de larvas y pupas en comparación con el “Juvenat” que no recibió intervención?	Estudios entomológicos previos a la intervención, octubre de 2015: los encuestadores buscaron y recogieron larvas y pupas de <i>Aedes aegypti</i> en cada concesión visitada en los dos distritos. De junio a principios de octubre de 2016: realización de actividades de intervención en Tampouy. Estudios entomológicos después de la intervención de finales de octubre a principios de noviembre de 2016: las personas investigadoras buscaron y recogieron larvas y pupas de <i>Aedes aegypti</i> en cada concesión visitada en los dos distritos.

Aplicación del método LISA

Los Indicadores de Asociación Espacial Local permiten la exploración espacial local del impacto de la intervención comunitaria para combatir el *Aedes aegypti* en el distrito de Tampouy de Uagadugú. Sobre la base de los datos de los estudios entomológicos realizados antes y después de la intervención, el procedimiento de análisis consta de tres pasos: 1) cartografía de la distribución espacial del número de larvas y pupas, 2) medición de esta distribución, y 3) cálculo de los indicadores de asociación espacial local y cartografía de los mismos.

Etapas 1: Análisis de la distribución espacial del número de larvas y pupas a través de la cartografía

El primer paso es producir un mapa de los valores en cada barrio antes y después de la intervención utilizando una escala de representación común. Estas representaciones gráficas se describen detalladamente en varios

manuales de cartografía, que permiten elegir la representación gráfica adecuada para cada tipo de variable y cada tipo de objeto geográfico (puntos, líneas, superficies) (Bertin 1967; Bailly 1992).

Se pueden utilizar varios programas de software, de pago (ArcGIS o Mapinfo) o libres (QGIS), para trazar un mapa del número de larvas y pupas utilizando un tamaño de punto proporcional al número. Es necesario disponer de las coordenadas geográficas de cada hogar en el que se haya realizado el estudio entomológico y para el que se haya calculado un número total de larvas y pupas. Por ello, a fin de evitar el sesgo de interpretación, es necesario un diseño de muestreo espacial (muestreo aleatorio en el espacio) para la recolección entomológica.

Los mapas (Figura 1) representan todos los hogares en los que se recogieron datos. Si no se han encontrado larvas, el hogar está representado por un punto azul en el mapa; no expresa ninguna cantidad. Para los demás hogares, el punto negro representado es gráficamente proporcional al número de larvas y pupas calculado. La comparación de los mapas permite observar varios fenómenos.



Foto: Identificación de los depósitos de larvas durante las visitas de puerta en puerta (Crédito: Diane Saré)

Durante la fase inicial, antes de la intervención, se observa un número importante de larvas y pupas en las zonas de intervención y control. En cada una de estas zonas aparecen zonas con un mayor número de larvas y pupas que en otras partes de los barrios. Esto ilustra que las larvas de *Aedes aegypti* están presentes en todas partes de Uagadugú y que hay, por múltiples razones, “bolsas” en las que estas larvas son más numerosas.

Durante el último análisis posterior a la intervención, se observó una disminución significativa del número de hogares con alojamientos de larvas y pupas en Tampouy, la zona de intervención. En el sector de control, Juvenat, que no fue objeto de intervención, se observa que los depósitos identificados durante la fase inicial están en

su mayoría presentes, y la aparición de nuevos depósitos. Cabe señalar que los efectos no se asocian con una menor precipitación en una zona en comparación con otra, y de un año a otro, y por lo tanto con una menor productividad de los depósitos larvarios.

Este primer paso es sencillo de llevar a cabo; proporciona una representación cartográfica del fenómeno. Sin embargo, para caracterizar los “bolsillos” y medir si existe una concentración del número de sitios de cría y larvas en ciertas partes de los barrios y cómo ha evolucionado después de la intervención, se deben hacer mediciones globales y luego locales de los valores asociados a los hogares.



Figura 2. Cartografía del número de larvas y pupas en las zonas de estudio antes y después de la intervención

Fase 2: Medición de la distribución espacial del número de larvas y pupas en las zonas de estudio de Uagadugú

En esta fase se mide la autocorrelación espacial para determinar si existe una correlación entre el valor de los objetos y las relaciones métricas o topológicas entre ellos. Los índices de correlación se utilizan para dar cuenta de los valores vecinos, el más utilizado de los cuales es el índice de Moran. Se

define como el promedio de los productos de los valores normalizados de los pares de puntos, ponderados por la distancia entre los dos puntos (Moran, 1948).

$$I_{Moran} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Donde ij = unidad espacial; n =numero de unidades espaciales; x_i es el valor de la variable en la unidad i ; \bar{x} es la media de x ; y w_{ij} son los elementos de la matriz de interacciones espaciales, definida como la cercanía, las distancias o las fronteras comunes.

Por consiguiente, este índice mide la medida en que, para la totalidad de la muestra, el valor de los individuos se aproxima o se aleja del de su vecindario en relación con la media. Esta medida es global (Cliff, 1969) ya que se relaciona con la similitud de los lugares con su vecindario para toda la muestra. En el caso de esta evaluación, el índice de Moran indica si las viviendas de los hogares con un número significativo de larvas y pupas están cerca o lejos, lo que revela una autocorrelación espacial positiva o negativa. Este es el primer paso antes de intentar investigar más a fondo la ubicación de los agregados.

El valor del índice de Moran en el barrio de Tampouy antes de la intervención muestra una autocorrelación espacial positiva (si el índice es superior a cero), relativamente baja ($i=0,10$), pero con una dispersión de los valores del índice que sugiere que los agregados pueden estar presentes. Tras la intervención, observamos que la autocorrelación espacial es cero ($i=-0,02$), lo que significa que los valores están distribuidos aleatoriamente, sin autocorrelación, y muestra una evolución entre las dos fases de recogida y por lo tanto un efecto de la intervención. En el sector de control, la autocorrelación espacial global es también positiva ($i=0,05$) antes de la intervención y nula ($i=-0,004$) después. Por lo tanto, la evolución general en los dos barrios es similar. Sin embargo, es importante verificar la evolución a nivel local, es decir, a nivel de la vivienda, en particular para comprobar si hay agregados y, en caso afirmativo, cómo han evolucionado.

Fase 3: Indicadores locales de la asociación espacial del número de larvas y pupas en los barrios

Los indicadores de asociación espacial local desarrollados por Luc Anselin permiten llevar a cabo estas evaluaciones. Analizan la concentración de valores similares y disímiles medidos en un conjunto de puntos o calculados por agregación espacial.

$$I_i = \frac{\sum_j w_{ij} (p_i - \bar{p}) (p_j - \bar{p})}{\sum_i (p_i - \bar{p})^2}$$

Donde p_i y p_j son los valores de las unidades espaciales i y j , de tal manera que i y j se consideran vecinos en base a la medición de su grado de proximidad, \bar{p} es el valor medio de las unidades espaciales y w_{ij} una medida de la proximidad de las unidades espaciales i y j .

El procedimiento se puede llevar a cabo con el software libre Geoda (Anselin, 1995). Permite llevar a cabo todos los pasos anteriores y no requiere el uso de un sistema de información geográfica para preparar los datos que se utilizarán. Por otra parte, los resultados pueden ser exportados a GIS. Antes de analizar los indicadores locales de asociación espacial, debe definirse una matriz de pesos (una tabla que enumera todas las distancias entre todos los vecinos) que identifique a los vecinos. Los pesos de los vecinos pueden construirse de varias maneras dependiendo del tipo de objeto geográfico (un punto o una superficie). En el caso de la intervención para el control del dengue que se presenta aquí, los hogares están representados por puntos. Por lo tanto, la matriz se calcula a partir de una distancia entre los puntos.

El último paso es ejecutar el análisis LISA usando esta matriz de peso. Se calcula una estadística para cada unidad, que se basa en el valor de las unidades vecinas. El resultado son dos tipos de mapas: uno que representa los valores del índice calculado y muestra cómo varía el índice en todo el territorio, y otro que pone de relieve las relaciones estadísticamente significativas de los índices mediante la cartografía de los agregados para ilustrar las similitudes y las disimilitudes de los valores.

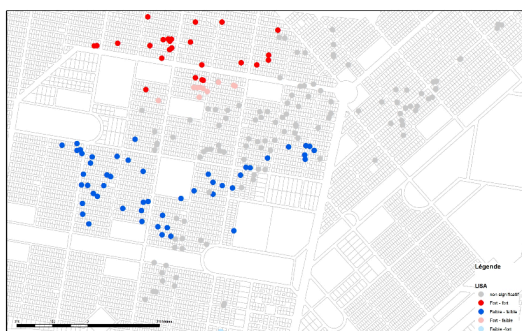


Figura 3a. Tampouy antes del procedimiento

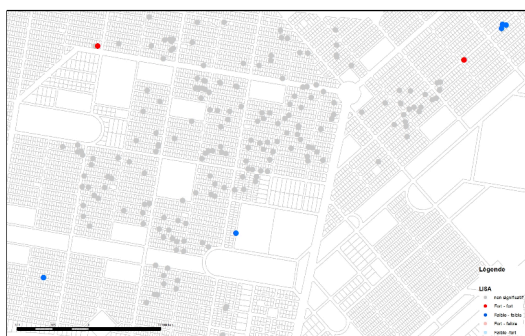


Figura 3b. Tampouy después del procedimiento

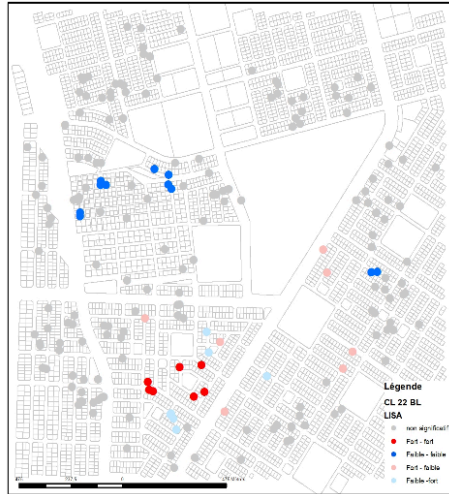


Figura 3c. Juvenil antes del procedimiento

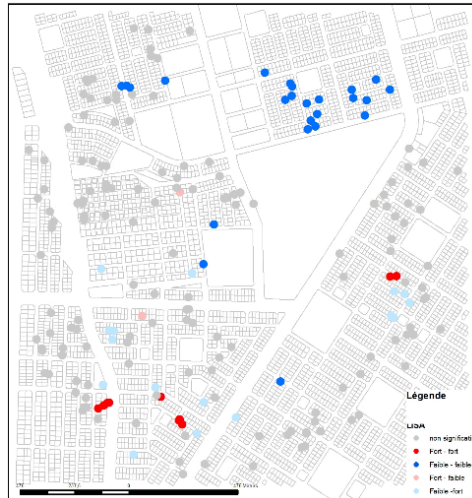


Figura 3d. Juvenil después de la cirugía

Los mapas LISA (figuras 3a, 3b, 3c y 3d) de los valores larvarios y de pupas de la zona de intervención de Tampouy (arriba) muestran que, antes

de la intervención, había agregados de valores negativos (azul) y positivos (rojo). Esto significa que había dos tipos de concentraciones, ambas muy claras en el vecindario. En el azul, había una concentración de hogares que tenían hábitats de larvas y pupas relativamente improductivos porque tenían un número bajo de larvas y pupas. Esta concentración de valores bajos también significa que los hogares vecinos tenían características similares (bajo número de larvas). Por el contrario, en el norte de la zona, la concentración de valores altos está representada por los puntos clasificados como fuertes-fuertes (valores fuertes del hogar y sus vecinos). Por lo tanto, el análisis LISA muestra una doble concentración de valores opuestos de larvas y pupas. Esto no significa una presencia o ausencia de larvas y pupas, sino que había zonas graduadas de producción de larvas y pupas en el área de estudio antes de la intervención.

Después de la intervención, los análisis de LISA muestran pocos agregados de cualquier tipo y la desaparición de los agregados que existían antes de la intervención. Por consiguiente, el análisis muestra que la intervención dio lugar a una reducción del número de zonas con concentraciones altas y bajas de larvas y pupas. El análisis general confirma esto con una autocorrelación espacial nula, lo que significa que la presencia de larvas y pupas en la zona de intervención es ahora aleatoria. Este análisis complementa los análisis de los epidemiólogos que observaron una reducción significativa de los índices entomológicos en Tampouy después de la intervención.

En la zona de control se observaron agregados que persistieron en los mismos lugares durante la recolección posterior a la intervención y confirmaron el aumento de la producción de larvas y pupas identificado por los demás análisis epidemiológicos.

Análisis reflexivo y conclusión

La evaluación de la dimensión espacial en las intervenciones de salud pública suele limitarse a un análisis de los efectos medios de la intervención que varían según las características geográficas y las medidas aproximadas de la distancia. Los métodos tradicionales utilizados para la evaluación no captan la variabilidad espacial de los lugares y los efectos de las intervenciones. El método utilizado en el presente capítulo es un ejemplo

de análisis espacial que puede utilizarse. Estos métodos se utilizan rara vez para evaluar las intervenciones y más a menudo para explorar la propagación espacial o la concentración de una enfermedad. Sin embargo, este ejemplo demuestra que los fenómenos pueden observarse a diferentes escalas y que los resultados globales pueden ocultar diferentes situaciones locales. La dimensión espacial de estos análisis complementa otros resultados en el contexto de la evaluación del impacto de una intervención. Por consiguiente, se recomienda generalizar el uso de análisis espaciales en la evaluación de las intervenciones a fin de aprehender todos los efectos a todas las escalas.

En el ejemplo presentado, el procedimiento utilizado demostró ser eficaz y no planteó ninguna dificultad siempre que la toma de muestras fuera geográfica. Este es uno de los puntos fundamentales de los métodos de análisis espacial. La calidad de los resultados depende de un plan de muestreo geográfico que asegure la correcta interpretación de los resultados. Uno de los retos de este método es encontrar los hogares en la fase posterior a la intervención. En los países en que no existen direcciones de calles, es necesario utilizar el GPS y asegurarse de que la geolocalización es correcta y está bien registrada. En entornos urbanos densos, algunas perturbaciones de la señal interfieren con las posiciones registradas en el GPS y hacen más difícil encontrar los hogares encuestados en la fase previa a la intervención. Al realizar encuestas de población, asegúrese de registrar el número de teléfono del encuestado para que sirva como una forma alternativa de localizar el hogar. La falta de rastreo del hogar de la encuesta original es la principal debilidad del método en lo que respecta a su aplicación en los países de ingresos más bajos. Otras deficiencias se refieren al propio método, que permite el análisis de una sola variable a la vez, lo que limita la inclusión de otras variables que pueden considerarse confusas. Sin embargo, hay otros métodos geográficos que combinan varias variables y serían complementarios (regresiones de GWR ponderadas geográficamente) de las evaluaciones estadísticas.

El acceso a estos métodos se facilita hoy en día por la disponibilidad de software gratuito de Sistemas de Información Geográfica como QGIS. Permiten cartografiar la geolocalización de los hogares en una encuesta, pero también realizar muestreos geográficos. Asociado a las herramientas libres “GeoDa” desarrolladas por Luc Anselin, también se facilita la realización de LISA.

Referencias

- Anselin, L., (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27(2), 93-115.
- Bailly A. S. (1999). Pumain D. et Saint-Julien T., *L'analyse spatiale T. 1. Annales de Géographie*, (605), 105.
- Bertin, J. (1967). *La sémiologie graphique : les diagrammes, les réseaux, les cartes*. La Haye ou Paris : Mouton ou Gauthier-Villars.
- Caloz, R. et Collet, C. (2011). *Analyse spatiale de l'information géographique* (1^e éd.). Lausanne : Presses polytechniques et universitaires romandes
- Cliff, A. D. et Ord, K. J. (1969). The problem of spatial autocorrelation. Dans A. J. Scott (dir.), *Papers in Regional Science, Studies in Regional Science* (vol. 1, p. 25-55). Londres: Pion.
- Moran, P. A. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society: Series B*, 10(2), 243-251.
- Observatoire des populations de Ouagadougou, Projet Wellcome Trust 2008-2013. (2013). *Rapport scientifique sur la caractérisation sociologique des quartiers de l'OPO*.
<http://www.issp02.issp.bf/opo/Publications/Rapport-scientifique-juin2013-Caracterisation-des-quartiers-Axe-qualite.pdf>
- Oliveau, S. (2010). Autocorrélation spatiale : leçons du changement d'échelle. *L'Espace Géographique*, 39(1), 51-64.

Resumen / Résumé / Abstract

Este capítulo presenta un análisis espacial de una intervención para evaluar si los valores similares de un resultado tienen más probabilidades de formar grupos en un área de estudio. Este método, denominado “indicadores locales de asociación espacial” (LISA), identifica áreas en las que los valores dependen espacialmente unos de otros. LISA se utilizó para evaluar el impacto de una intervención comunitaria para controlar al mosquito responsable de la propagación del dengue *Aedes aegypti* en dos barrios de

Uagadugú, la capital de Burkina Faso. La evaluación final de la intervención utilizando este método mostró que los grupos de larvas desaparecieron en el área que recibió la intervención mientras persistían en el área de control.

Ce chapitre présente une analyse spatiale d'une intervention qui vise à évaluer si des valeurs similaires d'un résultat sont plus susceptibles de former des grappes dans une zone d'étude. Cette méthode appelée « indicateurs locaux d'association spatiale » (local indicators of spatial association – LISA) permet d'identifier les zones où les valeurs sont spatialement dépendantes les unes des autres. LISA a été utilisé pour évaluer l'effet d'une intervention communautaire visant à contrôler le moustique responsable de la propagation de la dengue *Aedes aegypti* dans deux quartiers de Ouagadougou, capitale du Burkina Faso. L'évaluation finale de l'intervention à l'aide de cette méthode a montré que les groupes de larves ont disparu dans le quartier qui a reçu l'intervention alors qu'ils persistaient dans le quartier témoin.

This chapter present a spatial analysis of an intervention which aims at assessing whether similar values of an outcome are more likely to form clusters in a study area. This method called local indicators of spatial association (LISA) enable to identify areas where values are spatially dependant with one another. LISA has been used to evaluate the effect of a community-based intervention which aims at controlling the mosquito responsable for spreading Dengue, *Aedes aegypti* in two neighborhoods of Ouagadougou the capital city of Burkina Faso. The final evalutation of the intervention using this method showed that larva clusters disappeared in the neighborhood which received the intervention while they persisted in the control neighborhood.

Emmanuel Bonnet es geógrafo de la salud e investigador del IRD (Institut de Recherche pour le Développement). Ha realizado muchas investigaciones en el África occidental y tiene experiencia internacional en los ámbitos de vigilancia de epidemias, tecnologías geomáticas, sistemas de información geográfica y el análisis espacial.

Samiratou Ouédraogo es epidemióloga e investigadora de salud pública con experiencia internacional en evaluación. Trabaja en numerosos proyectos de investigación en Canadá y África.

Diane Saré es médico de salud pública. Tiene experiencia profesional clínica comunitaria y hospitalaria. También tiene experiencia en la aplicación y evaluación de intervenciones comunitarias.

Cita

Emmanuel Bonnet, Samiratou Ouédraogo y Diane Saré (2020). Análisis espacial. Un caso de intervención comunitaria para controlar el mosquito *Aedes aegypti* en Burkina Faso. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 347-364. Québec: Éditions science et bien commun y Marseille: IRD Éditions

14. Análisis coste efectividad

Evaluación de la eficiencia de un programa para incrementar el número de partos vaginales realizados en situaciones de feto en posición no cefálica

LETICIA GARCIA MOCHON Y CLARA BERMÚDEZ-TAMAYO

Análisis coste efectividad

Definición del método

Evaluación de la eficiencia de una tecnología, a través de un modelo de análisis a corto plazo para analizar la efectividad, los costes y la ratio coste efectividad incremental (RCEI). El análisis se realizó desde el punto de vista del sistema sanitario, es decir, que se incluyeron únicamente los costes directos soportados por el sistema.

Fortalezas del método

- El método de coste efectividad ofrece una información potente para tomar decisiones en el momento de decidir incluir nuevas intervenciones en los sistemas sanitarios, ya que evalúa conjuntamente los resultados en salud con el coste que supone la nueva intervención con la alternativa que se utiliza habitualmente.
- El modelo de decisión que se realizó para evaluar el RCEI permite combinar datos de distintas fuentes de información. Por un lado, se tomaron los datos de un ensayo clínico aleatorizado, pero para generalizar el modelo a otros contextos, se realizó adicionalmente una síntesis de evidencia de otros ensayos clínicos publicados en la literatura.

Desafíos del método

- Contar con datos que sean válidos para tomar decisiones en otros contextos
 - Tener en cuenta otros costes que pueden influir en la toma de decisiones como los riesgos asociados a la cesárea.
 - Comparar en costes y efectividad con la versión cefálica externa, técnica que cada vez más se utiliza para la versión cefálica.
-

La utilidad de la evaluación económica para tomar decisiones eficientes

En los últimos años hemos podido presenciar un aumento del uso de la evaluación de tecnologías sanitarias como instrumento de apoyo a la toma de decisiones para un uso más eficiente de los recursos públicos sanitarios. Uno de los principales retos a los que se enfrentan los sistemas sanitarios es asegurar la provisión de servicios a la población a la que ofrece

cobertura, teniendo en cuenta que los recursos con los que cuentan los gobiernos para la provisión de asistencia sanitaria son siempre limitados. La cobertura universal de la asistencia sanitaria implica que todas las personas con independencia del lugar en el que vivan y la posición social y económica que ocupen puedan tener acceso a la asistencia sanitaria (OMS,2010). Éste es un valor relevante en la mayoría de las sociedades desarrolladas. No obstante, el aumento del gasto público en salud, unido a las limitaciones cada vez más acuciantes en los presupuestos por la crisis económica, han puesto en duda la sostenibilidad de este sistema. Uno de los factores que más han influido en el incremento del gasto sanitario y que más pone en riesgo la sostenibilidad de la financiación de los sistemas de salud, es la inclusión de nuevas tecnologías sanitarias (Gonzalez 2007) El concepto de tecnologías sanitarias es amplio e incluye, medicamentos, dispositivos médicos, procedimientos clínicos y quirúrgicos e incluso nuevos modelos organizativos. No cabe duda que las tecnologías sanitarias suponen un elemento indispensable en la provisión de asistencia sanitaria. Sin embargo, no toda innovación supone un avance, mejorar la salud, o es costo-efectiva. El Informe sobre la salud en el mundo 2010 (OMS, 2010) identificaba el derroche de medicamentos y tecnologías sanitarias, como una de las principales causas de ineficiencia en la prestación de servicios de salud. Dirigir las sociedades hacia el principio universal de la salud, implica enfrentarse a la necesidad de priorizar decisiones eficientes de inversión a través de instrumentos que den soporte técnico a las decisiones políticas.

Las sociedades modernas intentan resolver esta priorización de asignación de recursos escasos a través de los mercados, partiendo de la identificación de asignaciones eficientes siempre que se cumplan una serie de exigencias técnicas, institucionales y de comportamiento de los agentes. Lamentablemente desde la obra seminal de Arrow (1963), se sabe que los mercados sanitarios no cumplen las expectativas de la “mano invisible” de Adam Smith aquejados de graves problemas de incertidumbre y asimetrías de información.

Que los mercados sanitarios no funcionen para asignar eficientemente los recursos, no implica que el problema desaparezca. Al contrario, vuelve más necesario e imprescindible el desarrollo de metodologías que sustituyan el papel que tienen los precios en los mercados de competencia perfecta. Es decir, ayuden en el proceso de toma de decisiones estableciendo qué tecnologías sanitarias son eficientes, y cuál es su coste de oportunidad

social. En este sentido la Evaluación Económica, que originariamente se encontraba asentada como aplicación práctica de la teoría económica del bienestar, actualmente se utiliza como parte de los procesos de Evaluación de Tecnologías Sanitarias.

¿Por qué combinamos los resultados en salud con los costes en las evaluaciones económicas? Simplemente, porque es necesario saber si las tecnologías valen lo que cuestan. Para ello es muy importante definir bien el resultado, que sea coherente con la intervención evaluada y la población a la que va dirigida, y sirva para tomar decisiones. Por otro lado, identificar adecuadamente el uso de recursos que suponen las alternativas que se están comparando y valorarlas económicamente mediante su coste.

Se pueden realizar evaluaciones económicas como parte de un estudio aleatorio, en el que además de evaluar la eficacia de la intervención, incluimos sus costes y los comparamos con los costes de la alternativa de comparación (idealmente la práctica habitual). Pero también se puede hacer evaluaciones económicas basadas en modelos, y en este caso, el uso de recursos y efectividad de la intervención la podemos tomar tanto de un único estudio como de la mejor evidencia disponible, a través de una revisión sistemática y metanálisis de estudios que aborden la efectividad de la técnica. El modelo simula una realidad clínica. La estructura del modelo debe reflejar nuestro 'modelo conceptual' de cómo la elección del tratamiento afecta el curso de la enfermedad y la relación entre la enfermedad y los costes y efectos. En el apartado X se presenta una descripción más detallada del modelo utilizado en este estudio.

En el caso de la evaluación aquí presentada, se optó por presentar la información relativa del ensayo clínico que se había realizado en el contexto de la Comunidad autónoma de Andalucía (España), para reflejar contextos asistenciales occidentales, en un país donde los cuidados obstétricos habituales incluyen técnicas posturales para abordar este problema. Pero también, se realizó una revisión de otra evidencia en la que incluyeron estudios realizados en otros países y contextos, como China. Con toda la evidencia, se realizaron varios análisis de coste efectividad para reflejar variabilidad en la efectividad de la técnica. Si un país quisiera adaptar estos datos a su medio, sólo tendría que actualizar el modelo con los costes propios de su sistema sanitario.

Las cesáreas como problema de salud global

Cada vez se hacen más cesáreas en el mundo, tanto en países desarrollados como en países en desarrollo. Los gobiernos y profesionales sanitarios de todo el mundo han expresado su preocupación al respecto, ya que las cesáreas cuando están justificadas son beneficiosas para disminuir la morbilidad materna y perinatal, pero hay muchos casos en los que no está indicada y supone un riesgo para la madre y el bebé (Chaillet 2015). Además, la cesárea está asociada con un mayor coste, principalmente por el mayor número de días de hospitalización tras el parto y la mayor tasa de rehospitalización (Bermudez Tamayo 2018).

Uno de los motivos para realizar una cesárea es cuando el bebé nace de cabeza (presentación no cefálica). Aproximadamente, entre un 3 y un 4% de los embarazos a término conllevan esta situación (Chaillet 2015). Debido a los riesgos que representa el parto vaginal en estos casos, suele proponerse la realización de una cesárea. La versión cefálica externa se realiza en los hospitales tras la exploración mediante cardiotocografía y ecografía. Esta técnica puede disminuir la cantidad de presentaciones de nalgas y de operaciones cesáreas, aunque la seguridad de la técnica no se ha evaluado mediante estudios aleatorios, si bien grandes estudios observacionales sugieren que son raras las complicaciones.

La moxibustión. Una terapia oriental para disminuir la tasa de cesáreas

La moxibustión es una terapia oriental que aprovecha la raíz de la planta *Artemisia vulgaris*, denominada ‘mogusa’ en Japonés (‘moxa’). La artemisia es secada, prensada y liada en papel de morera. Cuando se quema produce calor sin llama. La aplicación de calor mediante la combustión de la moxa se ha utilizado desde antaño como terapia en China. Entre otros efectos, parece que cuando se aplica en el punto de acupuntura V 67 situado en el ángulo ungueal externo del 5 dedo del pie puede corregir la presentación no-cefálica del feto. Revisiones sistemáticas recientes han informado sobre la efectividad de la aplicación de calor por moxibustión para promover la versión del feto en posición no-cefálica en comparación con los cuidados obstétricos habituales (Coyle 2012). Sin embargo, se recomienda precaución en la interpretación de los resultados debido a la considerable

heterogeneidad observada en los estudios. Se necesita más evidencia derivada de estudios clínicos controlados, aleatorizados, bien realizados para evaluar la efectividad y seguridad de la moxibustión.

En un centro de salud de la Comunidad Autónoma de Andalucía (España), un grupo de obstetras y matronas realizaron un estudio controlado y aleatorizado de moxibustión aplicada por no acupuntores en una población de estudio de 406 mujeres embarazadas (Vas 2013). El tratamiento se aplicó en primer lugar en los centros de salud por matronas que previamente habían recibido un curso de formación de 10 h, impartido por profesionales de la acupuntura con una experiencia de más de 10 años. El objetivo de esta primera sesión de entrenamiento era dotar de las habilidades necesarias, a la pareja o persona relacionada, para la aplicación de las sesiones de moxibustión restantes en el domicilio de la embarazada. Se pedía a la mujer embarazada que se acostase boca arriba, con las rodillas ligeramente flexionadas, para proceder a la aplicación del calor con el cigarro de moxa de artemisia durante 20 min en total en el punto de acupuntura V 67 de ambos pies. El calor se aplicó desde una distancia de 1,5-3,0 cm. Dos semanas más tarde se procedía a la realización de un examen clínico.

Este estudio confirmó la efectividad y la ausencia de efectos secundarios de la técnica de moxibustión en un contexto occidental. Se informó de una reducción del riesgo relativo (RR) de la presentación cefálica al nacer en el grupo de moxibustión verdadera en comparación con el grupo de atención habitual del 29,7% (intervalo de confianza del 95% [IC95%], 3,1-55,2%), con un número necesario a tratar de 8 (IC95%, 4-72).

Pero para tomar la decisión de incorporarla dentro de la cartera de servicios del centro de salud, hacía falta hacerse algunas otras preguntas, ¿Cuánto cuesta la moxibustión? ¿Cuántas cesáreas nos evitamos? ¿es costo-efectivo? ¿y si combinamos la demás evidencia de efectividad de otros estudios, llegaríamos a la misma conclusión? De todas estas preguntas, surgió la necesidad de completar este ensayo clínico con un estudio más amplio que planteara lo que después fue el objetivo del estudio: determinar si la moxibustión utilizada para corregir la posición no-cefálica es coste-efectiva para el sistema sanitarios público de Andalucía. Por tanto, la evaluación económica se realizó sobre la base de los datos de eficacia aportados por el ensayo clínico, pero también a partir de un metanálisis que

resumió la evidencia disponible en el momento sobre la técnica. De este modo, se evaluaron las posibles variaciones en los resultados atribuibles a diferentes medidas de eficacia clínica de la técnica de moxibustión.

Diseño del modelo para el análisis coste efectividad

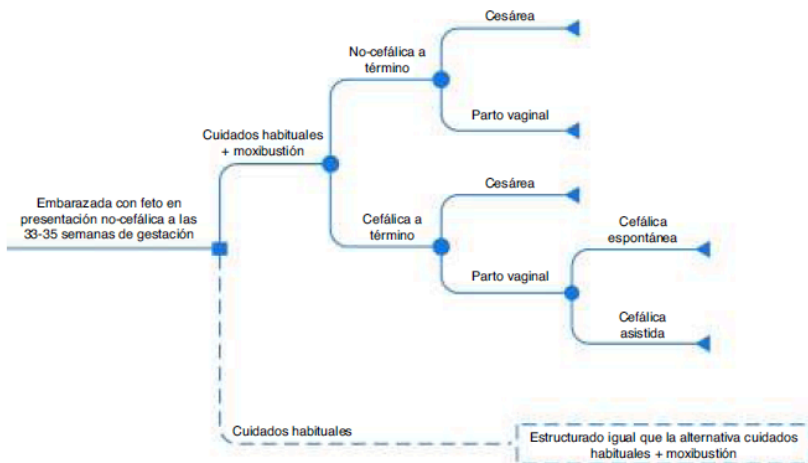
Se desarrolló un modelo de árbol de decisiones a corto plazo, con objeto de analizar la efectividad y el coste del tratamiento de la aplicación de calor por combustión de *A. vulgaris* (moxibustión) en el punto de acupuntura V 67, con el fin de facilitar la versión del feto y, por lo tanto, modificar la presentación no-cefálica en las 33-35 semanas de gestación (ver figura más adelante). El modelo se basó en una población de mujeres embarazadas que estaban una edad gestacional de 33-35 semanas, que presentaban biometría fetal normal y un feto en la posición no-cefálica, diagnosticado a las 32 semanas mediante exploración física y ecográfica.

Para esta población se presentaron 2 alternativas. La primera fue la de continuar con el tratamiento convencional establecido por el Servicio Andaluz de Salud para el embarazo, parto y puerperio, según el cual deben darse recomendaciones sobre las actividades cotidianas e instrucciones sobre la práctica diaria de la posición genupectoral; además, un examen ecográfico programado a la semana 38 y, en caso de que persista la presentación no-cefálica, programación de una cesárea a término. En la segunda alternativa se añadía, además de estas recomendaciones posturales, el tratamiento diario con moxibustión en el punto de acupuntura V 67 durante 2 semanas.

Como puede observarse, este modelo permite simular la realidad clínica de mujeres embarazadas con el feto en situación de nalgas en la semana 33-35 de gestación. La estructura refleja cómo la elección del tratamiento afecta el curso de la enfermedad, en este caso, la probabilidad de que el feto se dé la vuelta y se ponga en posición cefálica, y la relación de ésta con los costes y efectos (parto vaginal). El resultado final de este modelo es el ratio coste efectividad incremental (RCEI). El RCEI es el resultado que normalmente se muestra de una evaluación económica. Lo que varía de una modalidad de evaluación económica a otra, es la forma de medir el resultado (Drummond 2005). Una explicación más detallada se muestra en el siguiente apartado.

En este caso, el RCEI mide el coste de utilizar la moxibustion frente a la atención habitual por parto vaginal realizado. Cuando se consigue la versión fetal, esto incrementa el número de partos vaginales y reduce el número de cesáreas necesarias, con la terapia de moxibustión en comparación con el tratamiento habitual.

El análisis se realizó desde el punto de vista del sistema sanitario, lo que quiere decir que se incluyeron únicamente los costes directos soportados por el sistema. Los estudios de evaluación económica también contemplan la posibilidad de incluir los costes incurridos por los pacientes relacionados con la prestación del servicio sanitario, y así lo recomiendan las guías de evaluación económica (Bastida 2010, Cheers 2013), pero en este caso no se incluyó al ser poco relevante para el resultado del RCEI.



Modelo de árbol de decisiones para el tratamiento con moxibustión en las mujeres embarazadas con feto en una posición no-cefálica y una edad de gestación de 33-35 semanas.

Datos sobre efectividad y seguridad clínica

En evaluación económica hay distintas modalidades; coste/utilidad y coste/beneficio, coste/efectividad y minimización de costes (Drummond 2005). En un análisis de minimización de costes se asume que los resultados de las intervenciones que se están comparando son idénticos y por lo tanto

no se tienen en cuenta, y solo se comparan los costes. Las demás modalidades se diferencian por la forma de medir el resultado. Así, un análisis coste-utilidad mide el resultado en años de vida ajustado por calidad (AVAC), el análisis coste beneficio se caracteriza por medir el resultado en unidades monetarias, y el análisis coste efectividad mide el resultado en unidades naturales, es decir, unidades de resultado clínicamente relevantes y que tienen una relación directa con la intervención sujeta a evaluación. Por ejemplo, años de vida, o como en este caso (para la evaluar la moxibustión), partos vaginales realizados. Se consideró que ésta era la mejor forma de medir el resultado ya que lo que pretende el tratamiento de moxibustión es incrementar el número de partos vaginales realizados en situaciones de feto en posición no cefálica. Una consecuencia directa del incremento de partos vaginales es la reducción de cesáreas, pero este resultado se contemplará en el coste total del proceso, que se explica con más detalle en el apartado siguiente.

La principal fuente de datos para el análisis basal fue el estudio multicéntrico y aleatorizado llevado a cabo en el ámbito de la atención primaria de salud (VAS 2013). Incluyó mujeres embarazadas con una edad gestacional de 33-35 semanas, que fueron asignadas a una de las siguientes 3 opciones de tratamiento: moxibustión verdadera en el punto de acupuntura V 67 más el tratamiento habitual; moxibustión en un punto de acupuntura no específico (moxibustión simulada) más el tratamiento habitual, o tratamiento habitual solo. Todos los análisis se realizaron por intención de tratar.

Como fuente secundaria de información se pensó incluir una síntesis de la evidencia disponible, por lo que se realizó una búsqueda de revisiones sistemáticas que incluyeran estudios relacionados con la eficacia de la moxibustión en la promoción de la versión del feto a partir de una posición no-cefálica en comparación con la práctica habitual de recomendaciones posturales. La búsqueda se restringió a las revisiones publicadas en inglés o español desde 2004. Se consultaron las bases de datos de MEDLINE, Web of Knowledge, la Biblioteca Cochrane y el Centro de Revisiones y Disseminación. Se utilizaron los siguientes términos de búsqueda libres y con operadores booleanos: moxibustión, *Artemisia vulgaris*, acupuntura, moxa, investigación sobre efectividad comparativa, medidas de resultado, presentación de nalgas y posición de nalgas. Se encontraron 3 revisiones que describían 3 estudios que evaluaron la efectividad de la moxibustión en la provocación de la versión del feto a partir de una posición no-cefálica en comparación con

el tratamiento habitual (Vas 2009, Coyle 2009, Zhang 2013). Sobre la base de los datos de estos estudios y los reportados por el ECA, se realizó un metaanálisis para resumir toda la evidencia disponible y evaluar las posibles variaciones en la efectividad alcanzada en los diferentes estudios. El metaanálisis se realizó por el método de efectos aleatorios, dado que se detectó una considerable heterogeneidad entre los estudios. Esta heterogeneidad se identificó y se midió mediante una inspección visual inicial del pequeño grado de solapamiento entre los IC95% de los diferentes estudios. La heterogeneidad estadística se evaluó mediante los estadísticos X^2 e I^2 . La heterogeneidad se consideró alta si $I^2 > 50\%$. Los resultados del metaanálisis se presentaron como efecto promedio del tratamiento (con IC95%) y las estimaciones de I^2 .

En lo que se refiere a los datos de seguridad, dado que se considera una técnica segura, en el análisis de decisión no se tuvieron en cuenta los posibles efectos adversos resultantes del tratamiento de moxibustión.

Los costes incluidos en el modelo

Los costes incluidos fueron los relacionados con el Sistema Sanitario Público de Andalucía y la intervención que se está evaluando. Según el modelo de decisión ya definido, decidir qué costes incluir es sencillo, pues no hay más que seguir cuál es todo el proceso de intervención (mosxibustion+ asistencia habitual), la alternativa de comparación (asistencia habitual) y las consecuencias posteriores (parto vaginal en sus modalidades o cesárea).

Aquellos costes que son similares en ambas opciones de tratamiento (la asistencia habitual) no fueron tomados en cuenta. El análisis de coste efectividad es un análisis de comparación de costes y resultados en salud, por lo que no tiene sentido comparar costes que son iguales en ambas alternativas (el resultado sería 0). Sí se tuvieron en cuenta todos los costes relacionados con el tratamiento de moxibustión y los del parto, dependiendo del tipo (parto vaginal espontáneo, parto vaginal asistido o cesárea).

El coste del tratamiento con moxibustión se calculó asumiendo un precio de 13,90 € (20 unidades/puros) para obtener los 14 puros de artemisia necesarios para el tratamiento de 2 semanas. También se incluyeron el coste

de la primera sesión de moxibustión en el centro de salud que proporcionaba la formación necesaria antes de continuar el tratamiento en casa, y la consulta de seguimiento con la matrona del centro de salud a final del programa de tratamiento de 2 semanas.

Los costes correspondientes al parto y cesáreas, así como los de las visitas al centro de salud se tomaron de acuerdo con las tarifas publicadas en el Boletín Oficial de la Junta de Andalucía (BOJA 2005). Todas las fuentes consultadas para los se actualizaron para los valores de 2013 (año de realización del estudio de evaluación económica) de acuerdo con el índice de precios al consumo.

Evaluación de la incertidumbre del modelo

En una evaluación económica es muy importante realizar análisis de sensibilidad. Estos análisis evalúan la incertidumbre y permiten comprobar así el grado de estabilidad de los resultados y hasta qué punto pueden verse éstos afectados cuando se modifican algunas de las variables principales.

Para el estudio se realizaron varios análisis de sensibilidad deterministas, en los que se modificaron 2 parámetros variables. Uno de ellos fue la proporción de cesáreas llevadas a cabo debido a la ocurrencia de presentación no-cefálica entre el grupo de tratamiento con moxibustión, para hacer coincidir este porcentaje con el del grupo de tratamiento habitual. Este análisis fue necesario porque en el ECA la tasa de cesáreas realizadas en mujeres cuyos fetos se encontraba en la posición cefálica fue mayor en el grupo de moxibustión (19%) que en el grupo de tratamiento habitual (13%), sin que hubiera una explicación al respecto. En principio, no debería haber habido esta diferencia, como se comenta más adelante. Por lo tanto, se realizó un análisis de sensibilidad en el que para los 2 grupos se estableció la tasa de cesáreas en el 13%, es decir, la del tratamiento habitual. El segundo parámetro que se modificó fue la efectividad del tratamiento con moxibustión en la inducción de la versión fetal según el resultado de nuestro metaanálisis.

Además, se realizó un análisis de sensibilidad probabilístico, como recomiendan los métodos y guía de Evaluación Económica (Husereau 2013, WHO-CHOICE). Este tipo de análisis de sensibilidad es el enfoque habitual

para la estimación de la incertidumbre en los modelos. En este caso, las variables del modelo se consideran variables aleatorias compuestas por un valor medio y un término de error estocástico. De esta manera, se realiza una repetición de los análisis utilizando la variación de los “inputs” teniendo en cuenta ese término de error. La salida del modelo probabilístico se puede utilizar para construir intervalos de confianza para los costes y en los AVAC o para construir “curvas de aceptabilidad coste-efectividad”. Para nuestro estudio, se realizaron

1.000 iteraciones para reflejar la incertidumbre general en las entradas al modelo. En esta simulación a la probabilidad de un parto vaginal se le asignó una distribución beta y al riesgo relativo se le asignó una distribución log-normal. Como los datos de coste se obtuvieron de tarifas y, por lo tanto, no sujetos a incertidumbre muestral, no fue posible asignarles distribuciones a estos parámetros, y no se incluyeron para el probabilístico.

Principales resultados y análisis de sensibilidad

Nuestro análisis mostró un coste medio del tratamiento con moxibustión de 65,28 €. El tratamiento con moxibustión en las mujeres embarazadas cuyo feto se encuentra en presentación no-cefálica a las 33-35 semanas de gestación previene el 8,92% de los partos de nalgas en comparación con el tratamiento convencional. Esto equivale a un ahorro de coste medio de 107,11 € por parto, debido principalmente a la disminución del número de cesáreas requeridas. Estos resultados sugieren que el tratamiento con moxibustión es una alternativa eficiente, tiene menor coste y mayor efectividad que el tratamiento convencional, por lo que implantarla dentro de la cartera de servicios del centro de salud supondría un ahorro para el sistema sanitario, al mismo tiempo que una mejora en la salud de las embarazadas.

Los análisis de sensibilidad mostraron que la superioridad de la moxibustion seguía siendo evidente. El metanálisis mostro un RR de 0,34, lo que suponía un ahorro de costes de 414 euros por realizar la moxibustión en lugar de los cuidados habituales. Por otro lado, asumiendo el RR del análisis basal, pero disminuyendo la tasa de cesárea a valores similares al tratamiento habitual, el coste promedio ahorrado por parto sería de 240,28 €, con un 12% menos de cesáreas que con el tratamiento convencional. Y, por último,

el análisis de sensibilidad probabilístico mostro una probabilidad de que la moxibustión fuera costo-efectiva sin tener que pagar ningún euro adicional del 95%-99%.

Análisis reflexivo y conclusión

El análisis presenta dos puntos fuertes. En primer lugar, se recogió toda la evidencia que había disponible en torno a la efectividad de la técnica, variable fundamental para el análisis. Los estudios que se tuvieron en cuenta en la síntesis de evidencia, fueron evaluados por su calidad en la revisión Cochrane. En segundo lugar, la realización del modelo permite poder ajustar los datos a otros contextos en los que los costes son diferente. No obstante, es importante tener en cuenta que el hecho de que el coste del tratamiento de la moxibustión sea tan poco en relación a lo que cuesta una cesárea, hace que los resultados sean bastantes generalizables a otros contextos. Sólo con que la moxibustión evite una cesárea ya sería costo efectivo. En nuestro estudio el coste de la cesárea fue de 4.035 euros.

Una de las limitaciones es que no se tuvo en cuenta el coste del curso de entrenamiento de las matronas. El curso tuvo una duración de 10 horas impartido por un acupuntor experimentado y teniendo en cuenta únicamente el coste del personal sanitario y docente, podría estimarse en 700 euros cada 20 matronas. A nuestro juicio, el coste atribuible a cada mujer que se le practicara el tratamiento con moxibustión sería despreciable. De cara al cálculo de costes y la viabilidad es interesante que la técnica muestre su efectividad siendo autoadministrada y sin necesidad de ser supervisada por acupuntores, en general más escasos y con coste más alto que el de las matronas. Otra limitación está en la alternativa de comparación, una evaluación económica más completa para este análisis de decisiones, podría incluir una comparación con la versión cefálica externa a término, dada la efectividad demostrada de esta técnica para colocar al feto en posición cefálica. El hecho de no haberla incluido en este estudio fue por no disponer de ECAS en el que se comparen directamente esta técnica con la moxibustión. No obstante, según la evidencia que existe por el momento, merece la pena poner el foco en estas técnicas que con poca inversión pueden evitar los graves inconvenientes que supone la cesarea.

Por otro lado, señalar que llama la atención que la incidencia de cesáreas en presentaciones cefálicas a término en el grupo de moxibustión del ECA fue notablemente mayor entre el grupo moxibustión (19%) que en el grupo control (13,3%). Resulta difícil imaginar cualquier vínculo fisiológico entre la versión provocada por el tratamiento con moxibustión y un mayor riesgo de cesárea. No se encontró ningún efecto secundario que pudiera explicar esto, ya sea en el estudio o en la bibliografía. No tenemos ninguna explicación para este hecho, pero una hipótesis posible que debería ser probada en futuros estudios futuros proviene del marco conceptual de la economía del comportamiento a través de la llamada ceguera inducida por la teoría 21. Este trastorno significa que, una vez que aceptamos una teoría y la usamos como herramienta en nuestra forma de pensar, es extraordinariamente difícil darnos cuenta de sus defectos, confiando en la comunidad de expertos que ya la han aceptado. Esta hipótesis puede fortalecerse aún más si tenemos en cuenta el agotamiento cognitivo que implica la evaluación de una decisión ya tomada con anterioridad.

Por último, nuestra conclusión a raíz de la evidencia y los datos de coste, es que La aplicación de moxibustión durante 14 días en el punto de acupuntura V 67 a mujeres embarazadas con presentación no-cefálica durante las semanas 33-35 es coste-efectiva, reduciendo la necesidad de cesáreas por esta causa en un 9% en comparación con el tratamiento habitual.

Referencias clave

Drummond MF, Sculpher MJ, Torrance G, O'Brien J, Stoddart GL (2005). *Methods for the economic evaluation of health care programmes*. (3rd ed). Oxford: Oxford University Press.

Este libro incluye un detalle de los métodos para medir los resultados y costes en una evaluación económica. Como realizar modelos y evaluar la incertidumbre de los resultados.

WHO-CHOICE. (s.d). Quantities and unit prices (cost inputs)
<http://www.who.int/choice/cost-effectiveness/inputs/en/>

El sitio web del proyecto CHOICE presenta los métodos e instrumentos desarrollados por la OMS para las evaluaciones económicas.

Sacristán JA, Oliva J, Campillo-Artero C, Puig-Junoy J, Pinto-Prades JL, Dilla T, Rubio-Terrés C, Ortún V. (2020) *Gac Sanit.* Mar – Apr;34(2):189-193.

En este artículo se da respuesta al creciente interés por fijar el precio de las nuevas tecnologías en función del valor que estas proporcionan a los sistemas de salud. La interpretación de sus resultados del coste-efectividad requiere establecer unos valores de referencia que sirvan de guía sobre lo que constituye un valor razonable para el sistema sanitario. Los umbrales de eficiencia en España según este trabajo podrían estar comprendidos entre los 25.000 y los 60.000 euros por año de vida ajustado por calidad.

Referencias

Organización Mundial de la Salud (2010). Informe sobre la salud en el mundo 2010: Financiación de los sistemas de salud: el camino hacia la cobertura universal.

<http://www.who.int/whr/2010/es/index.html>.

González López-Valcárcel B (2007). La incorporación de nuevas tecnologías en el Sistema Nacional de Salud. Coste-efectividad y presiones sobre el gasto sanitario. *Revista Presupuesto y Gasto Público* 49/2007: 87-105.

Arrow KJ (1963). Uncertainty and the Welfare Economics of Medical Care. *American Economic Review*, 53(5); 941-73.

Bermúdez-Tamayo C, Johri M, Chaillet N. Budget impact of a program for safely reducing caesarean sections in Canada. *Midwifery*. 2018 May;60:20-26.

Cruikshank DP (1986). Breech presentation. *Clin Obstet Gynecol*, 29:255-63.

Chaillet N, Dumont A, Abrahamowicz M, Pasquier JC, Audibert F, Monnier P, Abenhaim HA, Dubé E, Dugas M, Burne R, Fraser WD; QUARISMA Trial Research Group (2015) A cluster-randomized trial to reduce cesarean delivery rates in Quebec. *N Engl J Med*. Apr 30;372(18):1710-21.

Hofmeyr GJ, Kulier R (1996). External cephalic version for breech presentation at term. *Cochrane Database Syst Rev*.

Coyle ME, Smith CA, Peat B (2012). Cephalic version by moxibustion for breech presentation. *Cochrane Database Syst Rev*. May 16;(5):CD003928.

- Vas J, Aranda-Regules JM, Modesto M, Ramos-Monserrat M, Barón M, Aguilar I (2013). Using moxibustion in primary healthcare to correct non-vertex presentation: a multicenter randomised controlled trial. *Acupuncture in Medicine*,31:31-8
- Vas J, Aranda JM, Nishishinya B, Mendez C, Martin MA, Pons J (2009). Correction of non-vertex presentation withm oxibustion: a systematic review and metaanalysis. *American Journal of Obstetrics & Gynecology*, 201:241-59.
- Zhang QH, Yue JH, Liu M, Sun ZR, Sun Q, Han C, (2013). Moxibustion for the correction of non-vertex presentation: a systematic review and meta-analysis of randomized controlled trials. *Evidence-Based Complementary and Alternative Medicine*,241027.
- López Bastida J, Oliva J, Antoñanzas F (2010). Propuesta de guía para la evaluación económica aplicada a las tecnologías sanitarias. *Gaceta Sanitaria*,24:154-70.
- González-González NL, Medina V, Jiménez A, Gómez Arias J, Ruano A, Perales A, et al. Base de datos perinatales nacionales 2004. *Prog Obstet Ginecol*. 2006;49:645-55.
- Orden de 14 de octubre de 2005, por la que se fijan los precios públicos de los servicios sanitarios prestados por Centros dependientes del Sistema Sanitario Público de Andalucía [consultado 23-8-2014].
<http://www.juntdeandalucia.es/boja/2005/210/28>
- Husereau D, Drummond M, Petrou S, (2013). Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. *European Journal of Health Economic*, 14:367-72.

Resumen / Résumé / Abstract

El método de coste efectividad ofrece una información potente para tomar decisiones sobre la inclusión de nuevas intervenciones en los sistemas sanitarios. Presentamos como ejemplo la evaluación de la eficiencia de un programa para incrementar el número de partos vaginales realizados en situaciones de feto en posición no cefálica. En este ejemplo se describe la importancia de un buen análisis de la evidencia disponible para garantizar

la robustez de los modelos utilizados, así como la suficiente descripción del análisis y parámetros que permita que se pueda ajustar a otros contextos con datos de entrada diferentes.

La méthode coût-efficacité offre des informations puissantes pour la prise de décision lorsqu'il s'agit d'inclure de nouvelles interventions dans les systèmes de santé. Nous présentons à titre d'exemple l'évaluation de l'efficacité d'un programme visant à augmenter le nombre d'accouchements par voie vaginale effectués dans des situations fœtales non céphaliques. Cet exemple décrit l'importance d'une bonne analyse des preuves disponibles pour garantir la robustesse des modèles utilisés, ainsi qu'une description suffisante de l'analyse et des paramètres pour permettre son ajustement à d'autres contextes avec des données d'entrée différentes.

The cost-effectiveness method provides powerful information for making decisions about the inclusion of new interventions by health systems. We present as an example the evaluation of the efficiency of a program to increase the number of vaginal deliveries performed in non-cephalic fetal situations. This example describes the importance of an appropriate analysis of the available evidence to ensure the robustness of the models used, as well as a sufficient description of the analysis and used parameters to allow it to be adapted to other contexts with different input data.

Leticia García Mochón es investigadora de la Escuela Andaluza de Salud Pública, desarrolla su actividad principalmente en economía de la salud, gestión sanitaria, y la evaluación económica de tecnologías sanitarias. Desde que comenzó su carrera en 2006, ha colaborado en proyectos internacionales financiados por la Comisión Europea, la Organización Mundial de la Salud (OMS), y la Organización Panamericana de la Salud (OPS), entre otras instituciones, la mayoría relacionadas con proyectos relacionados con las temáticas mencionadas. Es autora de más de 50 publicaciones entre artículos científicos, libros y capítulos.

Clara Bermúdez-Tamayo es investigadora en la Escuela Andaluza de Salud Pública y en el CIBERESP (Ciber de Epidemiología y Salud Pública). Es Directora Asociada de la Revista Gaceta Sanitaria. Sus líneas de investigación

se centran en la implementación y evaluación de modelos organizativos y tecnologías sanitarias eficientes, considerando una perspectiva integradora que minimice las desigualdades en salud. Ha colaborado en proyectos y estrategias para organizaciones internacionales y ONG, como la OPS/OMS, National Institute for Health Research-NETS programmes (UK), Agence Française de Développement. Directora Adjunta de la revista Gaceta Sanitaria (revista no anglófona más importante en el campo de la salud pública).

Cita

Leticia García Mochón y Clara Bermúdez-Tamayo (2020). Análisis coste efectividad. Evaluación de la eficiencia de un programa para incrementar el número de partos vaginales realizados en situaciones de feto en posición no cefálica. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 365-380. Québec: Éditions science et bien commun y Marseille: IRD Éditions

15. Ensayo clínico aleatorizado por conglomerados

Diseño para evaluar la efectividad de una intervención en población desfavorecida para mejorar el autocuidado en diabetes adaptado a un contexto de restricción de recursos

EVA MARTÍN RUIZ Y ANTONIO OLRÍ DE LABRY LIMA

Ensayo clínico aleatorizado por conglomerados

Definición del método

Los ensayos aleatorizados por conglomerados son muy útiles para evaluar intervenciones en servicios de salud. En este tipo de ECA la asignación al grupo de intervención o control se realiza por conglomerados. La configuración de los grupos a estudio; control (que recibe intervención habitual o ninguna intervención-placebo-), e intervención (que recibe la intervención a evaluar) se realiza de manera aleatoria atendiendo, por ejemplo, a demarcaciones geográficas, áreas de influencia, o pertenecer a un colegio, centro de salud.

Fortalezas del método

- Resulta un diseño válido para la investigación en determinados contextos y situación, como por ejemplo en el ámbito de las intervenciones comunitarias, salud pública, o en contextos de limitación de recursos, por la agilidad y sencillez que aporta al proceso de definición y configuración de los grupos.
- Las limitaciones económicas, así como la naturaleza de determinadas intervenciones, junto a su buena adaptabilidad en el campo de la investigación en servicios de salud hace que en los últimos años se está experimentando un aumento de este tipo de diseño.
- Son considerados como eficientes desde el punto de vista administrativo.
- Es un diseño útil cuando se considera que hay riesgo de contaminación o influencia entre las intervenciones realizadas en las dos ramas del ensayo.

Desafíos del método

- Requiere de un tratamiento estadístico particular tanto para la configuración de los tamaños muestrales como para el análisis de los datos obtenidos.
- La posible variabilidad entre los grupos que forman parte del ensayo, puede ser mayor que la existente dentro del grupo, por lo que se requiere de un mayor tamaño muestral que en otros ensayos clínicos.
- Existe cierto debate en torno a aspectos éticos de este diseño, ya que no siempre se solicita el consentimiento informado al conjunto de participantes en la investigación.

Los ensayos clínicos aleatorizados por conglomerados (ECA-C) para la evaluación en Servicios de Salud

La investigación en servicios de salud tiene como fin último mejorar la efectividad y eficiencia de los propios servicios implementando y evaluando para ello, mejoras en la organización, desempeño y planificación de dichos servicios, entre otros aspectos. La investigación en este campo no se centra estrictamente en lo que podríamos llamar áreas clínicas o asistenciales, sino que va más allá, incluyendo áreas de gestión y de provisión.

Los servicios sanitarios se caracterizan por su gran complejidad de organización, que tiene repercusiones en su gestión clínica, toma de decisiones y todo ello con un elevado grado de incertidumbre. Además, dichos servicios están expuestos a constantes cambios, a los que diariamente les supone enfrentarse a innumerables retos y, situaciones nuevas, no siempre propias, a las que es necesario adaptarse. Entre los diferentes retos destacan aspectos ligados a la universalidad y accesibilidad de la atención sanitaria, los cambios en las dinámicas poblacionales y sociales, el progresivo envejecimiento de la población, la incorporación la orientación comunitaria, social y familiar, el alto nivel resolución, así como la sostenibilidad del sistema.

Así, teniendo en cuenta este ámbito tan complejo, la investigación en servicios de salud tiene como objetivo responder a preguntas sobre el mejor tratamiento disponible para un determinado problema de salud, el mejor método diagnóstico, así como la provisión de la atención sanitaria requerida de la manera más eficiente, para la consecución de mejores resultados en la salud y de calidad de vida de las personas.

Las investigaciones de evaluación sobre elementos ligados más a la gestión y los procesos asociados a la misma en los servicios de salud, son limitadas y por lo general, muy circunscritas a contextos específicos donde es difícil la extrapolación de experiencias y de resultados a otros contextos. Es posiblemente uno de los mayores puntos débiles de la investigación en este campo. En las páginas que siguen, se mostrará el proceso seguido en el desarrollo de una investigación en Servicios de Salud a partir de un diseño epidemiológico concreto como es el ensayo aleatorizado por conglomerados, para evaluar la mejora de la salud y la calidad de vida de las personas con diabetes, a partir de una intervención de autogestión de la enfermedad, dirigida a personas con bajo nivel educativo.

Durante la planificación, gestión y desarrollo de este estudio de intervención, tuvo lugar la eclosión de una fortísima crisis económica, en la que los niveles de empleo se vieron especialmente reducidos, así como supuso una merma importante en torno a la capacidad económica de las familias y del estado. En cuanto a la capacidad de financiamiento de los servicios públicos, entre los que se encuentran los sanitarios, tuvieron lugar serias restricciones, por lo que se tuvieron que realizar cambios y ajustes en la planificación inicial de esta investigación, así como a la incorporación de importantes modificaciones en todas las etapas del estudio, que a continuación iremos detallando.

Diabetes como problema de salud a abordar

La diabetes mellitus (DM) es una enfermedad crónica que se presenta como un grupo heterogéneo de trastornos caracterizados por hiperglucemia e intolerancia a la glucosa. Hay tres tipos principales de diabetes: tipo 1, tipo 2 y gestacional. El tipo 1 (DM1) es causado por una reacción autoinmune que ataca las células beta pancreáticas, que son responsables de la producción de insulina. DM1 no se puede prevenir o curar, pero se pueden tomar medidas para evitar complicaciones y muerte prematura. En la diabetes tipo 2 (DM2), las células de los órganos diana desarrollan resistencia a la insulina. Además, su producción por las células beta pancreáticas es a menudo insuficiente. DM2 representa aproximadamente el 90% de todos los casos de DM. Sin embargo, ciertos factores de riesgo, como la obesidad, el consumo de bebidas azucaradas y la falta de actividad física, entre otros, pueden abordarse y modificarse para, retrasar o prevenir su aparición y/o complicaciones derivadas.

La DM presenta diversas complicaciones a nivel macro y microvasculares que tienen un alto impacto sobre el sistema sanitario. Es la cuarta causa de muerte más común en todo el mundo, llegando a producir unos cinco millones de muertes anuales en el planeta. Presenta a su vez, una elevada prevalencia: la cifra mundial se sitúa en torno a 8.8% en personas con edades comprendidas entre los 20 y 79 años, existiendo un alto número de casos infradiagnosticados, pudiendo llegar incluso a duplicar esta cifra. La Organización Mundial de la Salud (OMS) estima que para el año 2025, esta cifra de prevalencia habrá aumentado en un 42% en los países desarrollados y en un 170% en los países en desarrollo (OMS 2018).

Esta enfermedad no se distribuye por igual entre todos los grupos de población, ya que se ha observado una mayor incidencia, aparición de complicaciones y una mayor tasa de mortalidad entre las mujeres, las minorías étnicas, las personas con un bajo nivel socioeconómico y también en aquellos países de ingresos medios a bajos¹. La literatura científica señala que aquellas personas que padecen DM con un nivel socioeconómico bajo sufren un mayor retraso en el diagnóstico y son peores para controlar su condición (Danaei G et al 2011; Williams et al 2018). En esta población, que además cuentan con bajo nivel educativo, una buena comunicación entre el conjunto de profesionales sanitarios que le atienden es fundamental para obtener unos adecuados resultados en salud. Así, se requería de herramientas que permitiesen mejorar la comunicación entre el profesional sanitario y el paciente. Teniendo en cuenta las cifras ya comentadas de elevada prevalencia y su alto impacto económico, y que aún es mayor entre la población desfavorecida, se hace necesario desarrollar intervenciones, programas o tecnologías que frenen la progresión de la DM o mejoren la atención a las personas con diabetes, así como sus resultados en salud y calidad de vida asociados. Por ello, era pertinente trabajar en el desarrollo de una intervención que tuviera como objetivo final mejorar el autocuidado en aquellas personas con este diagnóstico y bajo nivel educativo.

Descripción de la intervención

El objetivo de la intervención a evaluar es la mejora del autocontrol de la diabetes en pacientes con un bajo nivel educativo y un control glucémico inadecuado mediante la mejora de la comunicación entre el paciente y el proveedor. El diseño de la intervención se efectuó a partir de una revisión de la literatura que permitió identificar las intervenciones con este propósito que ya estaban probadas como efectivas. Después de este análisis, el equipo de investigación se decidió por adaptar una intervención que tenía como base una herramienta desarrollada por Chapin et al (2003) consistente en un sistema de registro para monitorización de la diabetes del paciente en casa, denominado “Take-home Diabetes Record” (THDR), que el equipo tradujo por “Registro de diabetes para llevar a casa (RDPC)”. Esta herramienta

1. OMS. Informe mundial sobre la diabetes. Ginebra: OMS 2016 Disponible en <https://www.who.int/diabetes/global-report/es/>

consistía en registrar determinada información en formato determinado y de la manera más gráfica posible (por ello, se habla de herramienta visual), en relación al cumplimiento del tratamiento farmacológico y no farmacológico, así como en el registro gráfico de los valores de hemoglobina glicosilada (HbA1c) a lo largo del tiempo y relacionarlos con las actividades de autocontrol autoinformadas. Los resultados obtenidos por Chapin et al. mostraron que se trataba de una intervención efectiva: el 51% de los pacientes lograron un control glucémico adecuado (HbA1c <7%, 53.01 mmol/mol). Debido a su bajo costo de intervención y su fácil transferibilidad a la práctica clínica habitual, se decidió adaptar esta intervención e implementarla en un entorno de atención primaria y evaluar su impacto.

Los agentes de salud y los pacientes colaboran para completar el RDPC, que consta de dos partes: 1) un gráfico que muestra los niveles anteriores de HbA1c del paciente, y 2) cinco elementos comunicados por el paciente sobre las actividades de autocontrol. Al final de la sesión, se ofreció a los pacientes una copia del RDPC para llevar a casa y se les anima a que lo comenten con sus familiares. La intervención se realiza cada 3 meses, en el curso de un año (es decir, cuatro sesiones al final del período de estudio). Para estandarizar el estilo de intervención, los agentes recibieron entrenamiento en aspectos cognitivos, emocionales y de comunicación.

Un subgrupo de pacientes recibía refuerzo telefónico, que se centró específicamente en la actividad física y la alimentación sana y se basó en estrategias psicoeducativas y de resolución de problemas. Los refuerzos telefónicos fueron llevados a cabo por un miembro del equipo de investigación que había sido entrenado previamente en técnicas de cambio de comportamiento y habilidades de entrevista motivacional.

En el grupo control se proporcionaba una atención estándar, que incluía citas de seguimiento regulares cada 3 meses.

Procedimiento

Todas las etapas de este estudio (reclutamiento, mediciones de referencia, intervenciones y mediciones finales) se llevaron a cabo en las citas trimestrales de rutina. Este enfoque se utilizó para minimizar el costo del estudio y reducir la carga para los pacientes.

Medidas demográficas y de resultados: los participantes informaron por sí mismos su sexo, fecha de nacimiento, origen étnico, número de hijos, nivel educativo y apoyo social, que se midieron mediante un instrumento validado. La variable de resultados primaria fue la HbA1c; como resultados secundarios se examinaron el colesterol LDL y HDL, los triglicéridos, la presión arterial, el índice de masa corporal (IMC) y la circunferencia abdominal. Las medidas de resultado se recogieron al inicio, cada tres meses durante el período de intervención y después de la intervención. Todos los ensayos bioquímicos se ajustaron a las directrices nacionales españolas.

Inicialmente, como se menciona, en el protocolo de investigación se definió que el desarrollo de esta intervención sería desarrollada por agentes de salud, donde el papel del conjunto de profesionales de la salud del centro en el que se desarrollaba este trabajo, desarrollaría el papel de “captadores” de participantes, es decir, de las personas encargadas de la primera toma de contacto con las personas susceptibles a participar en esta investigación. Así, su papel a desarrollar consistía en la colaboración para el desarrollo de las tareas de selección y reclutamiento de pacientes que cumplieran el perfil requerido.

El diseño epidemiológico para la evaluación

Al consistir nuestra investigación en la evaluación de los resultados de la implementación de una intervención concreta, el diseño más adecuado con el de un estudio experimental o ensayo clínico. Por otra parte, al tratarse de un estudio bajo el paradigma de investigación en servicios de salud, el diseño del ensayo se realizó teniendo en cuenta los estándares y situaciones de la práctica clínica habitual en el centro de salud de atención primaria donde se iba a realizar dicha intervención, para así obtener una efectividad lo más cercana posible a las condiciones reales y por consiguiente de un ensayo clínico más pragmático. Así, este tipo de estudio son de corte pragmático donde se evalúa la efectividad de una intervención y no la eficacia.

Entre las diferentes familias de diseños de estudios experimentales, se decidió escoger el ensayo clínico por cluster o por conglomerados. Una de las opciones previamente valoradas fue la de adoptar un diseño cuasiexperimental aunque pudiera resultar una alternativa eficiente, se descartó esta opción por considerarse, aunque veremos más adelante que

se tuvieron que hacer ciertas modificaciones. En cuanto al grupo control, había que tener en cuenta la naturaleza de la intervención, la ausencia de información previa, la imposibilidad de controlar otros posibles factores de confusión y que era factible tanto logísticamente como por número de pacientes a incluir, finalmente se decidió por incluir un grupo control en el estudio.

En todo ensayo clínico, debemos asegurar la comparabilidad entre grupos, por ello, cualquier estudio que disponga de grupo de comparación, éste debe ser lo más similar posible al grupo intervención. De esta manera, a la hora de evaluar los resultados, si el grupo intervención y el grupo control no fuesen similares, pudiese generarse la duda de que diferencias en torno a las medidas de resultados evaluadas puedan ser debidas a la intervención desarrollada o por estos factores diferenciadores del grupo. Además, la diabetes al ser una enfermedad con alto impacto comunitario puede estar sometido a frecuentes estrategias o iniciativas para mejorar la salud de esta población, hace necesario disponer de un grupo control para mitigar o reducir esas posibles influencias.

Por otro lado, debe utilizarse un método que permita la asignación de los pacientes a los dos grupos del ensayo clínico y que este método permita tener ambos grupos lo más similares posible.

En este trabajo se decidió realizar un ensayo clínico por conglomerados, definiéndose el cluster como la consulta del profesional sanitario de atención primaria. Para la asignación de los participantes a los grupos del ensayo clínico, y con la finalidad de minimizar el **sesgo de selección**, dicha asignación se realizó de manera cegada y aleatorizada por personal independiente, con las consultas del profesional sin identificar, por tanto, anónimas, tanto por parte del profesional de dicha consulta como por los pacientes allí atendidos.

El siguiente sesgo a abordar fue el **sesgo de realización**, que puede producirse por dos factores. El primero de ellos es por un inadecuado cegamiento de los grupos de participantes en el estudio y el otro por fallos en el cegamiento del personal. El sesgo de realización se produce a causa del conocimiento de las intervenciones asignadas por parte de los participantes y/o del personal durante el estudio. Dependiente del tipo y naturaleza de la intervención a realizar, hay veces que no es posible cegar a los participantes en el estudio, a los profesionales que participan y/o a la persona encargada

del análisis de la información. Es importante minimizar el riesgo de ejecutar este sesgo, ya que puede influir de un modo u en otro en la alteración de los resultados del estudio por diferentes motivos:

- Diferentes tratamientos en el grupo intervención.
- Diferentes expectativas de los participantes.
- Conduce a cambios en los resultados reales.

En el estudio desarrollado, dada la naturaleza de la intervención, el cegamiento no fue posible ya que los participantes conocían si recibían la intervención, por tanto, se asume que se corre ese riesgo a sesgo y que es imposible su minimización y control.

El estudio se monitorizó activamente por investigadores internos y externos para garantizar la calidad de los datos.

Modificaciones al diseño de la intervención

En el momento de puesta en marcha de este estudio, el planteamiento inicial tuvo que ser modificado cobrando los profesionales del centro colaboradores, un papel más protagonista. Se decidió que la intervención fuese llevada a cabo por los propios profesionales sanitarios del centro de salud y esta decisión adoptó por diferentes motivos. En primer lugar, la crisis económica tuvo un negativo impacto sobre la realización del estudio, ya que limitó la cuantía de la financiación para realizar este estudio. Aunque inicialmente, el desarrollo de la intervención iba a ser desarrollada por agentes de salud que apoyaban a los médicos del propio centro de salud en el que se realizaba el estudio y contratos *ad hoc*, la reducción de financiación conllevó a una reestructuración y replanteamiento, que conllevó finalmente a que se acordara que la intervención a estudio sería desarrollada por los propios profesionales sanitarios del centro. Esto suponía cambios en su carga de trabajo, así como la adaptación de los modelos de análisis previos.

Antes de adoptar esta decisión, se realizó una estimación del riesgo de cometer un posible **sesgo de selección**. Así, se debatió con los profesionales sanitarios del centro de salud, y otras personas con experiencia en investigación en servicios de salud. Con la información recabada se decidió acometer las siguientes actividades a fin de minimizar este riesgo:

- La realización de actividades formativas de profesionales del centro de salud en el que se iba a desarrollar la intervención, con el objetivo último de homogeneizar el nivel de partida en cuanto a conocimientos e información, así como de establecer un protocolo común y estandarizado para el proceso de captación de participantes y la realización de la intervención.
- Desarrollo de entrevistas motivadoras con profesionales participantes, donde se ahondó en la relevancia del desarrollo de este tipo de investigaciones, así como los posibles beneficios a obtener en términos de resultados para la salud de las personas a las que tienden.
- Se les proporcionó retroalimentación sobre la información de pacientes en el estudio.

Señalar también que el equipo de investigación realizaba visitas regulares al centro de salud para solventar dudas y posibles incidencias que puedan darse en el transcurso de la investigación, así como por mantener una comunicación fluida y continua con el grupo de profesionales implicados en el proyecto sobre aspectos ligados a la captación de participantes, desarrollo de la intervención, recopilación de datos, entre otros.

Esta modificación incorporada al protocolo de investigación, en la puesta en marcha del estudio de intervención, conlleva una serie de ventajas aparejadas: menor interferencia en la dinámica habitual del centro y de sus profesionales, posiblemente mayor cercanía dado el conocimiento y relación previa entre profesional y paciente, así como una mayor información recopilada, es decir, se pudieron recoger un número mayor de variables.

El hecho de que la intervención fuese realizada por los propios profesionales del centro de salud, obligaba a simplificar parte del trabajo de estas personas durante el desarrollo de la intervención en cuanto a la recogida de información, por ello, se redujo el número de variables a recoger para que el tiempo a emplear en fuera el menor posible. Además, parte del equipo de investigación se encargó de facilitar al máximo la planificación y gestión del trabajo de campo: realizando las tareas relacionadas con la programación y gestión de citas, facilitar materiales, y todas otras que pudieran ser desarrolladas por personal externo al centro, y así agilizar en la medida de lo posible la carga del trabajo.

Para esta investigación, el clúster o conglomerado quedó definido por la consulta de cada profesional participante. Esta decisión fue adoptada en el seno del equipo con el propósito de facilitar en lo posible, la tarea del profesional en el proyecto ya que presenta una gestión más fácil del proyecto, así como la ventaja de minimizar el posible efecto de “contaminación” o posibles efectos cruzados de la intervención en el grupo control y/o de participantes. En relación a la contaminación, es decir, que los participantes en el estudio experimentan interacciones entre ellos que pueden influir en la efectividad de la intervención, siendo por ello más adecuado que el ensayo clínico tradicional. Concretamente, en el ejemplo que estamos desarrollando, diferentes elementos para mejorar el autocuidado de los pacientes respecto a la diabetes es muy probable que los pacientes hablen entre ellos de esta intervención produciéndose una modificación de su conducta y hábitos de vida que pueden influir en la variable dependiente del estudio. Dicha interacción haría inadecuado optar por un ensayo clínico con aleatorización de individuos, pues el riesgo de contaminación es muy elevado, siendo más conveniente en este caso optar por un diseño en conglomerados.

Hay que tener en cuenta que este estudio se trata de un ensayo clínico pragmático, siendo el objetivo de esta intervención mejorar la comunicación entre el personal facultativo y pacientes, por ello introducir un elemento extraño nos alejaría del pragmatismo buscado. Así, el modo de proceder seguido, y el no alejarnos de la propia infraestructura del centro y los profesionales del centro de salud en el que se desarrollaba el estudio, probablemente los pacientes vivieron el desarrollo de esta intervención como parte de la atención habitual.

Sesgo de desgaste

El contexto en el que tuvo lugar esta investigación como se ha comentado anteriormente, no fue el más facilitador o propicio para el desarrollo de intervenciones de este tipo y desarrolladas por el personal de los centros de salud. La crisis económica en España produjo un incremento de la desigualdad en la salud de la población, cambios en el acceso al sistema sanitario, el incremento del copago, que han afectado a la salud de la población. Entre las medidas adoptadas en materia de ajustes

presupuestarios y de organización de servicios sanitarios, algunas de ellas afectaron directamente al personal público sanitario: reducción salarial del 20% junto al aumento de la jornada laboral, el bloqueo de la carrera profesional, la limitación o supresión de las sustituciones, la pérdida de beneficios sociales y la reducción del número de días de descanso.

En este contexto adverso, la carga de trabajo del personal se vio incrementada, lo que pudo influir en los procedimientos de selección y captación del conjunto de participantes en el estudio, así como en el seguimiento y monitorización de los mismos. El abandono de participantes en el estudio pudo verse incrementado.

La retirada antes de la finalización o las pérdidas en el seguimiento de los participantes en el estudio se maneja de manera similar a cuando estos hechos se producen en un ensayo clínico aleatorizado: utilizando un análisis estadístico *ad hoc* para datos ausentes.

Como medida adoptada para la minimización de este sesgo, ya se ha comentado anteriormente el contacto periódico entre el equipo de investigación y el equipo de profesionales colaboradores del centro de salud a través de visitas regulares.

Análisis y principales resultados

Debido a la naturaleza jerárquica de los datos y a las características de los datos de panel, se utilizaron modelos de efectos mixtos de dos niveles (paciente y proveedor) para examinar si los dos grupos de pacientes tenían tasas de cambio diferenciales a lo largo del tiempo para la variable respuesta (HbA1c) y los resultados secundarios (es decir, tiempo por interacciones de grupo). Este análisis también se repitió para otras variables (HDL-c (mg/dL), LDL-c (mg/dL), triglicéridos (mg/dL), presión arterial sistólica (mmHg), presión arterial diastólica (mmHg), índice de masa corporal y circunferencia abdominal (cm), que también se controlaron para la comorbilidad, el sexo y el apoyo social. Por último, para evaluar un posible efecto de la dosis, se examinó el número de sesiones asistidas como predictor de la HbA1c. El nivel de significación se fijó en el 5% ($p < 0,05$).

Los datos faltantes (nada despreciables por las ausencias en el seguimiento de algunos pacientes) se evaluaron examinando los patrones de las principales variables. No se encontró ningún patrón, por lo que

concluimos que los datos que faltaban podían ser aleatorios y utilizamos el procedimiento de “última observación llevada a cabo”, suponiendo que la última observación es representativa del valor que falta. Los individuos con valores perdidos no mejoraron ni se deterioraron.

Nuestro enfoque analítico hizo uso de todos los datos disponibles, haciendo de este un análisis de intención de tratar. Se utilizó la estimación de máxima probabilidad de información completa. Esta forma de estimación proporciona estimaciones no sesgadas bajo el supuesto menos restrictivo de que los datos falten al azar, lo que permite que la probabilidad de que los datos falten dependa del resultado y de los predictores. Para tener más en cuenta los datos que faltan, se realizaron análisis adicionales en los que se investigó el carácter informativo de los patrones de datos que faltan como posibles factores de confusión. Como no se registró un abandono formal, tratamos los datos posteriores al tratamiento que faltaban debido al abandono y los datos que faltaban por otras razones como equivalentes.

Los pacientes tanto del grupo de intervención como del grupo de control, mejoraron significativamente su control glucémico después del período de estudio. Sin embargo, se observó una mayor reducción de la HbA1c en el grupo de intervención. La intervención produjo un mayor impacto positivo en los pacientes con mayores niveles de apoyo social y sin comorbilidades. Se observó un efecto de la dosis, ya que cada sesión adicional se asoció a una disminución del 0,2% de HbA1c.

Análisis reflexivo y conclusión

En este capítulo hemos visto el valor de realizar un ECA por conglomerados en un contexto de limitación de recursos. Por razones logísticas y organizativas resultaba el mejor diseño. El estudio presenta varios puntos fuertes metodológicos, a saber, un seguimiento adecuado de las medidas de resultados a lo largo del período de estudio (que incluye no sólo medidas de referencia y de resultados finales, sino también cada tres meses). Además, la asignación al azar se realizó a nivel de los médicos de cabecera (y no de los pacientes), lo que evitó un posible sesgo de contaminación. Además, la posible variabilidad entre los médicos generalistas (efecto de agrupación) se controló adecuadamente en nuestro análisis multinivel. Las reducciones de HbA1c fueron más notables después

de seis meses de seguimiento, estos resultados podrían justificarse por la necesidad de un número mínimo de sesiones para lograr un efecto de intervención, aunque no se pudo excluir el sesgo de rendimiento, porque las medidas de HbA1c se realizaron cada 3 meses, en lugar de cada 6 meses como la atención habitual. Además, cada seis meses se celebró una reunión conjunta de investigadores y médicos de cabecera para abordar los posibles problemas, supervisar el progreso del estudio y asegurar el cumplimiento del protocolo de estudio.

Sin embargo, el estudio también tiene varias limitaciones. En primer lugar, aunque la proporción de pacientes en el grupo de intervención y en el de control era similar, el número de proveedores en el grupo de intervención (seis médicos generales) era mayor que en el grupo de control (con sólo tres médicos generales). Sin embargo, los resultados del análisis de niveles múltiples sugirieron que la variabilidad atribuible al nivel de proveedores (efecto de grupo) no era significativa, y por lo tanto es poco probable que el número desequilibrado de proveedores por grupo haya sesgado nuestros resultados. En tercer lugar, no fue posible llevar a cabo un estudio “doble ciego” porque los propios médicos de cabecera eran los responsables de realizar la intervención. Por último, no podemos descartar la posibilidad de que los médicos de cabecera y/o los pacientes hayan modificado su comportamiento habitual durante el período de estudio porque sabían que estaban siendo estudiados (efecto Hawthorn), lo que explicaría, al menos parcialmente, los cambios positivos observados en el grupo de control.

Aunque es estadísticamente significativa, la reducción de la HbA1c observada después de la intervención no es clínicamente pertinente, y es sustancialmente inferior al efecto comunicado por Chapin y otros²¹ en su estudio anterior. Ello podría reflejar las diferencias entre los grupos de control en ambos estudios. Así pues, aunque la media de HbA1c disminuyó en los grupos de control de ambos estudios, la reducción fue sustancialmente mayor en nuestro estudio (-0,38%) que en el estudio de Chapin y otros²¹ (-0,18%). Además, cabe señalar que el número de sesiones a las que asistieron los participantes se asoció directamente con el impacto de nuestra intervención y, por lo tanto, el modesto efecto podría ser resultado de la baja asistencia observada en nuestro estudio.

Referencias clave

Campbell MJ (2019). Cluster randomised trials. *Med J Aust*;210 (4):154-156.e1.

Un artículo que presenta muy claramente el método de los ensayos aleatorios.

Campbell MJ (2019), Hemming K, Taljaard M. The stepped wedge cluster randomised trial: what it is and when it should be used. *Med J Aust*.;210(6):253-254.e1.

Un artículo que presenta muy claramente el método de los ensayos aleatorios y las especificidades de los ECAs por conglomerados.

Ricci-Cabello I, Olry de Labry-Lima A, Bolívar-Muñoz J, Pastor-Moreno G, Bermudez-Tamayo C, Ruiz-Pérez I, et al (2013). Effectiveness of two interventions based on improving patient-practitioner communication on diabetes self-management in patients with low educational level: study protocol of a clustered randomized trial in primary care. *BMC Health Serv Res*.;13:433.

En este trabajo se explica en detalle de la metodología resultante final del ejemplo presentado en el capítulo.

Referencias

Campbell MK, Piaggio G, Elbourne D, Altman DG (2012). Consort 2010 statement: extension to cluster randomized trials. *BMJ*; 345: e5661.

Chapin RB, Williams DC, Adair RF (2003). Diabetes control improved when inner-city patients received graphic feedback about glycosylated hemoglobin levels. *J Gen Intern Med*;18(2):1204.

Danaei G, Finucane M, Lu Y, Singh GM, Cowand MJ, Paciorek CJ, et al (2011). National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2,7 million participants. *The Lancet*;378(9785):31-40.

Donner A, Klar N (2004). Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*; 94: 416-22.

- Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G (2008). Internal and external validity of cluster randomised trials: systematic review of recent trials. *BMJ*; 336: 876.
- Eldridge A, Kerry S. A Practical Guide to Cluster Randomised Trials in Health Services Research. [Consultado 02 Febrero 2020].
<https://onlinelibrary.wiley.com/doi/book/10.1002/9781119966241>
- Giraudeau B, Ravaud P (2009). Preventing bias in cluster randomised trials. *PLoS Med*; 6: e1000065.
- Grau M, Elosua R, Cabrera-de-León A, Guembe MJ, Baena-Díez JM, Vega-Alonso T, et al (2011). Cardiovascular risk factors in Spain in the first decade of the 21st Century, a pooled analysis with individual data from 11 population-based studies: The DARIOS study. *Rev Esp Cardiol*;64:295-304.
- Lázaro P. La investigación en el sistema de salud. En: Repullo Labrador JR y Oteo Ochoa LA, editores: un nuevo contrato social para un Sistema Nacional de Salud sostenible. Ariel. Barcelona, 2005:245-252.
- Martin de los Santos E, Portolés-Pérez A (2014). Aspectos metodológicos y éticos de los ensayos clínicos en conglomerados. En: Luces y sombras en la investigación clínica. Dal-Ré R, Carné X y Gracia D (Dir). Triacastela, Madrid: 217-240.
- Ministerio de Sanidad, Servicios Sociales e Igualdad. La salud de la población en España. Resultados de la Encuesta Europea de Salud en España 2014 [consultado 10 Mayo 2020]. Madrid 2017.
http://www.mscbs.gob.es/estadEstudios/estadisticas/EncuestaEuropea/Enc_Eur_Salud_en_Esp_2014.htm
- Molina Arias M, Ochoa Sangrador C (2014). Ensayo clínico (I). Definición. Tipos. Estudios cuasiexperimentales. *Evid Pediatr*;10:52.
- OMS (2016). Informe mundial sobre la diabetes. Ginebra: OMS
<https://www.who.int/diabetes/global-report/es/>
- Soriguer F, Goday A, Bosch-Comas A, Bordiú E, Calle-Pascual A, Carmena R, et al (2012). Prevalence of diabetes mellitus and impaired glucose regulation in Spain: The Di@bet.es study. *Diabetologia*.55:88-93.

Williams J, Allen L, Wickramasinghe K, Mikkelsen B, Roberts N, Townsend N (2018). A systematic review of associations between non-communicable diseases and socioeconomic status within low- and lower-middle-income countries. *J Glob Health*;8(2):020409.

Resumen / Résumé / Abstract

Los ensayos aleatorizados por conglomerados son muy útiles para evaluar intervenciones comunitarias y en contextos de limitación de recursos. Este es el caso de la evaluación de la intervención que se presenta. Su objetivo es la mejora del autocontrol de la diabetes en pacientes con un bajo nivel educativo y un control glucémico inadecuado, mediante la mejora de la comunicación entre el paciente y el proveedor. Se describe el proceso seguido para el diseño, los ajustes efectuados con la restricción de recursos y los retos metodológicos que se tuvieron que abordar.

Les essais randomisés en grappes aléatoires sont très utiles pour évaluer les interventions communautaires et dans des contextes à ressources limitées. C'est le cas de l'évaluation de l'intervention présentée. L'objectif est améliorer l'autogestion du diabète chez les patients ayant un faible niveau d'éducation et un contrôle glycémique insuffisant en améliorant la communication entre le patient et le prestataire de soins. Il décrit le processus suivi pour la conception, les ajustements apportés avec la restriction des ressources et les défis méthodologiques qui ont dû être relevés.

Randomised cluster trials are very useful for evaluating community interventions and in resource-constrained settings. This is the case of the evaluation of the intervention being presented. Their aim is to improve self-management of diabetes in patients with a low level of education and inadequate glycaemic control by improving patient-provider communication. We describe the process followed for the design, the adjustments made with the restriction of resources and the methodological challenges that had to be addressed.

Eva Martín Ruiz es socióloga y enfermera, doctora en ciencias de la salud por la universidad de Sevilla, con diplomas de especialización en gestión de cuidados y en gestión estratégica de profesionales, con experiencia investigadora de más de 15 años que se ha desarrollado en la EASP, a cuya plantilla pertenece, formando parte de equipos de investigación como IC, en proyectos financiados de ámbito autonómico, nacional y europeo, en cuyas líneas destaca la revisión sistemática de intervenciones de cuidados relacionadas con diferentes problemas de salud, y el desarrollo de herramientas metodológicas para la aplicación de tecnologías sanitarias. Los últimos proyectos en los que ha participado son de carácter multidisciplinar y multicéntrico, han obtenido financiación europea y sus principales resultados han quedado recogidos en publicaciones en formato artículo, libros y recursos web.

Antonio Olry de Labry Lima es técnico en la Escuela Andaluza de Salud Pública, Licenciado en Farmacia y Doctor en Salud Pública por la Universidad de Granada, Máster en Salud Pública y Gestión Sanitaria y Diploma en instituciones de salud y seguridad del paciente por Escuela Andaluza de Salud Pública. Es miembro investigador del Centro de Investigación Biomédica en Epidemiología y Salud Pública (CIBERESP). Tiene amplia experiencia en soporte metodológico, evaluación de tecnología de salud, revisiones sistemáticas y análisis de costo-efectividad. En los últimos años, ha trabajado en varios proyectos europeos sobre evaluación de tecnologías sanitarias.

Cita

Eva Martín Ruiz y Antonio Olry de Labry Lima (2020). Ensayo clínico aleatorizado por conglomerados. Diseño para evaluar la efectividad de una intervención en población desfavorecida para mejorar el autocuidado en diabetes adaptado a un contexto de restricción de recursos. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 383-399. Québec: Éditions science et bien commun y Marseille: IRD Éditions

PARTE V

EVALUACIÓN DE LOS PROCESOS Y FIDELIDAD DE IMPLEMENTACIÓN

16. Análisis de los procesos de implementación

Una intervención compleja en Burkina Faso: financiación basada en los resultados

VALÉRY RIDDE Y ANNE-MARIE TURCOTTE-TREMBLAY

Análisis de los procesos de implementación

Definición del método

El análisis de implementación tiene por objeto comprender cómo uno o más procesos organizados contribuyen a los cambios deseados en un contexto particular con una intervención con respecto a los medios utilizados.

Fortalezas del método

- Comprender cómo funciona una intervención en su contexto real y cómo evoluciona a lo largo del tiempo
- Evita el error de tipo III, es decir, la evaluación de una intervención cuando no se ha aplicado según lo previsto
- Utiliza diseños rigurosos, como estudios de casos longitudinales múltiples que combinan varios tipos de recogida de datos complementarios, como la observación, las entrevistas formales e informales o la documentación

Desafíos del método

- El volumen de datos cualitativos a veces importante, el compromiso entre el fortalecimiento de la validez interna frente a la externa, el tiempo necesario para el análisis y la complementariedad de lo cualitativo con lo cuantitativo
- El uso de modelos lógicos (teoría de la intervención) y marcos conceptuales para guiar la recogida y el análisis de los datos, pero sin centrarse solo en ellos
- Innovar y mantenerse abierto a los datos empíricos emergentes

“Sabemos lo que hay que hacer, pero no sabemos cómo hacerlo” es una frase que parece repetirse en los informes de los actores de salud global durante mucho tiempo. En 2010, por ejemplo, las Naciones Unidas declararon que al ocuparse de la salud de las mujeres y la infancia “sabemos lo que funciona” (Secretario General de Naciones Unidas 2010). Sin embargo, aunque el contenido de las intervenciones teóricamente efectivas es

relativamente conocido, su nivel de cobertura sigue siendo bajo (Peters et al. 2009). Además, hay poca comprensión de las condiciones para su implementación. Un meta-análisis muestra que la efectividad teórica de las intervenciones se reduce en un 50% debido a múltiples factores contextuales que afectan a la aplicación (J. A. Durlak y DuPre 2008). Así pues, no basta con saber si una intervención sanitaria global es eficaz en un entorno controlado (eficacia teórica: eficacia), sino que también es necesario comprender por qué funciona, cómo funciona, para quién funciona y en qué contextos reales (eficacia práctica: efectividad). Es aquí donde la ciencia de la implementación es de innegable ayuda. Se define como un campo de investigación destinado a comprender los “factores asociados con la aplicación efectiva” (Franks y Schroeder 2013). La aplicación efectiva se refiere a la capacidad de la intervención para organizar las actividades esenciales para lograr sus objetivos, pero también, y quizás lo más importante, a su capacidad para adaptarse a los actores y contextos a lo largo de su historia (Ridde, Delormier y Gaudreau 2007).

En palabras de un autor destacado en este campo: “estudiar la aplicación de las intervenciones no es fácil, pero es esencial” (Durlak 2015b). En una revisión de la literatura sobre el concepto, Pfadenhauer y sus colegas (2015) definen la aplicación como “un esfuerzo iniciado deliberadamente y planificado activamente con la intención de poner en práctica un objeto específico”. En este libro sobre métodos de evaluación de las intervenciones sanitarias globales y en el presente capítulo, por aplicación se entiende uno o más procesos organizados en un contexto particular para producir los cambios deseados por una intervención (en el sentido más amplio del término: política, programa, proyecto) en relación con sus medios. La aplicación puede distinguirse de la aplicación, que algunos consideran un anglicismo, en el sentido de que esta última es más normativa o “simple”, mientras que la primera es verdaderamente compleja y dinámica. La evaluación de la aplicación será útil para verificar que las actividades se han organizado según lo previsto (análisis del cumplimiento), mientras que la evaluación de los procesos permitirá comprender la aplicación, su complejidad, el papel de los agentes sociales y el contexto.

Por consiguiente, la ciencia de la implementación tiene por objeto movilizar teorías, conceptos y métodos para “comprender mejor qué, quién y cómo funciona una intervención en un contexto real” (Peters et al. 2013). Obviamente, no existe un “contexto irreal”, sino que deseamos referirnos

aquí al contexto natural en el que se lleva a cabo una intervención (Petticrew y otros, 2005), en contraposición a los contextos en los que los investigadores/as y profesionales intentan controlar las situaciones (contexto y acción) para demostrar la efectividad, a menudo utilizando un enfoque experimental y especificaciones como los ensayos controlados aleatorios.

Por consiguiente, el presente capítulo se aparta del concepto de “investigación sobre la implementación”, que a menudo es imposible de traducir y que se centra en el análisis de la forma en que las intervenciones tienen en cuenta e integran las pruebas en la formulación de su contenido (Nilsen y otros, 2013). La revista *Implementation Science* es el estandarte para esto, aunque al usar este título trae confusión adicional a un área que no la necesitaba. Por otra parte, este campo también se aproxima al de la transferencia de conocimientos, que también está muy poco desarrollada en los países del Sur (Siron, Dagenais y Ridde 2015). Sin embargo, estamos de acuerdo con el hecho de que “la investigación en estos dos campos se ocupa de los desafíos de convertir las intenciones en cambios deseados” (Nilsen et al. 2013).

Por lo tanto, en este capítulo nos centramos en los procesos de implementación de las intervenciones. Sin embargo, las investigaciones sobre los procesos de implementación en la salud global deben aprovechar mejor las teorías, los marcos conceptuales y los enfoques de las ciencias sociales (Van Belle, van de Pas y Marchal 2017). Por ejemplo, Durlak ofrece una interesante síntesis al sugerir ocho áreas de estudio que retoman los famosos “resultados de la implementación” de Proctor et al (2011): aceptabilidad, adopción, pertinencia, viabilidad, fidelidad, coste, cobertura, sostenibilidad. También sugiere 23 factores que influyen en la aplicación y 14 medidas que deben adoptarse para llevar a cabo una intervención (Joseph A. Durlak 2015a; Figura 1). Si bien los enfoques inductivos y empíricos son esenciales para generar discurso e identificar prácticas eficaces, el uso de desarrollos teóricos y conceptuales es igualmente esencial (Olivier de Sardan y Ridde 2015), ya que permite un enfoque más sistemático y completo y una mejor organización de los datos para facilitar el análisis y la presentación.

Recordemos que entre los estudios de promoción de la salud que analizan las políticas públicas, sólo el 18% hace referencia a un marco teórico en este campo (Breton y De Leeuw 2011). Los trabajos han demostrado, por ejemplo, que la teoría de las corrientes de Kingdon (Ridde 2009), la teoría

de Lipsky sobre los actores de primera línea (Walt y otros, 2008) o la teoría de la innovación de Rogers (Pérez y otros, 2013), todas ellas elaboradas en los Estados Unidos, pueden adaptarse al contexto de determinados países del Sur para comprender la aplicación. Además, todavía no se ha abordado el papel de las ideas en la aplicación (en contraposición al surgimiento) de las intervenciones en los países del Sur (Béland y Ridde 2016), mientras que la literatura sobre este tema en los países del Norte es abundante (Cox 2001).

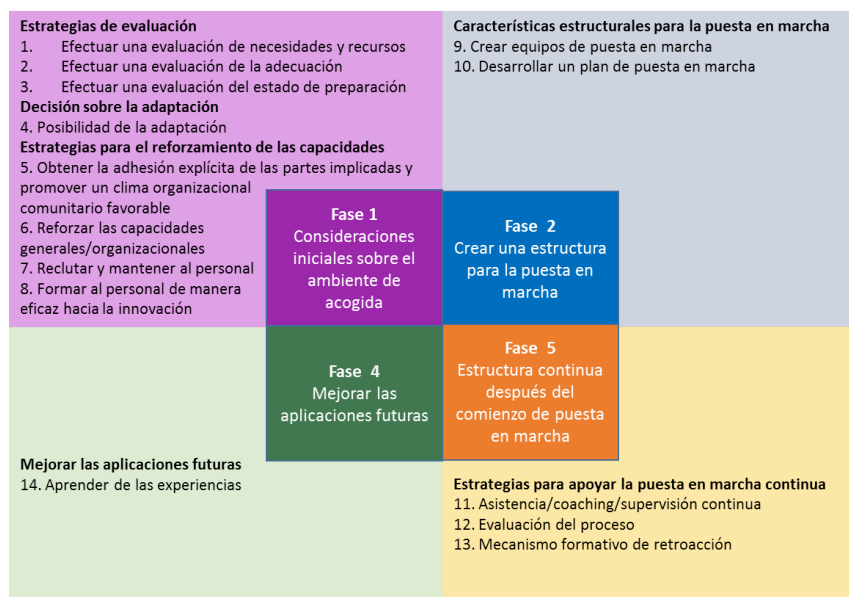


Figura 1. 14 pasos para una implementación exitosa según Joseph A. Durlak (2015a)

Es evidente que la ciencia de la implementación no puede desarrollarse por sí sola, sin considerar los efectos de las intervenciones (o los principios de eficacia, en palabras de Patton [2010]), a riesgo de caer en el error de tipo III, en el que una intervención se evalúa cuando no se ha aplicado como se pretendía (Dobson y Cook 1980). Es importante, por ejemplo, describir mejor el contenido de las intervenciones e informar sobre la fidelidad (tradicional para los epidemiólogos) y la intensidad de la aplicación (véase el capítulo de Pérez y otros).

Por lo tanto, es hora de tomar en serio la petición formulada hace 10 años con respecto a los países del Sur: “más investigación sobre la

implementación y especialmente sobre los retos de la realización de intervenciones que tengan como objetivo una mayor equidad” (Gilson y Raphaely 2008). Por consiguiente, en este contexto hemos realizado un análisis de la puesta en marcha de una compleja intervención en Burkina Faso, que describimos en la siguiente sección.

Contexto y programa evaluados

En los últimos años en África, muchos países han emprendido intervenciones de financiación basada en los resultados para mejorar el rendimiento de los sistemas de salud. Existen muchos modelos de FBR, pero en este capítulo describimos el promovido por el Banco Mundial en Burkina Faso. En este contexto, el objetivo es vincular la financiación de los centros de salud y el pago de primas al personal sanitario con el logro de indicadores predefinidos de cantidad y, a menudo, de calidad de la atención. Estos indicadores están sujetos a un proceso de control por parte de actores externos al establecimiento sanitario. Las pruebas científicas de la eficacia o eficiencia del FBR siguen siendo limitadas tanto en el Norte (Milstein y Schreyoegg 2016) como en el Sur (Turcotte-Tremblay et al. 2016; Witter et al. 2012). Las investigaciones sobre su aplicación siguen siendo muy escasas en el África occidental francófona (Paul, Sossouhounto y Eclou 2014). Los métodos y resultados son muy variados, pero todos confirman la falta de conocimientos sobre la aplicación del marco de referencia en África.

En Burkina Faso, tras un proyecto pre-piloto en tres distritos, el Ministerio de Salud y el Banco Mundial desplegaron una intervención de FRB en 12 distritos adicionales en enero de 2014. Burkina Faso ha decidido innovar con estrategias para mejorar la equidad y apoyar el acceso a la atención de los más pobres. Así pues, en 10 distritos se organizó en ciertos centros de salud seleccionados al azar un proceso de selección de base comunitaria para personas indigentes que les permitía quedar exentas del pago de la atención. Otros dos distritos organizan un seguro médico comunitario combinado con la exención de pago para personas indigentes. Todos los pacientes seguían pagando por la atención en el punto de servicio, excepto los pacientes indigentes que están exentos o las mujeres embarazadas cuyos

partos han sido subvencionados (80%) por el Estado desde 2006 y gratuitos desde 2016. Escrito con el apoyo de los consultores del Banco Mundial, se dispone de una guía de intervención de más de 300 páginas.

Se creó una dependencia central denominada Servicio Técnico de la FBR (ST-FBR) para apoyar la aplicación de la FBR mediante la elaboración y la vigilancia de indicadores, la supervisión de auditorías, la organización de cursos prácticos de capacitación y el análisis de los datos de rendimiento. Se crearon agencias de contratación y verificación (CVA) para verificar los resultados cuantitativos de los establecimientos sanitarios, apoyar a los trabajadores de la salud en el uso de los instrumentos de FBR, introducir datos y presentar facturas de los establecimientos sanitarios para su validación. Los LCA realizan controles de cantidad mensualmente para contar el número de servicios específicos prestados (~23 indicadores). Se realizan controles de calidad trimestrales en varias dimensiones de la calidad de la atención, como las condiciones de recepción de los pacientes, la disponibilidad de equipo, el mantenimiento de los materiales, la gestión de las drogas y la gestión financiera (~104 indicadores). Esto último requiere la colaboración de los miembros de los comités de gestión (COGES) de los centros de salud debido a su participación en la tesorería del Centro de Salud y Promoción Social (CSPS). Se han establecido diversos instrumentos (por ejemplo, el plan de mejora del rendimiento, el instrumento del índice) para promover la planificación de las actividades y las finanzas. Los indicadores de cantidad se adquieren a precios unitarios fijos y los de calidad se adquieren de acuerdo con un objetivo mínimo a alcanzar.

La organización del FBR se diseñó para facilitar su evaluación de impacto financiado por el Banco mundial, basado en un ensayo controlado aleatorio de las estructuras de primera línea, a saber, el CSPS. Se definen cuatro categorías de aleatoriedad, como muchos brazos de intervención: i) FBR 1: se paga al centro de salud por los indicadores de actividades logrados sobre la base de precios fijos, pero que pueden fluctuar en el curso del proyecto; ii) FBR 2: FBR 1 junto con una intervención de selección comunitaria para personas indigentes exentas del pago de la atención adquirida a un precio moderado; iii) FBR 3 : FBR 2 a un precio de compra más alto para los indigentes para fomentar iniciativas innovadoras para aumentar su uso de los servicios; iv) FBR 4: FBR 1 combinado con un seguro de salud de base

comunitaria junto con una intervención de selección de indigentes de base comunitaria. Este último componente de la FBR se aplica en una sola región sanitaria.

Para las cuatro modalidades, se ha establecido una “prima de equidad” entre los distritos y los centros de salud. Como resultado, los precios de compra varían significativamente para los centros de salud dentro de estas cuatro categorías. Estos precios se calcularon teniendo en cuenta la distancia de los centros de salud a la capital del distrito y las necesidades de recursos. Por ejemplo, una consulta curativa para una persona de más de cinco años puede adquirirse entre 100 y 180 francos CFA (100 francos CFA = 0,15 euros) dependiendo de los CSPS del mismo distrito.

El análisis de la fidelidad de la aplicación del FBR realizado 12 meses después de su inicio muestra que, en general, se han aplicado los elementos importantes y esenciales de la teoría de la intervención (Belaid et al. 2016). Sin embargo, observamos déficits de aplicación y retrasos en la verificación de la ejecución y los pagos de las subvenciones del Fondo Fiduciario para el fomento de la capacidad, lo que compromete en cierta medida la eficacia teórica de la lógica de los incentivos. Aunque este análisis de fidelidad es esencial, no es suficiente para comprender el funcionamiento de la intervención en todas sus sutilezas. Así pues, emprendimos un análisis cualitativo de la aplicación del RBF en Burkina Faso de acuerdo con una especificación que se explica en la siguiente sección.

Estimación de la evaluación

Este estudio forma parte de la corriente de investigación y evaluación de la intervención basada en la teoría de la intervención (Weiss 1998). La teoría de intervención del FBR guió nuestra recogida de datos y análisis. Metodológicamente, nos encontramos en la tercera generación de investigación de implementación de intervenciones, incluyendo el uso de múltiples estudios de casos longitudinales (Saetren 2014). Sin embargo, en el momento de redactar el presente capítulo, sólo se ha completado la primera de las tres fases de recogida de datos previstas para los estudios de casos y es el tema de esta presentación.

El enfoque metodológico es el de un estudio de casos múltiples (los CSPS) y contrastado con varios niveles anidados de análisis (los grupos de

actores involucrados) (Yin y Ridde 2012). La selección de los casos se ha explicado en otra parte (Zombré y otros 2016). La investigación se llevó a cabo en tres distritos sanitarios diferentes a fin de representar la diversidad de los procesos de aplicación del BFR. La recogida se llevó a cabo en un hospital regional, dos hospitales de distrito y 18 centros de salud de primera línea (CSPS). El análisis para los hospitales es objeto de otro informe porque el contexto es diferente. Un debate de grupo con los miembros de los tres equipos de gestión de distrito (EGD), apoyado por un análisis estadístico de los datos de rutina, permitió clasificar los EGD según su nivel de rendimiento antes del inicio del FBR. Se incluyó aquí el desempeño con respecto a los indicadores de utilización de los servicios disponibles en estos datos de rutina, cuya calidad se verificó. En los dos distritos que organizaron la FBR1, FBR2 y FBR3, se seleccionaron un CSPS con rendimiento y otro sin rendimiento ($n = 3 \times 2 \times 2$). En el distrito que organizó la FBR 1 y la FBR4, se seleccionaron un CSPS de alto rendimiento, un CSPS de rendimiento moderado y un CSPS de rendimiento nulo ($n = 3 \times 2$) (Figura 2). En cada uno de los distritos, la investigación estuvo a cargo de un investigador experimentado del país, un socio-antropólogo, que coordinó la recogida de datos llevada a cabo por cuatro asistentes de investigación, quienes se encargaron de la recogida de datos en dos o tres CSPS del mismo tipo de FBR, excepto en los hospitales en los que sólo un asistente llevó a cabo la recogida de datos. Todos ellos vivieron durante 10 a 15 días en cada pueblo/ciudad donde se encontraban los centros de salud. Las tres recopilaciones de datos comenzaron en el otoño de 2014, 2015 y 2016, respectivamente. Un taller de tres días reunió a todos los investigadores y asistentes de inicio de la investigación para coordinar el proceso, con algunas sesiones de debate con los responsables de la toma de decisiones y las organizaciones que aplican la intervención.

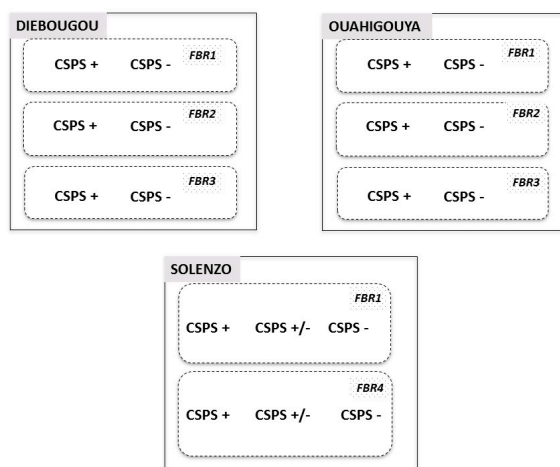


Figura 2. 18 casos de tres distritos sanitarios

Se utilizaron cuatro instrumentos cualitativos de reunión de datos.

1- Se llevaron a cabo observaciones (participantes y no participantes) en las aldeas (mercados, reuniones comunitarias, etc.) para redactar una monografía del contexto local y en los centros de salud para estudiar las interacciones entre los agentes durante las consultas, las visitas de control y supervisión, etc. Los resultados de las observaciones se registraron en un cuaderno de bitácora para cada asistente.

2. Se realizaron entrevistas individuales formales utilizando una guía de prueba previa común a todos los casos y relativa a las dimensiones de la teoría de la intervención. Las personas fueron seleccionadas después de varios días sobre el terreno según su capacidad y disponibilidad para arrojar luz útil sobre la BFR y según el principio de muestreo que hace hincapié en la variación y la intensidad dentro del mismo grupo de agentes estratégicos (Palinkas et al. 2013): trabajadores sanitarios, trabajadores sanitarios de la comunidad, miembros del comité de gestión (COGES), funcionarios electos y representantes de asociaciones locales, usuarios y poblaciones. En el cuadro 1 se presenta la distribución de las 459 entrevistas individuales formales según los distritos para esta primera ronda (recuerde que se ha previsto un

total de tres entrevistas para realizar estudios de casos longitudinales). En promedio, realizamos 22 entrevistas por área de salud de un centro de salud. La foto 1 ilustra una entrevista formal con el presidente de un SMC.

3- Se realizaron entrevistas informales: a diferencia de las anteriores, se trataba de entrevistas no planificadas, realizadas sin grabación, pero esenciales para comprender el marco de derechos humanos en un contexto de asistencia al desarrollo. Las personas fueron seleccionadas en función de las oportunidades (Palinkas y otros, 2013) y de su contribución a la comprensión de la



Entrevista formal con el presidente de un SMC

situación estudiada. Posteriormente se tomaron notas para conservar los elementos más importantes del debate. La última fila de la Tabla 1 muestra la distribución de las 224 entrevistas individuales informales por distrito. Los retos asociados a este gran número de entrevistas se examinan al final del capítulo.

4- Se analizaron numerosos documentos: informes de actividades y de formación, documentos de formación, guía de aplicación, registros de consulta, etc. En los cuadernos de bitácora de los asistentes e investigadores se anotaron resúmenes de elementos importantes.

Tabla 1 : Número de entrevistas por distrito

	Distrito 1	Distrito 2	Distrito 3
Miembros del Equipo de Liderazgo de Distrito	8	2	5
Profesional de la salud	36	46	52
Técnicos de salud comunitarios	8	12	22
Los miembros de los SMC	14	27	25
Funcionarios electos, ONG y líderes locales	28	20	52
Usuarios de los servicios	17	23	59
TOTAL DE ENTREVISTAS FORMALES	114	130	215
TOTAL DE ENTREVISTAS INFORMALES	168	30	26

Todas las entrevistas fueron transcritas en su totalidad. Usamos el software QDA Miner Lite 4 para asignar temas a los segmentos de texto relevantes para organizar y extraer datos. El segundo autor de este capítulo impartió capacitación sobre el programa informático a los miembros del equipo de investigación, lo que incluyó el ensayo y la adaptación de los códigos y el desarrollo de una comprensión común del proceso de análisis. El análisis siguió un enfoque deductivo-inductivo asignando temas predefinidos a los datos basados en nuestro marco analítico (Ritchie y Spenser 1994), a saber, la teoría de la intervención, y desarrollando nuevos temas mediante la lectura de las transcripciones. Los asistentes escribieron una monografía y un informe de investigación para sintetizar los resultados de sus estudios de casos (n= 4 asistentes en los 3 distritos). Posteriormente, los tres investigadores del distrito realizaron un análisis transversal intra-distrital (n = 3 informes). Por último, el primer autor llevó a cabo el análisis transversal interdistrital que fue validado por los coinvestigadores. La figura 3 presenta este proceso analítico para este primer análisis transversal. El presupuesto general para la recogida de datos (de los tres previstos) asciende a unos 35.000 euros, incluidos los gastos de recogida de datos para los asistentes y sus viajes, los gastos de transcripción de las entrevistas, el pago de su tiempo de análisis y el pago de los honorarios de los tres investigadores principales de Burkina Faso.

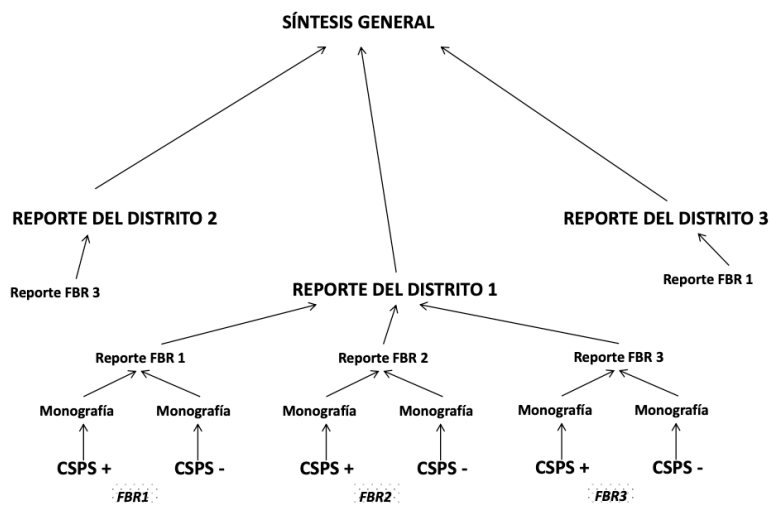


Figura 3. Proceso de análisis transversal

Algunos resultados

Resumimos en unas pocas líneas los principales hallazgos de esta investigación cualitativa.

Siguiendo un enfoque de “capacitación en cascada” a gran escala (incluida la capacitación fuera del país para algunos directivos), todos los que participaron eran los más conocedores de los principios generales y el funcionamiento de la FBR. Todos los SF han designado un punto focal de la FBR (es decir, una persona focal para la intervención que ha sido entrenada y que se supone que debe compartir información. Pero más allá de este círculo, el nivel de conocimiento disminuye hasta llegar a ser inexistente para la población o los líderes de la comunidad. En algunos CSPS, la concesión de primas al personal no se ha anunciado deliberadamente a la población y a los líderes de la comunidad. La visión general de las personas reunidas que conocen el FBR sigue siendo la de la presencia de un “proyecto”, como tantos otros antes y después de él. Así pues, los encuestados saben que

es temporal, financiada y dirigida por la ayuda internacional, y que puede dar lugar a muchas lógicas de monopolización, especialmente de los recursos financieros.

El establecimiento de organismos de vigilancia y verificación y de contratos de ejecución se ha retrasado mucho, especialmente en lo que respecta al pago de subsidios y primas. Estos retrasos debidos a los engorrosos procedimientos administrativos se acumulan (contratación de ACV, firma de contratos, etc.) con el resultado de que los agentes locales no conocen realmente el funcionamiento de los contratos, los indicadores o los planes de mejora del rendimiento (utilizados para planificar las inversiones de las subvenciones del FRB y otros ingresos para mejorar los servicios). Las primeras auditorías en todas partes condujeron a discusiones de equipo y reajustes rápidos de los trabajadores de la salud para que los siguientes fueran mejores para recibir los máximos subsidios posibles, porque “cuando el FBR se usa para ser evaluado, son demasiado rigurosos, pequeño error, ¡no hay perdón en el FBR! ». Sin embargo, la auditoría cuantitativa suele ser una fuente de controversia (entre los auditores y los trabajadores de la salud, entre el personal de atención de la salud y el personal de apoyo, y dentro del personal de atención de la salud). Además, en algunos CSPS de varios distritos, observamos la preparación de un escenario de preauditoría. Esto último siempre se anuncia o simplemente se conoce, ya que la información circula rápidamente en estos entornos donde el nivel de interconocimiento es alto. Asimismo, algunos datos se registran en los registros porque forman parte de las comprobaciones, mientras que nuestras observaciones muestran que no son reales. Es el caso, por ejemplo, de la toma de los signos vitales del paciente (temperatura, presión sanguínea, peso) o el relleno de los partos, que algunos trabajadores de la salud llaman “postogramas” en privado (es decir, rellenos después del hecho), según nos informaron durante las entrevistas informales.

La distribución de las primas individuales varía entre los centros de salud. Depende del liderazgo local y de las cuestiones de poder. En algunos casos, hay una absoluta falta de transparencia. En otro lugar, pero esto parece ser una excepción, en un CSPS de alto rendimiento, el gerente decidió involucrar a todos los agentes de su equipo para calcular las primas con total transparencia.

En su diseño, el FBR no preveía que se dieran bonificaciones a los miembros de los SMC, que sin embargo administran el presupuesto, la

cuenta bancaria del CSPS y la movilización de la comunidad. La mayoría de los centros de salud respetaban esta norma, pero otros se oponían a ella porque la consideraban injusta, dada la importante labor realizada por estos miembros de la comunidad local. Además, son los que reciben los grandes subsidios de la compra de servicios por parte de la FBR. Como dijo un presidente del SMC: “El FRB necesita pensar en motivar a los miembros del SMC también. ¿Cómo va alguien a retirar más de 400.000 F CFA [600 euros] para venir a dárselos a los agentes que van a compartir? ».

En resumen, los resultados demuestran que la adaptación y la evolución son la norma y no la excepción en la aplicación de las intervenciones de salud pública, lo que obviamente plantea problemas de evaluación (Pérez y otros, 2013; Ridde y Robert, 2014). La intervención de la FBR fue muy centralizada, su contenido planeado y definido en detalle, pero su encuentro con el contexto, los actores y los eventos condujo a que las cosas no sucedieran como se habían planeado. Los importantes retrasos en los pagos y lo que está en juego en la distribución de las primas son preocupantes. El débil patrimonio local y la financiación externa plantean problemas para la sostenibilidad, ya que hasta que los funcionarios gubernamentales se comprometan con este tipo de intervención y proporcionen recursos financieros estatales, la continuidad del enfoque se verá comprometida.

Análisis reflexivo

La fuerza de nuestro análisis es que se basa en la triangulación empírica de una muestra muy grande de datos y casos contrastantes, y en la colaboración interdisciplinaria para asegurar la calidad de los datos. Todos los asistentes han sido capacitados y supervisados para garantizar el rigor de su labor, y el conocimiento detallado de los investigadores del contexto y las políticas sanitarias de Burkina Faso hace que sus interpretaciones sean acertadas. Este tipo de proceso es relativamente raro en África, ya que a menudo se pide a los asistentes que “simplemente” reúnan datos que luego son analizados por los investigadores del país o, peor aún, sólo por los investigadores del Norte (Ouattara y Ridde 2013). Evidentemente, esto no ha sido fácil y el contexto de Burkina Faso puede ser más adecuado que en otros lugares, ya que hemos estado trabajando allí durante mucho tiempo, con una red de investigadores amplia, generalmente adscritos a instituciones.

Por lo tanto, tienen una gran red de estudiantes dentro de la cual hemos podido encontrar personas competentes. En Malí, donde estamos tratando de organizar el mismo proceso, la situación es más compleja, ya que los recursos humanos parecen ser más escasos y nuestros compañeros de investigación participan menos en los programas de formación para tener una red de estudiantes.

Los resultados han sido ampliamente compartidos y validados con muchas partes interesadas en la FBR en el país. Por lo tanto, creemos que pueden ser transferibles a todos los distritos del país involucrados en la aplicación del RBF.

Además, los resultados serán sin duda útiles para comprender mejor la evaluación de los efectos. Desde el comienzo de esta investigación de implementación, hemos mantenido estrechos vínculos con los investigadores del equipo de evaluación de impactos. Nuestra experiencia en materia de aplicación ha sido útil para ayudar a este equipo a comprender mejor el curso real de la intervención (más allá de los manuales de intervención inicial) y a adaptar parte de sus cuestionarios de salida normalizados para mejorar su pertinencia en el contexto local. Esta colaboración representa un valor añadido para nuestras respectivas investigaciones y desarrollo de conocimientos al salvar la brecha entre los enfoques cualitativos y cuantitativos.

En el caso de este estudio de implementación, el principal desafío sigue siendo la gran cantidad de datos cualitativos (¿demasiados?) y, por lo tanto, su consecuencia, que fue la longitud del análisis. Se tardó más de un año en disponer de los primeros informes de análisis de casos para iniciar un análisis colectivo y transversal de las 459 entrevistas formales y las 224 entrevistas informales de los 21 centros de salud. La utilización de un proceso piramidal en el que cada miembro del equipo analizaba y sintetizaba los datos a un nivel diferente (CSPS, distrito y general) facilitaba la gestión de los datos mediante la distribución del volumen de trabajo. Al utilizar una cantidad tan grande de datos recogidos en diferentes sitios, esperamos que los resultados puedan explicar la heterogeneidad de la aplicación, además de convencer a un amplio público de la validez de los resultados, incluidos los encargados de adoptar decisiones y algunos investigadores para los que el tamaño de la muestra sigue siendo esencial. De hecho, nuestra experiencia en la región muestra que los responsables de la toma de decisiones suelen desacreditar la investigación cualitativa debido al pequeño tamaño de las muestras, la

falta de comprensión de los conceptos de saturación empírica o el desconocimiento de los criterios de la ciencia cualitativa (Laperrière 1997). En este contexto, ¿cómo puede encontrarse un buen equilibrio entre la necesidad de disponer de datos suficientes para responder a una compleja pregunta de investigación en torno a una intervención en gran escala con múltiples componentes (armas), los desafíos éticos de atender a tantas personas y movilizar un equipo tan grande, y las demandas de los responsables de la toma de decisiones de una mayor validez externa de las evaluaciones de los programas? Este análisis de la aplicación es un intento de reconciliar estos mundos sin pretender cerrar la discusión.

Otro reto en el contexto de esta colaboración internacional ha sido mantener una comunicación dinámica entre los diferentes miembros del equipo, dispersos en diferentes regiones de Burkina Faso y de todo el mundo. Se han comunicado malentendidos, en particular en relación con las funciones de cada miembro del equipo, los métodos de reunión de datos, los plazos de los estudios y las disposiciones de indemnización de los asistentes de investigación. La distancia, el tamaño del equipo y los problemas de la red de comunicaciones, así como los costos asociados, ciertamente contribuyeron a la magnitud del desafío. Además, en el contexto de la colaboración internacional, las demoras en el análisis y el intercambio de informes nunca son fáciles de gestionar, ya que hay que ser a la vez investigador (¡del Norte, además!) y trabajar con otros compañeros, pero al mismo tiempo hay que ocuparse de cuestiones financieras y de tiempo para disponer de productos suficientemente robustos. Estos desafíos son bien conocidos (Ouattara y Ridde 2013).

Uno de los puntos fuertes de nuestro enfoque ha sido la organización de la formación en el uso de software libre de análisis cualitativo (QDA miner lite) para los miembros del equipo (ver Foto 2). Esta capacitación, poco habitual en Burkina Faso, contribuyó a fomentar la capacidad de los asistentes para esta investigación y para su futura carrera en este campo. También fue una oportunidad para integrarlos en todo el proceso y no convertirlos en simples recolectores de datos, como suele ocurrir, lo que plantea importantes problemas para la investigación cualitativa. Varios participantes expresaron su interés y gratitud a este respecto.

Uno de los principales retos durante la capacitación fue la gestión de los diferentes niveles de conocimientos informáticos de los participantes. Teníamos que encontrar un ritmo que se ajustara a todo el grupo. Otra

dificultad con que se tropezó durante la capacitación fue la de acordar las definiciones y la pertinencia de ciertos códigos de análisis. Por ejemplo, tuvimos una discusión sobre la creación de códigos separados para la “utilidad del FBR” y la “relevancia del FBR”. Estas discusiones son importantes para desarrollar un entendimiento común del proceso analítico, pero pueden ser difíciles cuando hay cientos de códigos por desarrollar. Además, la utilización de programas informáticos de análisis cualitativo para organizar los datos facilitó la realización de análisis secundarios para estudiar posteriormente temas específicos (por ejemplo, la calidad de la atención, la motivación de los agentes).



Capacitación en la utilización de programas informáticos gratuitos de análisis cualitativo

A pesar de los esfuerzos de algunos miembros del equipo, la centralización de los numerosos documentos de investigación (grabaciones de entrevistas, notas de observación, transcripciones de entrevistas, datos codificados) fue un gran desafío para este estudio. La centralización de los documentos es importante para

evitar la pérdida de datos, verificar los datos en bruto y permitir análisis secundarios. Sin embargo, la centralización resultó difícil durante toda la investigación, en particular debido a la dispersión geográfica de los asistentes de investigación y a la deficiente conexión a la Internet, que obstaculizó el rápido intercambio de archivos grandes. Más de tres años después del comienzo de esta investigación, seguimos pidiendo la centralización de todos los documentos.

Este estudio fue parte de la corriente de evaluación basada en la teoría de la intervención. Así pues, la teoría de la intervención guió la reunión y el análisis de datos más que los marcos o las teorías de la ciencia de la aplicación. Dicho esto, en nuestras reflexiones y análisis se han utilizado ampliamente conceptos de este campo, como la noción de contexto, el papel de los actores y las cuestiones de poder, los déficits de aplicación, el papel central de los actores de primera línea entendidos como los verdaderos “hacedores” de políticas (Lipsky 2010; Pfadenhauer y otros 2015; Pressman y Wildavsky 1984; Shoveller y otros 2015), etc. En Malí, sin embargo, se utilizó ampliamente un marco conceptual específico para el análisis de la aplicación

para estudiar la intervención de la FBR en una región del país (Damschroder et al. 2009). Pronto podremos sacar lecciones de él, en particular en los desafíos de su uso por parte de los antropólogos que no están acostumbrados, y esto es normal, a enmarcar excesivamente las dimensiones que estudian. Sin embargo, dada la complejidad de las intervenciones, la necesidad de múltiples estudios de casos para aumentar la validez externa de los resultados que requieren los encargados de la adopción de decisiones, que están más acostumbrados a los datos cuantitativos, y la fenomenal cantidad de datos cualitativos necesarios, puede ser inevitable el uso de este tipo de marco conceptual y de análisis asistido por computadora. El intento de innovar en esta área se hace esencial (Belaid et al. 2016).

En el capítulo se presentan los resultados de una reunión de datos transversales, que reflejan la situación en un momento dado y no su evolución (aunque, incluso durante los 12 meses del estudio, se observaron en gran medida adaptaciones). Pero queríamos compartir los resultados sin esperar a las otras fases, ya que la investigación sobre el tema es rara en el África occidental francófona. Las próximas fases de nuestro programa de investigación nos permitirán comprender mejor los cambios en la intervención, su adaptación y la forma en que el contexto y los actores actúan a lo largo del tiempo. Los estudios de casos longitudinales son, en efecto, esenciales para comprender la política, pero difíciles de organizar (Saetren 2014; Ridde y otros 2014; Hill y Hupe 2014).

La evaluación de los procesos de ejecución de las intervenciones es esencial, no sólo para comprender los efectos (o la falta de ellos) de las acciones emprendidas, sino, lo que es más importante, para captar la complejidad de los procesos, el papel de los agentes y el contexto. En este capítulo se presenta un ejemplo de cómo se puede movilizar a este respecto la teoría de la intervención y, más ampliamente, los marcos conceptuales de la ciencia de la aplicación. Ciertamente, necesitamos colectivamente más y mejores evaluaciones de la aplicación de las intervenciones sanitarias globales, incluida la utilización de marcos conceptuales y datos cualitativos.

Referencias clave

Patton, M. Q. (2008). *Implementation Evaluation: What happened in the program? Utilization-focused Evaluation*. 4th Edition. Saint-Paul: SAGE Publications.

Patton argumenta de forma convincente la importancia de estudiar la aplicación de las intervenciones. El autor relata la inverosímil historia de un grupo de evaluadores que evaluaron los efectos de una intervención que nunca se había aplicado realmente, sacando conclusiones erróneas sobre su ineficacia. El autor presenta cinco tipos de evaluación: evaluación del esfuerzo, supervisión, evaluación del proceso, evaluación de los componentes y especificación del tratamiento. El capítulo concluye con una lista de ejemplos de preguntas de evaluación relacionadas con la aplicación.

Damschroder, L. et al., (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4(50).

En este artículo se presenta un modelo conceptual, el Marco Consolidado para la Investigación de la Aplicación, que se basa en diferentes teorías. El modelo se compone de cinco dominios: las características de la intervención, el entorno externo, el entorno interno, las características de los individuos y el proceso de aplicación. Los autores presentan definiciones útiles para los diferentes conceptos y sus construcciones. En el siguiente sitio web se pueden consultar varios otros recursos: <https://cfirguide.org/>.

Olivier de Sardan, J.-P. (2008). *La rigueur du qualitatif. Les contraintes empiriques de l'interprétation socioanthropologique*. Louvain-La-Neuve: Bruylant.

¡Este es un libro de lectura obligada! Permite comprender la importancia de la inserción prolongada en un ambiente para entender la implementación de una intervención. También le da las herramientas necesarias para enfrentar los desafíos de la observación participativa. El libro aborda varias cuestiones importantes: ¿Cuáles son los procedimientos de observación? ¿Cuáles son las condiciones de rigor en los enfoques cualitativos? ¿Qué lugar debe darse a la subjetividad del investigador? ¿Pueden minimizarse los escollos de la sobreinterpretación o los sesgos de las ideologías científicas?

Ridde, V., Yaogo, M., Somé, P.-A., Turcotte-Tremblay, A.-M., (2017). Twelve months of implementation of health care performance-based financing in Burkina Faso: A qualitative multiple case study. *The International Journal of Health Planning and Management*. 33 (1), e153-e167.

En este artículo se presenta nuestro estudio sobre la aplicación de la financiación basada en los resultados en Burkina Faso. Proporcionará un ejemplo concreto del tipo de artículo científico que puede publicarse al estudiar la aplicación de una compleja intervención sanitaria global. El artículo contiene una referencia a nuestro protocolo completo que puede servirle de ejemplo para desarrollar un protocolo para entender cómo funciona una intervención, para quién y en qué escenarios de la vida real.

Referencias

- Belaid, L., Bodson, O., Ridde, V., Samb, O. et Turcotte-Tremblay, A.-M. (2016). Innovations qualitatives dans l'évaluation des interventions en santé. Dans J. Kivits, F. Balard, C. Fournier et M. Winance (dir.), *Les recherches qualitatives en santé* (p. 260-276). Malakoff : Armand Colin.
- Béland, D. et Ridde, V. (2016). Ideas and policy implementation: Understanding the resistance against free health care in Africa. *Global Health Governance*, 10(3), 9-23.
- Breton, E. et De Leeuw, E. (2011). Theories of the policy process in health promotion research: a review. *Health Promotion International*, 26(1), 82-90.
- Cox, R. H. (2001). The social construction of an imperative: Why welfare reform happened in Denmark and the Netherlands but not in Germany. *World Politics*, 53(3), 463-498.
- Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A. et Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science*, 4, 50-64.
<https://doi.org/10.1186/1748-5908-4-50>
- Dobson, D. et Cook, T. D. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning*, 3(4), 269-276.

- Durlak, J. A. et DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(34), 327-350.
<https://doi.org/10.1007/s10464-008-9165-0>
- Durlak, J. A. (2015a). What everyone should know about implementation. Dans J. A. Durlak, C. E. Domitrovich, R. Weissberg et T. Gullotta (dir.), *Handbook of social and emotional learning: research and practice* (p. 395-405). New York: The Guilford Press.
- Durlak, J. A. (2015b). Studying program implementation is not easy but it is essential. *Prevention Science*, 16(8), 1123-1127.
<https://doi.org/10.1007/s11121-015-0606-3>
- Franks, B. et Schroeder, J. (2013). Implementation science : What do we know and where do we go from here?. Dans A. Metz et I. Martinez-Beck (dir.), *Applying implementation science in early childhood programs and systems* (p. 5-21). Baltimore: Paul H. Brookes Publishing Company.
- Gilson, L. et Raphaely, N. (2008). The terrain of health policy analysis in low and middle income countries: a review of published literature 1994-2007. *Health Policy and Planning*, 23(5), 294-307.
- Hill, M. J. et Hupe, P. L. (2014). *Implementing public policy: an introduction to the study of operational governance* (3^e éd.). Thousand Oaks, CA: Sage Publications.
- Laperrière, A. (1997). Les critères de scientificité des méthodes qualitatives. Dans J. Poupart, J-P. Deslauriers, L-H. Groulx, A. Laperrière, R. Mayer et A.P. Pires (dir.), *La recherche qualitative : enjeux épistémologiques et méthodologiques* (p. 365-389). Montréal : Gaëtan Morin.
- Lipsky, M. (2010). *Street-level bureaucracy: Dilemmas of the individual in public services*. New York : Russel Sage Foundation.
- Milstein, R. et Schreyoegg, J. (2016). Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy*, 120(10), 1125-1140.
<https://doi.org/10.1016/j.healthpol.2016.08.009>

- Nilsen, P., Ståhl, C., Roback, K. et Cairney, P. (2013). Never the twain shall meet? – a comparison of implementation science and policy implementation research. *Implementation Science*, 8, 63.
<https://doi.org/10.1186/1748-5908-8-63>
- Olivier de Sardan, J.-P et Ridde, V. (2015). Public policies and health systems in Sahelian Africa: Theoretical context and empirical specificity. *BMC Health Services Research*, 15(Suppl. 3), S3.
<https://doi.org/10.1186/1472-6963-15-S3-S3>
- Ouattara, F. et Ridde, V. (2013). Expériences connues, vécues... mais rarement écrites : à propos des relations de partenariat Nord-Sud. *Nouvelles Pratiques Sociales*, 25(2), 231-246.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N. et Hoagwood, K. (2013). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533-544.
<https://doi.org/10.1007/s10488-013-0528-y>
- Patton, M. Q. (2010). *Developmental evaluation : Applying complexity concepts to enhance innovation and use*. Lieu d'édition: The Guilford Press.
- Paul, E., Sossouhounto, N. et Eclou, D. S. (2014). Local stakeholders' perceptions about the introduction of performance-based financing in Benin: A case study in two health districts. *International Journal of Health Policy and Management*, 3(4), 207-214.
<https://doi.org/10.15171/ijhpm.2014.93>
- Perez, D., Lefevre, P., Castro, M., Toledo, M. E., Zamora, G., Bonet, M. et Van der Stuyft, P. (2013). Diffusion of community empowerment strategies for Aedes aegypti control in Cuba: A muddling through experience. *Social Science and Medicine*, 84, 44-52.
<https://doi.org/10.1016/j.socscimed.2013.02.003>
- Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A. et Tran, N. (2013). Implementation research: what it is and how to do it. *British Medical Journal*, 347, f67-53.
<https://doi.org/10.1136/bmj.f6753>

- Peters, D. H., El-Saharty, S., Siadat, B., Janovsky, K. et Vujicic, M. (dir.). (2009). *Improving health service delivery in developing countries: from evidence to action*. Washington: World Bank.
- Petticrew, M., Cummins, S., Ferrell, C., Findlay, A., Higgins, C., Hoy, C., ... et Sparks, L. (2005). Natural experiments: an underused tool for public health?. *Public Health*, 119(9), 751-757.
- Pfadenhauer, L. M., Mozygemba, K., Gerhardus, A., Hofmann, B., Booth, A., Bakke, K. B., ... et Rehfuess, E. A. (2015). Context and implementation: A concept analysis towards conceptual maturity. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, 109(2), 103-114.
<https://doi.org/10.1016/j.zefq.2015.01.004>
- Pressman, J. L. et Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland* (3^e éd.). Berkeley, Los Angeles, London : University of California Press.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., ... et Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, 38(2), 65-76.
<https://doi.org/10.1007/s10488-010-0319-7>
- Ridde, V. (2009). Policy implementation in an African state: an extension of the Kingdon's multiple-streams approach. *Public Administration*, 87(4), 938-954.
- Ridde, V., Delormier, T. et Gaudreau, L. (2007). Evaluation of empowerment and effectiveness: Universal concepts?. Dans D. V. McQueen et C. M. Jones (dir.), *Global Perspectives on Health Promotion Effectiveness* (vol. 1, p. 389-403). New York: Springer Science et Business Media.
- Ridde, V. et Robert, É. (2014). *Real World Evaluation Strategies*. Oxford University Press.
<http://www.oxfordbibliographies.com/display/id/obo-9780199756797-0140>
- Ridde, V., Turcotte-Tremblay, A.-M., Soares, A., Lohmann, J., Zombré, D., Koulidiati, J., ... et De Allegri, M. (2014). Protocol for the process evaluation of interventions combining performance-based financing with health equity in Burkina Faso. *Implementation Science*, 9(1), 149-160.
<https://doi.org/10.1186/s13012-014-0149-1>

- Ritchie, J. et Spenser, L. (1994). Qualitative data analysis for applied policy research. Dans A. Bryman et R. G. Burgess (dir.), *Analyzing qualitative data* (p. 173-194). Londres and New York: Routledge.
- Saetren, H. (2014). Implementing the third generation research paradigm in policy implementation research: An empirical assessment. *Public Policy and Administration*, 29(2), 84-105.
<https://doi.org/10.1177/0952076713513487>
- Shoveller, J., Viehbeck, S., Di Ruggiero, E., Greyson, D., Thomson, K. et Knight, R. (2015). A critical examination of representations of context within research on population health interventions. *Critical Public Health*, volume(numéro), 114.
<https://doi.org/10.1080/09581596.2015.1117577>
- Siron, S., Dagenais, C. et Ridde, V. (2015). What research tells us about knowledge transfer strategies to improve public health in low-income countries: A scoping review. *International Journal of Public Health*, 60(7), 115.
<https://doi.org/10.1007/s00038-015-0716-5>
- Turcotte-Tremblay, A.-M., Spagnolo, J., De Allegri, M. et Ridde, V. (2016). Does performance-based financing increase value for money in low- and middle- income countries?: A systematic review. *Health Economics Review*, 6(1).
<https://doi.org/10.1186/s13561-016-0103-9>
- United Nations Secretary-General. (2010). *Global Strategy for Women's and Children's health*. Geneva : World Health Organization.
http://www.who.int/pmnch/knowledge/publications/fulldocument_globalstrategy/en/.
- Van Belle, S., van de Pas, R. et Marchal, B. (2017). Towards an agenda for implementation science in global health: there is nothing more practical than good (social science) theories. *BMJ Global Health*, 2(2), e000181.
<https://doi.org/10.1136/bmjgh-2016-000181>
- Walt, G., Shiffman, J., Schneider, H., Murray, S. F., Brugha, R. et Gilson, L. (2008). "Doing" health policy analysis: Methodological and conceptual reflections and challenges. *Health Policy and Planning*, 23(5), 308-317.
- Weiss, C. H. (1998). *Evaluation: methods for studying programs and policies* (2^e éd.). Upper Saddle River, NJ : Prentice Hall.

- Witter, S., Fretheim, A., Kessy, FL. et Lindahl, AK. (2012). Paying for performance to improve the delivery of health interventions in low- and middle-income countries. *Cochrane Database of Systematic Review*, 2, CD007899.
- Yin, R.K., et Ridde, V. (2012). Théorie et pratiques des études de cas en évaluation. Dans V. Ridde et C. Dagenais (dir.), *Approches et pratiques en évaluation de programme* (2^e éd., p. 177-193). Montréal: Presses de l'Université de Montréal.
- Zombré, D., Turcotte-Tremblay, A.-M., Sangaré, A., Zongo, S., Somé, P.-A., Coulibaly, A., ... et Ridde, V. (2016). *Méthode mixte et participative de sélection d'études de cas multiples pour l'évaluation d'intervention de santé au Burkina Faso et au Mali: De l'éthique à l'émique*. Communication présentée au Congrès de l'Association francophone pour le savoir, Montréal.
- https://docs.google.com/viewerng/viewer?url=http://www.equitesante.org/wp-content/uploads/2016/05/Poster.David_.ACFAS_.pdf&hl=fr.

Resumen / Résumé / Abstract / Résumé en moore / Résumé en dioula

Se necesitan más investigaciones sobre la aplicación para comprender cómo funcionan las intervenciones en su contexto real y cómo evolucionan con el tiempo. El objetivo de este capítulo es presentar un análisis de la aplicación de una intervención compleja, denominada financiación basada en los resultados, en Burkina Faso. Realizamos un estudio de casos múltiples (18 centros de salud) y lo contrastamos con varios niveles de análisis entrelazados. Los datos empíricos se obtuvieron a partir de observaciones, entrevistas informales y formales y documentos. Para facilitar el análisis de una gran cantidad de datos, utilizamos un proceso piramidal en el que cada miembro del equipo analizó y sintetizó los datos a un nivel diferente (centro de salud, distrito y general). De esta manera, los asistentes de investigación participaron en todo el proceso de análisis. La organización de un curso práctico sobre la utilización de programas informáticos de análisis cualitativo de libre acceso contribuyó a que se llegara a un entendimiento común del proceso de análisis. Los resultados de la investigación muestran

la importancia de abrir la “caja negra” para comprender la aplicación de las intervenciones de salud pública en África y descubrir sus adaptaciones en función de los contextos y los agentes sociales.

Plus de recherches sur la mise en œuvre sont nécessaires pour comprendre comment les interventions fonctionnent dans leur contexte réel et comment elles évoluent dans le temps. L'objectif de ce chapitre est de présenter l'analyse de la mise en œuvre d'une intervention complexe, intitulée le financement basé sur les résultats, au Burkina Faso. Nous avons effectué une étude de cas multiples (18 formations sanitaires) et contrastés avec plusieurs niveaux d'analyse imbriqués. Les données empiriques proviennent d'observations, d'entrevues informelles et formelles et de documents. Pour faciliter l'analyse d'une quantité importante de données, nous avons utilisé un processus pyramidal par lequel chaque membre de l'équipe a analysé et synthétisé des données à un niveau différent (centre de santé, district et général). Ainsi, les assistants de recherche ont participé à l'ensemble du processus d'analyse. L'organisation d'un atelier sur l'utilisation d'un logiciel d'analyse qualitative accessible gratuitement a permis de développer une compréhension commune du processus d'analyse. Les résultats de la recherche montrent l'importance d'ouvrir la « boîte noire » pour comprendre la mise en œuvre des interventions de santé publique en Afrique et mettre au jour leurs adaptations selon les contextes et les acteurs et actrices sociaux.

More research on implementation is needed to understand how interventions work in their real context and how they evolve over time. The objective of this chapter is to present the process of analyzing the implementation of a complex intervention, called performance-based financing, in Burkina Faso. We conducted a multiple and contrasted case study (18 healthcare facilities) with several levels of analysis. The empirical evidence comes from observations, informal and formal interviews and documents. To facilitate the analysis of a large amount of data, we used a pyramid process whereby each team member analyzed and synthesized data at a different level (healthcare facility, district, general). Thus, the research assistants participated in the whole process of analysis. The organization of a workshop on the utilization of a free qualitative analysis software allowed

the development of a common understanding of the analysis process. The results of the study highlight the importance of opening the “black box” to understand the implementation of interventions in global health and to shed light on their adaptations according to the contexts and the social actors.

Sen na yile ti d wum sōngrā na-kendr ā vōor la a sen teesde, tūmdā siglgl toga ne vaēsɜg ti seke. Zāms- kānga yegre yaa sen na yl n wilg bumb ning sen be sōng-kaseng sen sigend tūm- biis zug Burkina faaso ka wā . Tōnd talla bvgsg sen keed ne yel buud toor-toor logtoy-yi ziis piig la nii(18) n ti yek ne vaēsɜg bēd a taaba.

Vaēs kāng biisā tuuda ne ziig puge gesgo, sōsg ne neb la seb wvsg me karengo. Sen na yl ti veenem zind vaēs kaseng kānga zugu faag-faaga tōnd kītam ti tūm- tūmdb sen be yingr n tāag tengr kō b tagsgo (logto-yi la bēd la boonese). Rend bāngdba taoor dāmba zīnda tūmda singre n ti tāag a baasgo. Tigsg sen n kō zāmsg tūma-teed zug zaalem sōngam ti bāng sōng paam vaēsɜg zugu. Tūmda biis wilgam ti yaa tilae ti b pak koglg buud sen na n sōng ninsables zamaana laafi wa koglgo, la b le wilg sen tokd ne tūm-tūmdba fāa raabo.

Fēfē yaala ka taga lɔɲni fanfē min be na ni kow famuyali basigili ye, ka kan ka sira damanin ta min sababura famuyali kerēn kerēn be dafali sɜɜ, Nin baara nin kunsiyɜɜ, ye, ka se ka segesegeli kē ka taga kow gwelēma fēfē fanfē a be fɜ, min ma ko harajige jaga bɜ, n'o basigira tɜɜ, sɜɜli kan Burukin Faso yan.

An ye segesegeli kē koo caaman le kan dɜgɜɜɜso tan ani segi(18) le kan minw bɜlɛn tɛ yɜɜɜɜ ma.ɛɛgesegeli baara kera ka taga baro ani seben damani kɜɜ, ka se ka lɜɜni sɜɜ, o senkɜɜ, Ka se ka famuyali sɜɜ, o segesegeli kan nin kan, an ye tagamasiyen jɜɜjɜn ta minw sabababura baarakela minw bēɛ tun be jenkulu nin kɜɜ, ka se k'o lɜɜɜɜ, baara kē.(Dɜgɜɜɜɜodennin, a n'afabaw). ka tugu minw tun be lɜɜɜɜɜlagana nunu dɛmɛ ka baara segesegeli nin kē, o fana ye tali kē baara mumɛ bēɛ ra.

Yɜɜɜɜyebaara dɜ, sigi ra ka se ka lɜɜni segesegeli nin baarakeminanw labila gwansan min senkɜɜ, bēɛ ka se ka famuyali kelen sɜɜ, a ra ani o hakilinata ka kē kelen ye. Baara nin ka fēfē kera sababu ye ka gondon damani famuyali nɜgɜya ka taga jamakulu ka hɛrɛ kow lasigili kan farafina jamana kɜɜ, Baarakeminaw min bilara lɜɜɜɜɜ bolo dira o ma gwansan. A ti san

ni wari ye.egesegele nin tɛnc, lɛnnikelaw ka min basigi ka taga hɛrɛko fanfe farafina bolo kɛnc,ka se ben jamakulu ka hakilinata ye ,ani k'o ferew kerɛnkeren ta ka k'o ka sininyasigi jagabc,wala lɛnnikelaw ya yira cogomina.

Agradecemos a Alain Ouédraogo, Ahmed Sie Barro, la Sra. Namoano Solange y el Abbe Anselme Yanogo por las traducciones a los idiomas Moore y Dioula.

Valéry Ridde es Director de Investigación en el CEPED, una Unidad Conjunta de Investigación asociada a la Universidad de París Descartes y al Instituto de Investigación para el Desarrollo (IRD). Es investigador habitual del *Institut de recherche en santé publique de l'Université de Montréal* (IRSPUM) e investigador asociado del *Centre interdisciplinaire de recherche en développement international et société* (CIRDIS) de Montreal. Es editor asociado de las revistas *BMJ Global Health* y *Global Health Promotion*, miembro del consejo editorial del *Canadian Journal of Program Evaluation* y *Qualitatif! Revue africaine de sciences sociales*. Su investigación se centra en la cobertura universal de la salud, la financiación de los servicios de salud, la evaluación de programas, las políticas de salud pública y la promoción de la salud.

Anne-Marie Turcotte-Tremblay es un candidata a doctor en salud pública con una especialización en organización de cuidados en la Escuela de Salud Pública de l'Université de Montréal (ESPUM). También está afiliada al Instituto de Investigación en Salud Pública de la Universidad de Montreal (IRSPUM). Su tesis doctoral se centra en las consecuencias imprevistas de la financiación basada en los resultados combinada con medidas de protección social en Burkina Faso. Anne-Marie es becaria del Instituto Canadiense de Investigación de la Salud (CIHR). Ha coordinado varios proyectos internacionales de investigación sobre la salud de la población. Anne-Marie está interesada en la evaluación de las intervenciones para reducir las desigualdades sociales en materia de salud.

Cita

Valéry Ridde y Anne-Marie Turcotte-Tremblay (2020). El análisis de los procesos de implementación. Una intervención compleja en Burkina Faso:

la financiación basada en los resultados. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 403-431. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

17. Evaluación de la fidelidad de implantación

Un proyecto para distribuir el carbón ártico a las mujeres embarazadas en Nunavik

LARA GAUTIER, CATHERINE M. PIRKLE, CHRISTOPHER FURGAL Y MICHEL LUCAS

Evaluación de la fidelidad de implantación

Definición del método

El marco de evaluación de la fidelidad de la implantación elaborado por Carroll y sus colaboradores (2007) pone de relieve la brecha entre las intenciones iniciales de los diseñadores de los proyectos de intervención y los medios con los que se ejecutan. Cuanto más amplia sea la brecha, menor será la “fidelidad en la implantación”. Las características del proyecto y los elementos “moderadores” (por ejemplo, las estrategias de facilitación, la calidad de la entrega, la receptividad de los participantes a la intervención) pueden influir en el grado de fidelidad de la ejecución.

Fortalezas del método

- Ayuda a los/as investigadores/as y a responsables de la toma de decisiones a mejorar la ejecución de los proyectos o programas de intervención.
- Permite una evaluación sistemática de los obstáculos y los facilitadores de la ejecución de los proyectos. Esto es particularmente pertinente en la esfera de la salud de las poblaciones vulnerables debido a los altos niveles de complejidad del contexto de implantación.

Desafíos del método

- Insuficiente atención al contexto sociopolítico y al entorno en el que se lleva a cabo la intervención (Hasson 2010). Esta limitación surgió durante las entrevistas a través de la expresión de preocupaciones sobre la sostenibilidad financiera y ambiental del proyecto.
 - Un enfoque “individualista” que ignora el contexto humano y los vínculos entre las personas involucradas. Esto es aún más problemático en la evaluación de las intervenciones realizadas en el contexto indígena, en el que la frontera entre el individuo y la comunidad es borrosa.
-



Una aldea Nunavik aislada en el invierno de 2014. Créditos: Isabelle Veillette.

Con el desarrollo del campo de la evaluación de procesos, están surgiendo nuevos enfoques para explorar la caja negra de la implantación. Estos enfoques ayudan a comprender las formas en que las intervenciones logran resultados eficaces. En la esfera de la salud de los aborígenes en el Canadá, hay un creciente interés en comprender mejor el grado de implantación de las intervenciones

y la forma de mejorarlas. En particular, han surgido nuevos métodos para explorar en profundidad el grado de implementación de los programas o “proyectos” de intervención. Este estudio cualitativo utiliza un marco innovador para la evaluación de la implantación: el de Carroll y colaboradores (2007). Los autores sugieren que los medios por los que se implementan los programas pueden diferir considerablemente de las intenciones originales de los diseñadores de programas. Esta diferencia representa el grado de “fidelidad de ejecución”, que indica que cuanto mayor sea la diferencia entre las intenciones y los medios de ejecución, menor será la fidelidad de ejecución.

Desde el punto de vista conceptual, según Carroll y sus colaboradores (2007), la fidelidad en la ejecución incluye las características de un proyecto (contenido, cobertura, frecuencia y duración), así como los moderadores que pueden influir en el grado de ejecución (complejidad del proyecto, estrategias de facilitación, calidad de la ejecución y capacidad de respuesta de los participantes en el proyecto). Hemos adaptado este marco teórico para evaluar la fidelidad de la implantación en el contexto del proyecto de distribución del carbón del Ártico en Nunavik. Se compone de seis elementos temáticos clave. Los cinco primeros se refieren a los diferentes moderadores, mientras que el sexto se refiere a los indicadores del proyecto:

- el impacto esperado del proyecto
- la complejidad inherente de la intervención (especialmente la logística)
- estrategias de comunicación y coordinación para la ejecución del proyecto

- la calidad de la prestación de la intervención (información sobre la diferencia entre lo que se supone que los beneficiarios deben recibir y lo que realmente reciben)
- la receptividad al proyecto, es decir, la forma en que los actores involucrados en el proyecto lo aprecian. La receptividad se analiza no sólo entre los beneficiarios, sino también entre las personas o grupos de personas responsables de la ejecución del proyecto y las organizaciones implicadas: ¿en qué medida influyen sus características estructurales (liderazgo, recursos humanos, rendición de cuentas, financiación, etc.)?
- la correspondencia entre las características del proyecto (cobertura del proyecto, dosis, frecuencia, duración, etc.) y la realidad.

Estos seis temas constituyen los elementos básicos para medir la fidelidad de implantación del proyecto evaluado en este capítulo.

Contexto y proyecto evaluado

Descripción del contexto en el que se desarrolla el proyecto

Los trastornos socioeconómicos y ambientales han afectado a las comunidades aborígenes del norte del Canadá en los últimos 60 años. En particular, estos cambios han tenido un impacto negativo en muchos sistemas alimentarios basados en productos locales de la caza, la pesca y la recolección. En la actualidad, entre las poblaciones aborígenes de los países industrializados, las comunidades inuit del Canadá son las que tienen los niveles más altos de inseguridad alimentaria (Consejo de Academias Canadienses 2014). Además, la exposición a contaminantes ambientales, como el mercurio, amenaza la seguridad alimentaria de esas poblaciones (Pirkle, Muckle y Lemire 2016; Van Oostdam y otros, 2005). Dada una dieta que incluye carne de mamíferos marinos, están desproporcionadamente expuestos al mercurio en comparación con los canadienses del sur (Pirkle, Muckle y Lemire 2016). El mercurio es peligroso para el feto en desarrollo y también tiene efectos negativos a largo plazo en el desarrollo del niño (Boucher y otros, 2012; Jacobson y otros, 2015; Pirkle, Muckle y Lemire, 2016).

Es necesario realizar actividades de promoción de la salud basadas en los sistemas alimentarios de los inuit y reducir la exposición a los

contaminantes ambientales. Algunos alimentos locales (por ejemplo, la carne de beluga) contienen altos niveles de nutrientes, pero también pueden contener altos niveles de mercurio (Lemire et al. 2015). Una opción de intervención es promover el consumo de alimentos inuit locales ricos en nutrientes pero pobres en contaminantes ambientales. El trucha alpina (*Salvelinus alpinus*), un pez de América del Norte, es una fuente de alimento que cumple precisamente estos criterios.



Folleto distribuido como parte del Día de la Alimentación Tradicional, 2014. Créditos de la foto: Isabelle Veillette.

Descripción de la lógica del proyecto

En Nunavik, una región del norte de Quebec (Canadá), existe un proyecto comunitario que tiene por objeto mejorar la calidad nutricional del suministro de alimentos, reducir la inseguridad alimentaria y minimizar la exposición a los contaminantes ambientales. El proyecto es sencillo en su concepto: consiste en distribuir gratuitamente carbón del Ártico a las mujeres embarazadas que viven en las aldeas Nunavik, habitadas principalmente por personas de origen inuit. Se conoce como el Proyecto de

Distribución del Carbón Ártico (AC/DP). En la década de 2000, el proyecto se probó en tres comunidades Nunavik. En los años siguientes, se intentó expandir el proyecto a todos los pueblos de la costa de la Bahía de Hudson.

En septiembre de 2011, diez años después de su primera iteración, la Junta Regional de Salud y Servicios Sociales de Nunavik (NRBHSS) quiso poner en marcha el proyecto de distribución de carbón del Ártico para todas las mujeres embarazadas que viven en las aldeas de la costa de la bahía de Hudson. Teóricamente, el pescado es capturado por pescadores locales en esta zona y luego se vende ligeramente por debajo del precio de mercado a la Junta. Luego se coloca en neveras (grandes, difíciles de manejar) para evitar el riesgo de quemaduras por congelación. Los refrigeradores son

transportados en avión por Air Inuit. Los inuits aéreos transportan las cajas de pescado a las aldeas de los alrededores gratuitamente y, en teoría, llevan las cajas vacías a las aldeas de abastecimiento por la modesta suma de 25 dólares canadienses. Las maternidades de estas aldeas deben organizar la distribución de pescado fresco a las mujeres embarazadas. El plan es que todas las mujeres embarazadas reciban un pez de 24 pulgadas o más una vez a la semana en el hospital de maternidad o en sus casas (en los pueblos más pequeños). Este proyecto tiene por objeto reducir la exposición a los contaminantes del medio ambiente y mejorar la seguridad alimentaria, dos importantes cuestiones de salud pública en las comunidades inuit de Nunavik.

Sin embargo, la implantación del proyecto está incompleta. El pescado no se distribuye regularmente (menos de una vez por semana) y no se cubren todas las zonas geográficas seleccionadas. Además, su sostenibilidad ambiental y financiera está amenazada.

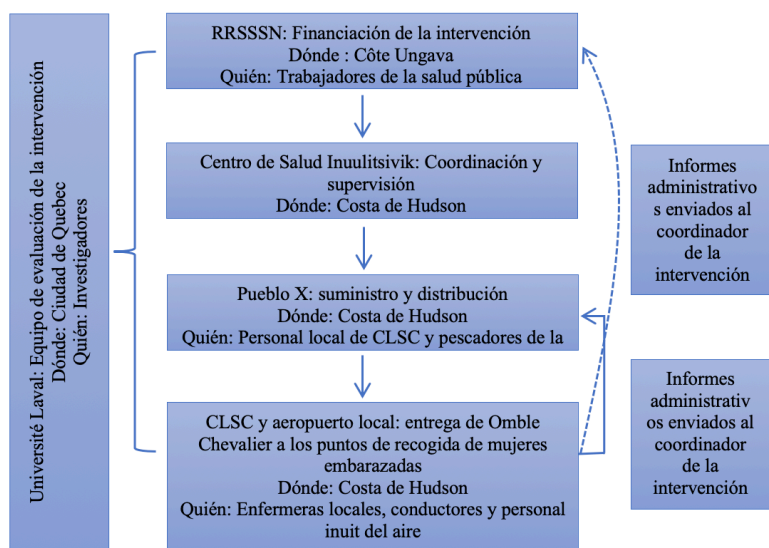


Figura 1. Diagrama de la l gica del proyecto y sus componentes

Adaptación del marco teórico al estudio de casos

Hasson (2010) constató que los componentes del marco de Carroll y colaboradores (2007) no tenían en cuenta el contexto socioeconómico, cultural y ambiental en el que se ubica un proyecto (Hasson 2010). Por lo tanto, creamos un tema adicional, que reflejaba las diferentes dimensiones que se emitieron, que probablemente afectaría a la implantación del proyecto: “Valores asociados a la intervención” (véase la sección 4 infra). Los “valores” reflejan los diferentes preconceptos naturalmente asociados al proyecto por las diferentes categorías de participantes. Se refieren al contexto macrosocial, económico, cultural y ambiental en el que se realiza la intervención. Los participantes hablan de varios tipos de valores asociados a la intervención:

- el valor de la “salud pública” (mejora del estado nutricional y la seguridad alimentaria)
- el valor “basado en la comunidad” (intervención de abajo hacia arriba, implantación a nivel de la comunidad)
- el valor “tradicional” (abogando por los alimentos tradicionales en la dieta)
- el valor “salud materna” (con referencia a la población destinataria: las mujeres embarazadas)
- el valor de la “conexión con los ancianos” (las generaciones mayores aconsejan a las mujeres embarazadas que coman pescado)

Diseño de la evaluación

Propósito de la evaluación

En este estudio, llevamos a cabo una evaluación de la fidelidad de la implantación del proyecto AC/DP. Concretamente, examinamos cómo los siete elementos para medir la fidelidad de la implantación afectan al logro de los objetivos originales de este proyecto, que se inició hace 15 años.

Proceso

Evaluamos la implantación del proyecto sobre la base de los datos recogidos en los documentos del proyecto, las actas de las reuniones, las notas de campo y las entrevistas cualitativas con los beneficiarios y los ejecutores del proyecto. Los temas emergentes de los datos se examinan a la luz del marco de fidelidad de la implantación elaborado por Carroll y sus colaboradores en 2007.

Tabla 1 : Resumen de la estimación de la evaluación de la implantación del proyecto de distribución del carbón ártico para mujeres embarazadas en Nunavik

Fuentes de datos	Descripción
Entrevistas individuales	<p>Guía de entrevistas para los beneficiarios del proyecto – mujeres embarazadas (de cualquier edad), explorando su nivel de aceptación del proyecto.</p> <p>Guía de entrevistas para los encargados de la adopción de decisiones y de la ejecución – encargados de la adopción de decisiones y de la ejecución, explorando las percepciones en torno a la visión, la financiación, la operatividad y la sostenibilidad a largo plazo del proyecto).</p>
Documentación	Viajes al campo (segundo autor) para recoger documentos.
Entrevistas individuales	<p>13 transcripciones literales – entrevistas con 14 mujeres embarazadas (una entrevista se realizó con dos mujeres).</p> <p>4 transcripciones literales – entrevistas con los responsables de la toma de decisiones y los ejecutores.</p>
Documentación	<p>7 documentos: notas de observación, informes de reuniones y documentos de proyectos.</p> <p>6 intercambios de correo electrónico</p>
Transcripción y temas	Tras la transcripción de todos estos datos, se llevó a cabo un enfoque de codificación estándar (deductivo-inductivo). Los códigos fueron revisados en colaboración por el primer y segundo autores. Los posibles temas se debatieron a la luz de los objetivos de la investigación y se establecieron por consenso.

Breve descripción de los resultados de la evaluación

Nuestros resultados indican que las mujeres embarazadas aceptaron plenamente el proyecto debido a su relevancia cultural. Sin embargo, hay una implantación incompleta del proyecto en las siete dimensiones:

1. **Impacto previsto del proyecto:** falta de una visión común del impacto previsto del proyecto: un proyecto destinado a la seguridad alimentaria para algunos o un proyecto destinado a reducir la exposición a los contaminantes para otros.
2. **Complejidad inherente a la intervención:** logística (pescado sólo disponible en determinadas épocas del año, transporte a los beneficiarios y mantenimiento de la cadena de frío) y multiplicidad de agentes encargados de la ejecución (gestores, coordinadores, enfermeros, etc.).
3. **Estrategias de comunicación y coordinación:** falta de estrategias y dificultades de comunicación y coordinación entre los responsables de la ejecución de los proyectos.
4. **Calidad de la prestación de la intervención:** la frescura era a veces insatisfactoria, ya que el pescado no llegaba regularmente. Por otra parte, el tamaño de los peces siempre parecía satisfactorio, incluido el hecho de que los beneficiarios podían elegir un tamaño que correspondiera a sus deseos.
5. **Receptividad al proyecto:** Los receptores informaron de altos niveles de receptividad al proyecto, en parte porque es un recurso gratuito para ellos. El proyecto también fue bien percibido por la mayoría de los profesionales de la salud, que eran conscientes de la utilidad de esa intervención de salud pública en esas comunidades. Por otra parte, en lo que respecta a algunos de los agentes políticos, los administradores y el personal que participan directamente en la ejecución, se puede hablar de una falta de compromiso y de visión a largo plazo: el liderazgo no siempre estaba a la altura (a cualquier nivel), los recursos humanos fluctuaban y la estructura de financiación del proyecto era inestable.
6. **Vínculos correspondientes entre las características del proyecto:** el pescado era siempre de tamaño satisfactorio, pero con una cobertura geográfica parcial e irregularidades en la distribución (el pescado no llegaba a los beneficiarios una vez por semana por razones principalmente logísticas).
7. **Valores asociados:** intervención cultural y socialmente adaptada al contexto, especialmente porque integra los códigos tradicionales (referencia a compartir, “comida local”, antepasados, etc.). La intervención también es válida desde el punto de vista de la salud pública (nutrientes y reducción de la exposición a los contaminantes). Posibles riesgos desde el punto de vista ambiental, relacionados con la posible

sobreexplotación de los peces.

Tabla 2 : Retrato resumido de los puntos fuertes y débiles del proyecto

Fortalezas del proyecto	Debilidades del proyecto
<ul style="list-style-type: none"> • Apreciadas y solicitadas por los beneficiarios (cf. n° 5) • Pertinencia cultural para las comunidades beneficiarias (cf. n° 7) • Rico en nutrientes (ver n° 7) • Bajo en contaminantes (ver n° 7) • Aceptado por la mayoría de los profesionales de la salud (cf. n° 5) • Costo limitado (véase el n° 5) 	<ul style="list-style-type: none"> • Complejidad logística, incluyendo discontinuidades en la cadena de transporte y distribución para asegurar la distribución de pescado fresco a los beneficiarios del proyecto (ver n° 6) • Dificultades de comunicación y coordinación entre los responsables de la implantación del proyecto (véase el n° 3) • Falta de una visión común en torno a los efectos previstos del proyecto (es decir, proyecto destinado a la inocuidad de los alimentos para algunos y a la reducción de la exposición a los contaminantes para otros) (véase el n° 1) • Insuficiente sostenibilidad financiera a largo plazo (véase el n° 5) • Los frecuentes cambios en los recursos humanos (en particular en las salas de maternidad y entre los transportistas) impiden la participación a largo plazo de las personas responsables de la implantación del proyecto. • Estacionalidad del pescado y frescura irregular del recurso vinculada a la manipulación (cf. n° 2) • Posible vulnerabilidad de los peces a la sobreexplotación (véase el n° 7)

Discusión de los resultados

Nuestros resultados muestran que el proyecto fue apoyado y apreciado por muchos actores y actrices. Sin embargo, esta labor pone de relieve el hecho de que la complejidad de un proyecto puede obstaculizar su ejecución, en particular en lo que respecta a la comunicación y la coordinación.

La logística (transporte, envío, almacenamiento, distribución) necesaria para ejecutar el proyecto requiere un alto nivel de coordinación entre los interesados, como las enfermeras y los transportistas comunitarios, en particular porque el pescado es rápidamente perecedero.

La falta de compromiso de ciertos interesados ha causado dificultades en la implantación, así como frecuentes cambios de personal en la región de Nunavik. Además de la rotación del personal inuit del aire y de la maternidad, los coordinadores del proyecto han cambiado a menudo de puesto. Las frecuentes rotaciones del personal impidieron el desarrollo de redes coordinadas de personas comprometidas a largo plazo y dispuestas a superar las dificultades logísticas que supone la distribución de recursos en una zona remota. Por ello, la fiabilidad del proyecto se deterioró con el tiempo: los principales interesados a menudo no pudieron hacer comentarios sobre la disponibilidad de peces. Estos problemas encontrados en Nunavik se repiten en el África subsahariana, donde también es común la inestabilidad del personal. En varias evaluaciones realizadas en el África subsahariana se informa de frecuentes cambios de personal y condiciones de trabajo difíciles tras la ejecución de intervenciones sanitarias específicas en el Senegal y Sudáfrica, por ejemplo (Rouleau y otros, 2012; Van Damme, Kober y Kegels, 2008). Estas situaciones afectan fuertemente a la fiabilidad de los proyectos.

Los problemas estructurales también contribuyen a la falta de comunicación entre los centros del proyecto, en particular los coordinadores del proyecto y las enfermeras de maternidad, incluso cuando trabajan en estrecha proximidad. También se identificaron problemas de comunicación en la evaluación de la implantación de otra intervención en materia de nutrición, la Nutrición del Norte del Canadá (Gobierno del Canadá 2013). Las investigaciones de este fenómeno en otros proyectos o programas de nutrición en áreas remotas de Canadá deben continuar en futuras investigaciones.

Es probable que la complejidad de la intervención no sea la única causa de una implantación incompleta. De hecho, como señaló un responsable de la RRSSSN, la distribución de carbón del Ártico a las mujeres embarazadas de esta región nunca se ha considerado un “programa formal” dirigido por la RRSSSN. Más bien, la intervención se consideró una simple “iniciativa”, apoyada ocasionalmente por el RRSSSN. Según este responsable, esto explicaría las dificultades de implantación y la falta de visibilidad a largo plazo.

Además, la falta de una visión común en torno al impacto previsto del proyecto provocó, en nuestra opinión, una dilución de la responsabilidad de las partes interesadas. Sin embargo, el compromiso de alcanzar un objetivo común es la clave del éxito del proyecto. En varios casos, los agentes que

participaban en la implantación no estaban preparados para tomar la iniciativa de encontrar la fuente de los problemas de distribución. El personal de la maternidad solía depender de los coordinadores de proyectos y de otras personas en puestos de autoridad para resolver cuestiones logísticas. El personal también demostró un compromiso limitado debido a su escaso conocimiento de la financiación de los proyectos y a la insuficiente retroalimentación de los resultados de las reuniones a las instituciones asociadas a los proyectos. Por consiguiente, la responsabilidad de todos los interesados ante los beneficiarios del proyecto se vio dificultada por la falta de liderazgo.

Creemos que el marco desarrollado por Carroll y sus colaboradores (2007) es una herramienta útil y valiosa para guiar a los investigadores y a los responsables de la toma de decisiones para mejorar la implantación de los proyectos o programas de intervención. Concebido como una “receta” que sólo debe seguirse en cada etapa, el marco ha demostrado ser útil para evaluar sistemáticamente los obstáculos y los facilitadores de la ejecución de los proyectos. Esto es particularmente pertinente en la esfera de la salud de los aborígenes debido a los altos niveles de complejidad del contexto de implantación. Era necesario profundizar en estos procesos a fin de informar mejor la toma de decisiones.



Las costas de la aldea de Puvirnituk en el invierno de 2014. Créditos de la foto: Isabelle Veillette.

Sin embargo, la ausencia de “valores” y elementos contextuales en el marco de Carroll y sus colaboradores (2007) apunta a una visión tecnocrática de la evaluación de la ejecución de proyectos o programas que tiene por objeto equilibrar los facilitadores y los obstáculos. Un problema importante es que el marco ignora

casi por completo el contexto humano y los vínculos entre las personas involucradas. En particular, no tiene en cuenta la necesidad de contar con personas motivadas en el diseño y la implantación de proyectos o políticas. Sin embargo, Durlak y DuPre (2008) habían demostrado al mismo tiempo que los factores relacionados con la capacidad organizativa (motivación del personal que realiza la intervención, liderazgo, seguimiento y apoyo de los superiores, etc.) eran esenciales para garantizar la plena implantación

(Chouinard y Cousins 2007). Esto plantea un desafío para la implantación del marco en el contexto de la salud de los aborígenes. Conectar a la gente es primordial, ya que el colectivo es muy importante. Por lo tanto, en el contexto aborigen, es necesario reconocerlo en cualquier evaluación que se realice.

En general, el proyecto KT/PD no adolece de falta de apoyo de los interesados, aunque los objetivos específicos a menudo se han puesto en tela de juicio. El desafío consiste más bien en aprovechar la popularidad del proyecto para superar las barreras logísticas y de comunicación.

Referencias clave

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40-48.

<https://doi.org/10.1186/1748-5908-2-40>

El artículo principal de Carroll y sus colegas que detalla el famoso marco teórico de la fidelidad al sitio: leer y releer.

Gautier, L., Pirkle, C. M., Furgal, C. et Lucas, M. (2016). Assessment of the implementation fidelity of the Arctic Char Distribution Project in Nunavik, Quebec. *BMJ Global Health*, 1(3), e000093.

<https://doi.org/10.1136/bmjgh-2016-000093>

Nuestro artículo de investigación original, publicado en inglés, en el que describimos los diversos componentes del programa de distribución de carbón del Ártico en Nunavik y los analizamos a la luz del marco de fidelidad al sitio.

Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5(1), 67-75.

<https://doi.org/10.1186/1748-5908-5-67>

La adaptación de Hasson del marco teórico de la fidelidad de implantación, teniendo en cuenta de manera más convincente el contexto sociopolítico y el entorno en el que se implanta una intervención compleja.

Referencias

- Boucher, O., Jacobson, S. W., Plusquellec, P., Dewailly, E., Ayotte, P., Forget-Dubois, N., ... et Muckle, G. (2012). Prenatal methylmercury, postnatal lead exposure, and evidence of attention deficit/hyperactivity disorder among Inuit children in Arctic Quebec. *Environ Health Perspectives*, 120(10), 1456-1461.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40-48.
<https://doi.org/10.1186/1748-5908-2-40>
- Chouinard, J. A. et Cousins, J. B. (2007). Culturally competent evaluation for Aboriginal communities: A review of the empirical literature. *Journal of Multidisciplinary Evaluation*, 4(8), 40-57.
- Council of Canadian Academies. Expert Panel on the State of Knowledge of Food Security in Northern Canada. (2014). *Aboriginal Food Security in Northern Canada: An Assessment of the State of Knowledge*.
<http://www.scienceadvice.ca/en/assessments/completed/food-security.aspx>
- Durlak, J. A. et DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350.
- Gautier, L., Pirkle, C. M., Furgal, C. et Lucas, M. (2016). Assessment of the implementation fidelity of the Arctic Char Distribution Project in Nunavik, Quebec. *BMJ Global Health*, 1(3), e000093.
<https://doi.org/10.1136/bmjgh-2016-000093>
- Government of Canada. (2013). *Implementation Evaluation of the Nutrition North Canada Program*. Site de Aboriginal Affairs and Northern Development Canada.
<http://www.aadnc-aandc.gc.ca/eng/1395347953550/1395348287432>
- Hasson, H. (2010). Systematic evaluation of implementation fidelity of complex interventions in health and social care. *Implementation Science*, 5, 67-75.
<https://doi.org/10.1186/1748-5908-5-67>

- Jacobson, J. L., Muckle, G., Ayotte, P., Dewailly, É. et Jacobson, S. W. (2015). Relation of Prenatal Methylmercury Exposure from Environmental Sources to Childhood IQ. *Environmental Health Perspectives*, 123(8), 827-833.
- Lemire, M., Kwan, M., Laouan-Sidi, A. E., Muckle, G., Pirkle, C., Ayotte, P. et Dewailly, E. (2015). Local country food sources of methylmercury, selenium and omega-3 fatty acids in Nunavik, Northern Quebec. *Science of the Total Environment*, 509, 248-259.
- Pirkle, C. M., Muckle, G. et Lemire, M. (2016). Managing mercury exposure in northern Canadian communities: State of the evidence and guidance for health professionals. *Canadian Medical Association Journal*, 188(14), 1015-1023.
- Rouleau, D., Fournier, P., Philibert, A., Mbengue, B. et Dumont, D. (2012). The effects of midwives' job satisfaction on burnout, intention to quit and turnover: a longitudinal study in Senegal. *Human Resources for Health*, 10(1), 9-22.
- Van Damme, W., Kober, K. et Kegels, G. (2008). Scaling-up antiretroviral treatment in Southern African countries with human resource shortage: How will health systems adapt?. *Social Science et Medecine*, 66(10), 2108-2121.
- Van Oostdam, J., Donaldson, S. G., Feeley, M., Arnold, D., Ayotte, P., Bondy, G., ... et Loring, E. (2005). Human health implications of environmental contaminants in Arctic Canada: A review. *Science of the Total Environment*, 351, 165-246.

Resumen / Résumé / Abstract / Résumé en inuktituk

En 2011, la Junta Regional de Salud y Servicios Sociales de Nunavik comenzó a proporcionar apoyo financiero al Arctic Char Distribution Project (AC/DP) para mujeres embarazadas. Esta iniciativa promovió el consumo de alimentos tradicionales de los inuit, el salmón del ártico, para las mujeres embarazadas que viven en las aldeas de Nunavik, una región del norte de Quebec poblada principalmente por personas de origen inuit. Esta intervención tenía por objeto reducir la exposición a los contaminantes ambientales y mejorar el estado nutricional y la seguridad alimentaria. La

ejecución del proyecto se evaluó sobre la base de los datos recogidos de los documentos del proyecto, las notas sobre el terreno y las entrevistas cualitativas con los beneficiarios y ejecutores del proyecto. Los temas que surgieron de los datos fueron discutidos a la luz del marco de fidelidad de implementación desarrollado por Carroll et al. (2007). Las mujeres embarazadas han adoptado plenamente la iniciativa por su relevancia cultural. Sin embargo, la ejecución del proyecto fue incompleta: no abarcó todas las zonas geográficas previstas y había una incoherencia recurrente en el suministro y la distribución de pescado. Además, la iniciativa no se financió sobre una base estable. Este trabajo muestra que la complejidad del proyecto puede dificultar una ejecución exitosa, especialmente en términos de comunicación y coordinación.

En 2011, la Régie régionale de la santé et des services sociaux du Nunavik a commencé à appuyer financièrement le Projet de Distribution d'Omble chevalier (*Arctic Char Distribution Project* ou AC/DP dans son acronyme en anglais) pour les femmes enceintes. Cette initiative favorisait la consommation d'une nourriture traditionnelle inuite – le poisson « Omble chevalier » – pour les femmes enceintes vivant dans les villages du Nunavik, une région du nord du Québec peuplée majoritairement de personnes d'origine inuite. Cette intervention visait à réduire l'exposition aux contaminants environnementaux et à améliorer l'état nutritionnel et la sécurité alimentaire. La mise en œuvre du projet a été évaluée sur la base des données collectées à partir de documents du projet, de notes de terrain et d'entretiens qualitatifs avec les bénéficiaires et les exécutants du projet. Les thèmes émergeant des données ont été discutés à la lumière du cadre de fidélité à la mise en œuvre développé par Carroll et coll. (2007). Les femmes enceintes ont pleinement adopté l'initiative pour sa pertinence culturelle. Toutefois, la mise en œuvre du projet était incomplète : elle ne couvrait pas toutes les zones géographiques prévues, et il y avait une incohérence récurrente dans l'approvisionnement et la distribution du poisson. En outre, l'initiative n'était pas financée de façon stable. Ce travail montre que la complexité du projet est susceptible d'entraver la réussite de la mise en œuvre, en particulier en termes de communication et de coordination.

Résumé en inuktituk

[illegible][illegible]

[illegible][illegible]

Lara Gautier es candidata al doctorado en salud pública, opción de salud global, en la Escuela de Salud Pública de la Universidad de Montreal. Formada en ciencias políticas y economía de la salud, Lara se interesa por las cuestiones de la gobernanza y el poder en las intervenciones sanitarias mundiales, utilizando un enfoque interdisciplinario para el análisis de los procesos políticos, socioculturales y organizativos. Sus áreas de investigación incluyen el África occidental francófona y las poblaciones vulnerables de Quebec y Francia. También participa en el sector asociativo, prestando apoyo social a los migrantes en Montreal y París. Lara ha publicado más de quince artículos científicos y capítulos de libros. Tiene una beca de investigación doctoral de la FRQ Société et Culture, y enseña salud global, evaluación de la implantación y métodos cualitativos y mixtos en Quebec, Malí y varios países europeos.

Catherine Pirkle es Profesora Adjunta en la Oficina de Estudios de Salud Pública de la Universidad de Hawaï'i a Mānoa. Es una de las principales investigadoras del equipo de evaluación de la Iniciativa Healthy Hawai. Esta iniciativa del Estado hawaiano, que comenzó en 2000, ha llevado a cabo varias intervenciones importantes en materia de salud pública, entre ellas la promoción de la actividad física, la alimentación sana y el fortalecimiento de los sistemas de salud. Financiado por el Departamento de Salud de Hawai y los Centros de Control de Enfermedades de los Estados Unidos, el equipo de Pirkle y sus colaboradores es responsable de evaluar las intervenciones clave de la Iniciativa Healthy Hawai. Las áreas de investigación del Dr. Pirkle son la salud global, la salud de la mujer y el envejecimiento. Además de su trabajo para el equipo de evaluación de la Iniciativa Healthy Hawai, es la

investigadora principal o coinvestigadora de varios proyectos financiados por los Institutos Nacionales de Salud de los Estados Unidos y los Institutos Canadienses de Investigación sobre la Salud.

Christopher Furgal es profesor asociado del Programa de Estudios y Ciencias Ambientales Indígenas de la Universidad de Trent. Tiene formación en ciencias naturales, sociales y de la salud. Es nombrado por la Escuela de Estudios Nativos Chanie Wenjack y la Escuela de Medio Ambiente de Trent. Es Director del Grupo de Investigación sobre Salud, Medio Ambiente y Comunidades Aborígenes y Director del Centro Nasivvik para la Salud de los Inuit y los Entornos Cambiantes. El Dr. Furgal está interesado en la salud ambiental de los aborígenes y en la interacción entre la ciencia y el conocimiento indígena, particularmente en el Ártico canadiense. Ha trabajado durante 25 años con comunidades aborígenes (en particular con las comunidades inuit) y organizaciones sobre cuestiones de salud ambiental, como el cambio climático y la salud, la seguridad alimentaria, la exposición a contaminantes y la percepción de riesgos. Es coautor o autor principal de varias evaluaciones nacionales e internacionales de la salud ambiental.

Michel Lucas es profesor asociado del Departamento de Medicina Social y Preventiva de la Facultad de Medicina de la Universidad Laval, investigador del CHU de Québec-Université Laval e investigador visitante de la Escuela de Salud Pública T.H. Chan de Harvard. El área de investigación del Dr. Lucas es la calidad de los alimentos, más específicamente los índices multidimensionales de calidad de los alimentos. Sus actividades de investigación sobre la salud mundial se centran en el efecto general de la dieta en la salud de las poblaciones, basándose en datos de diversos estudios de población (niños, adultos canadienses, aborígenes de Quebec, polinesios, etc.). Como investigador principal designado, dirige el gran proyecto de intervención VisezEau®, que reúne a 14 investigadores de diversas universidades de Quebec, Ontario y los Estados Unidos. El Dr. Lucas ha recibido varias becas y premios en reconocimiento a su trabajo. Es miembro de varios consejos editoriales y revisores de prestigiosas revistas científicas y ha producido más de 300 publicaciones y presentaciones.

Cita

Lara Gautier, Catherine M. Pirkle, Christopher Furgal y Michel Lucas (2020). Evaluación de la fidelidad de implantación. Un proyecto para distribuir el carbón ártico a las mujeres embarazadas en Nunavik. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 433-451. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

18. Evaluación de la fidelidad y la adaptación

Un ejemplo de la puesta en marcha de las intervenciones en salud global

DENNIS PÉREZ, MARTA CASTRO Y PIERRE LEFÈVRE

Evaluación de la fidelidad y la adaptación

Definición del método

Se trata de un análisis holístico que examina la forma en que los actores implementan las intervenciones según lo previsto o la forma en que provocan el cambio. Esto permite una aplicación de calidad y eficaz.

Fortalezas del método

- Permite identificar los componentes esenciales que deben aplicarse para que la intervención produzca los efectos previstos.
- Permite un equilibrio adecuado entre la fidelidad y la adaptación en el caso de las intervenciones adaptables.
- Identifica las estrategias de facilitación para hacer frente a las adaptaciones inadecuadas.
- Ayuda a distinguir entre el fracaso de la intervención (falta de eficacia) y el fracaso de la aplicación.

Desafíos del método

- Cuando no se dispone de un método normalizado para evaluar la fidelidad y las medidas propuestas son limitadas y a menudo específicas de los programas que se están evaluando.
 - Falta de consenso sobre cómo poner en práctica las adaptaciones.
-

El presente capítulo trata de la evaluación de la fidelidad y las adaptaciones en la aplicación de las intervenciones sanitarias mundiales. Se centra en las intervenciones adaptables. La mayoría de las intervenciones sanitarias mundiales son susceptibles de adaptación porque son complejas, abordan problemas de gran envergadura, implican diferentes niveles de organización y se dirigen a comportamientos colectivos. La supervisión de su aplicación por los diseñadores es a menudo imposible o incluso indeseable. Además, dependen del contexto, ya que algunos programas pueden funcionar bien en algunos contextos y fallar en otros.

La fidelidad con que se aplica una intervención puede afectar a su efectividad. Por lo tanto, lograr un alto nivel de fidelidad es una preocupación importante cuando se pasa de la investigación sobre la eficacia de una intervención en condiciones ideales a su aplicación en el mundo real. Cuando se pone en práctica una intervención, es más que probable que se produzcan adaptaciones. El control total del proceso de aplicación será a menudo imposible o incluso deseable, y será necesario adaptar muchas intervenciones.

Al igual que la fidelidad, las adaptaciones pueden tener un impacto positivo o negativo en la eficacia de una intervención sanitaria mundial. Por lo tanto, la evaluación de ambos es importante.

La ejecución es un conjunto específico de procesos y actividades destinados a poner en práctica una intervención, cuyas dimensiones se conocen (Fixsen y otros, 2005). Para evaluarlo, los resultados de la aplicación deben considerarse conceptual y empíricamente distintos de los utilizados para evaluar la eficacia de la intervención (Proctor y otros 2011; Peters y otros 2013). La distinción entre los efectos de la aplicación y los efectos de una intervención es crucial. Cuando una intervención fracasa, esta diferenciación ayuda a determinar si se debe a la ineficacia de la propia intervención (fracaso de la intervención), o si se ha aplicado incorrectamente (fracaso de la aplicación) (Proctor et al. 2011).

La fidelidad, o el grado en que una intervención se aplica según lo previsto por sus diseñadores (Dane y Schneider 1998; Dusenbury y otros, 2003; Schwarz y otros, 2014; Sundell y otros, 2015), es un resultado del proceso de aplicación (Proctor y otros, 2011). Está diseñado específicamente para asegurar que la intervención mantenga los efectos previstos (Dane y Schneider 1998; Backer 2002; Dusenbury y otros 2003; Schwarz y otros 2014; Sundell y otros 2015). Esta noción se acerca a términos como integridad (Dane y Schneider 1998, Dusenbury et al. 2003), fundación (Fixsen et al. 2005, Rebchook et al. 2006, Schwarz et al. 2014, Gagliardi et al. 2014, Sundell et al. 2015) o los componentes esenciales de la intervención (Carroll et al. 2007), es decir, los que deben aplicarse necesariamente para que la intervención produzca sus efectos.

La adaptación, en cambio, es la modificación del diseño original de una intervención (Backer 2002; Schwarz et al. 2014; Sundell et al. 2015). En la teoría de la difusión de las innovaciones (Rogers 2003), la adaptación se

entiende como el resultado de los procesos cognitivos que se producen cuando los actores tienen dificultades para dar sentido a una intervención durante su aplicación. En esta interpretación del término, este tipo específico de adaptación, que va más allá de la simple adaptación a las características locales, se denomina muy a menudo reinención.

Las adaptaciones pueden ser deliberadas o accidentales e incluyen varias modalidades: la adición de nuevos componentes, la eliminación o la modificación radical de un componente de la intervención de modo que ya no se parezca al original, o modificaciones menores o mayores de un componente (Backer 2002; Rebchook et al. 2006).

La fidelidad con que se aplica una intervención afecta a su eficacia (Dane y Schneider 1998; Backer 2002; Dusenbury y otros 2003; Carroll y otros 2007; Schwarz y otros 2014; Sundell y otros 2015). En consecuencia, el logro de un alto nivel de fidelidad ha sido la principal preocupación de muchos investigadores que se han esforzado por pasar de los estudios sobre la eficacia de las intervenciones en condiciones ideales a su aplicación en el mundo real (Bellg y otros, 2004; Borrelli, 2011). Sin embargo, en la práctica, la adaptación de las intervenciones es la regla y no la excepción (Sundell y otros, 2015). Además, algunas intervenciones deben adaptarse necesariamente durante la aplicación (Backer 2002; Schwarz y otros 2014; Sundell y otros 2015). Este es el caso de las intervenciones adaptables.

Las intervenciones adaptables son aquellas en las que se permite, o incluso se alienta, a los agentes de ejecución a que hagan modificaciones en el modelo inicial. Esta definición incluye el tipo de intervenciones adaptables, según la definición de Collins (et al. 2004), para las que se permiten cambios predeterminados y anticipados por los diseñadores de la intervención. En este tipo particular de intervención, la fidelidad es importante para asegurar que las adaptaciones predefinidas se produzcan según lo previsto. Sin embargo, las intervenciones de adaptación también incluyen aquellas en que los cambios no predefinidos son inducidos por actores de primera línea. En los últimos 15 años, se ha informado de la existencia de cinco dimensiones para medir la fidelidad: i) adhesión – la aplicación del programa según lo previsto, ii) dosis – la frecuencia y la duración de la exposición al programa, iii) calidad de la prestación – la forma en que se imparte el programa, iv) capacidad de respuesta de los participantes – el grado de compromiso de los participantes, y v) diferenciación del programa – las características críticas que distinguen

al programa (Dane y Schneider 1998; Backer 2002; Dusenbury et al. 2003; Carroll y otros, 2007; Schwarz y otros, 2014; Sundell y otros, 2015). Sin embargo, hay pocas investigaciones u orientaciones prácticas sobre la forma de adaptar una intervención manteniendo los componentes y mecanismos esenciales para su eficacia (Schwarz y otros, 2014).

Sobre la base de evaluación sistemática y crítica de las conceptualizaciones existentes sobre la fidelidad en la aplicación, Carroll y sus colegas (2007) propusieron un marco conceptual para dar significado al concepto de fidelidad, pero también para aclarar y explicar la función de cada una de las cinco dimensiones clásicas y sus vínculos. En su marco conceptual, también incluyeron dos elementos adicionales: la complejidad de las intervenciones y las estrategias de facilitación.

Según este marco, “la medida de la fidelidad en la ejecución es la medida de la adhesión” (Carroll y otros, en línea), es decir, la medida en que los responsables de la ejecución de la intervención se adhieren a la misma según lo descrito por sus diseñadores. La adhesión incluye el contenido, la frecuencia, la duración y la cobertura (dosis). El nivel de logro del contenido esperado o de la frecuencia deseada es el grado de fidelidad de la aplicación logrado para esa intervención. Este nivel puede verse influido por otras variables: la complejidad de la intervención, las estrategias de facilitación, la calidad de la aplicación y la capacidad de respuesta de los participantes (figura 1). La relación entre una intervención y su impacto (representada por una línea quebrada en la figura) es externa a la fidelidad de la ejecución, pero el grado de fidelidad alcanzado puede afectar a esta relación. En la evaluación de los resultados se debe tener en cuenta el contenido de la intervención determinando los requisitos mínimos (componentes esenciales) necesarios para mantener un alto nivel de fidelidad. Pero hasta que se conozcan estos componentes, sigue siendo necesaria la fidelidad a todos los elementos de la intervención (Carroll et al. 2007).

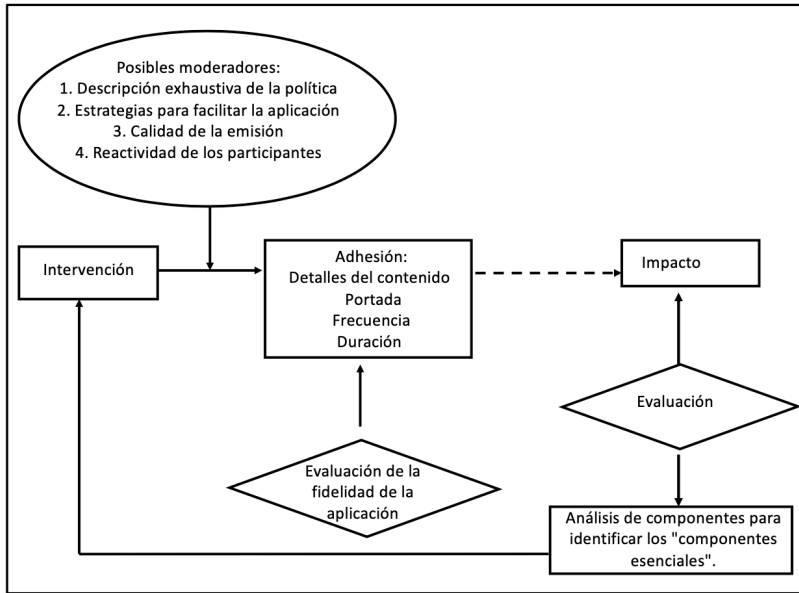


Figura 1. Modelo conceptual para evaluar la fidelidad de ejecución propuesto por Carroll et al. 2007

Apoyamos la idea de que la fidelidad y la adaptación coexisten (Rogers 2003). En nuestra opinión, las adaptaciones pueden conducir a mejorar o amenazar la teoría del cambio que subyace a la intervención (Hernández y Hodges 2003) y, por lo tanto, tener un impacto positivo o negativo en su eficacia. Por consiguiente, es necesario lograr un equilibrio adecuado entre la fidelidad y la adaptación durante la aplicación. Una forma de abordar esta cuestión es buscar sistemáticamente los aspectos del programa que han sido objeto de adaptaciones al evaluar su aplicación. La lealtad y la adaptación no son necesariamente mutuamente excluyentes. Los programas pueden incorporar tanto la lealtad como la adaptación. Por lo tanto, la evaluación debe considerar cuidadosamente ambos aspectos. Para ello, los investigadores necesitan: a) una descripción clara y completa de la intervención, incluido su contenido (qué hacer, durante cuánto tiempo y por quién), cómo poner en práctica la intervención y sus principios operativos, es decir, la teoría del cambio subyacente a la intervención, y b) medidas de

fidelidad y adaptación en relación con las características de la intervención (complejidad, capacidad de innovación), instrumentos de evaluación y fuentes de verificación.

En el resto de este capítulo, ofrecemos un ejemplo de la aplicación de una evaluación de la fidelidad de la intervención adaptativa en el contexto específico de Cuba y extraemos una serie de lecciones de ella.

Estableciendo el contexto

El dengue es una enfermedad de transmisión vectorial que se transmite principalmente por un mosquito hembra del género *Aedes aegypti* infectado con uno de los cuatro virus del dengue (Stoddard y otros, 2013). Un solo episodio de la enfermedad puede evolucionar hacia un dengue hemorrágico grave o mortal (Guzmán y otros, 2006). Todos los métodos de control actualmente disponibles apuntan al vector. Son casi imposibles de aplicar sin la aceptación de la comunidad o la participación activa (Parks y Lloyd 2004; Toledo y otros, 2008).

Se ha informado de que las estrategias de potenciación son eficaces para la prevención y el control del dengue de base comunitaria en Cuba (Toledo y otros, 2007; Sánchez y otros, 2009; Vanlerberghe y otros, 2009; Castro y otros, 2012) y en otros lugares (Kay y Nam, 2005; Kay y otros, 2010). La potenciación es un proceso por el cual se habilita a los individuos, grupos y comunidades para que tomen medidas sobre las decisiones que afectan a sus vidas (Rifkin 1996); en nuestro caso, estos son los determinantes de las epidemias de dengue. Sin embargo, estas estrategias siguen siendo controvertidas debido a cuestiones de transferibilidad y ampliación. Además, las descripciones de los procesos que intervienen en su aplicación son poco frecuentes en la literatura (Rifkin 2014).

Las pruebas científicas que vinculan directamente la participación de la comunidad con los efectos en la salud se consideran en general limitadas e insuficientes. Cuando se encuentran esos vínculos causales, se considera que son específicos del contexto y del proceso (contenido). Sin embargo, la aplicación y el papel del contexto se ignoran o se describen y analizan deficientemente en la literatura (Rifkin 2014). Así pues, se ha llevado a cabo un conjunto de estudios para proporcionar una descripción detallada y sólida de los procesos de aplicación de una estrategia de creación de

capacidad para la prevención y el control del dengue en Cuba. Nos enfrentamos a los retos de fidelidad y adaptación que se encuentran en su reproducción y aplicación a mayor escala.

Se ha diseñado una estrategia de creación de capacidad basada en pruebas para la prevención del dengue, junto con el Programa Cubano de Control del *Aedes aegypti* (AaCP), que se ha reproducido a mayor escala (Pérez et al. 2010; Castro et al. 2012). La AaCP, establecida en 1981, tiene una cobertura nacional. Está estructurado según los niveles de organización del sistema nacional de salud (nacional, provincial, municipal y primario). Está estructurada verticalmente y gestionada principalmente a nivel nacional, aunque en cierta medida es posible descentralizar la toma de decisiones en función de las características locales y la situación epidemiológica. Durante años, miles de trabajadores sobre el terreno han llevado a cabo actividades normalizadas como la vigilancia entomológica, la reducción de la fuente larvaria y el control de los mosquitos adultos, la educación sanitaria y la aplicación de la legislación de control de mosquitos (es decir, el uso de multas cuando se encuentran criaderos de mosquitos en los hogares). Sin embargo, en el enfoque de la AaCP, el papel de la población sigue siendo limitado y pasivo (Kourí y otros, 1989).

Descripción de la estrategia innovadora de creación de capacidad para la prevención del dengue

La estrategia fue desarrollada por investigadores del Instituto de Medicina Tropical Pedro Kourí en La Habana. Se basa en las teorías de empoderamiento y educación popular (Freire 1972). Su objetivo es la participación de la comunidad en la toma de decisiones para las actividades de control de vectores. Concebido como un componente horizontal que se integrará en el programa de control vertical existente, tiene cuatro componentes principales: organización y gestión, creación de capacidad, trabajo comunitario y vigilancia. Para cada componente se definieron las actividades (es decir, los subcomponentes) que debían realizarse (Cuadro 1).

La estrategia se reprodujo entre octubre de 2004 y diciembre de 2007 en 16 comunidades (circunscripciones) seleccionadas al azar en tres Consejos Populares del Municipio de La Lisa (Castro y otros 2012). La circunscripción es el nivel más descentralizado de gobierno local (aproximadamente 1.000

habitantes). El Consejo del Pueblo es una estructura de gobierno intermedia entre el nivel municipal y la circunscripción. Las circunscripciones electorales y los consejos populares eran heterogéneos en cuanto a la composición sociodemográfica de la población, las experiencias previas de participación comunitaria, las características de la dirección, los recursos y la dinámica del gobierno local.

Tabla 1 : Estrategia de creación de capacidad para el control del *Aedes aegypti* por componentes y subcomponentes. Municipio de La Lisa, La Habana, 2004-2007

Componentes	Definición operativa	Subcomponentes
1. Fomento de la capacidad	Desarrollo de los conocimientos, las aptitudes y los valores y prácticas asociados que necesitan los miembros de la comunidad para impulsar la estrategia.	1.1 Diagnóstico, trabajo en grupo y participación
		1.2 Vigilancia de los riesgos y del comportamiento
		1.3 Planes de acción y estrategia de comunicación
		1.4 Evaluación participativa
2. Organización y gestión	La forma en que los actores involucrados en el control del dengue se definen, organizan, comprometen y establecen roles, identifican recursos y toman decisiones	2.1 Presencia de grupos de trabajo comunitarios (GTC) que lideran la estrategia
		2.2 Presencia de personal del programa de control de vectores en los CTC
		2.3 Identificación de los recursos de la comunidad
		2.4 Movilización de recursos externos
3. Trabajo comunitario	Un ciclo repetitivo de acciones desarrolladas por un grupo de base comunitaria para cambiar las condiciones que aumentan la probabilidad de transmisión del dengue	3.1 Cartografía de riesgos
		3.2 Diagnóstico de los problemas
		3.3 Plan de acción
		3.4 Acciones realizadas
		3.5 Estrategia de comunicación
		3.6 Desarrollo de materiales de comunicación
		3.7 Vigilancia y evaluación
4. Vigilancia	Recopilación y análisis sistemáticos y oportunos de datos sobre los riesgos de transmisión del dengue y los comportamientos conexos para la adopción de medidas	4.1 Identificación de los riesgos ambientales
		4.2 Identificación de los riesgos de la vivienda
		4.3 Identificación de los comportamientos asociados

Protocolo de investigación y marco de análisis

Se llevó a cabo una evaluación retrospectiva de la aplicación como parte de la evaluación final de la reproducción de la estrategia. Sus objetivos eran: 1) evaluar el equilibrio entre la fidelidad y el ajuste en la reproducción de la estrategia; 2) comprender las dificultades y obstáculos encontrados durante la aplicación con miras a mejorar el proceso; y 3) avanzar en los conocimientos teóricos y prácticos sobre la realización de investigaciones sobre la fidelidad en las intervenciones de adaptación.

Para lograr nuestros objetivos, realizamos una evaluación de la fidelidad utilizando un método mixto de cuantificación de datos cualitativos (Creswell y Clarck 2011).

Por definición, una estrategia de empoderamiento debe modificarse y revisarse en función de su interacción con el marco institucional y los contextos en los que se aplica (Castro y otros, 2008). Dada la naturaleza adaptativa de la estrategia, anticipamos que las dimensiones convencionales de la fidelidad serían de utilidad limitada para evaluar las adaptaciones. De hecho, promover la participación de la comunidad en la toma de decisiones implica, por definición, incertidumbre sobre lo que se planificará y/o logrará.

Por consiguiente, optamos por evaluar la fidelidad y la adaptación en la aplicación utilizando la metodología de Rebchoock (et al. 2006). Según estos autores, se pueden identificar tres tipos de adaptación al diseño inicial de una intervención: (1) la omisión o modificación de un componente, de modo que el programa ya no se ejecute según lo previsto, (2) un componente ejecutado con modificaciones menores o mayores respetando el objetivo inicial, o (3) actividades adicionales o componentes añadidos. En teoría, las implicaciones de estos tres tipos de adaptaciones en la fidelidad son diferentes. Cuando se retira un componente, la fidelidad se ve amenazada. En el segundo caso, dependiendo de lo que impliquen los cambios, la fidelidad puede o no verse afectada. Por último, al añadir algo nuevo, la fidelidad se puede mantener fácilmente.

Aplicación del método y los procedimientos para evaluar la fidelidad y el ajuste

Se llevó a cabo una evaluación en tres etapas (Pérez et al. 2010). La primera etapa fue una evaluación individual por parte de los tres diseñadores de la estrategia que participaron como facilitadores de la aplicación de la estrategia. Se basó en un análisis de los procedimientos del proyecto, las actas de los talleres celebrados y un examen de los documentos rellenos por los agentes de ejecución. La segunda etapa consistió en un debate de las conclusiones con un grupo más amplio de expertos (los tres que participaron en la primera etapa y tres nuevos). Si todos estaban de acuerdo en que un componente o subcomponente se había aplicado según lo previsto en un determinado grupo, se clasificaba como aplicado. Si había consenso en cuanto a que un componente o subcomponente no se aplicaba, se clasificaba como no aplicado. Por último, si uno de los expertos consideraba que un componente o subcomponente había sido modificado, se clasificaba como modificado. También se identificaron actividades adicionales. El tercer paso fue una evaluación que se perfeccionó progresivamente de manera consensuada tras las conversaciones con los responsables de la aplicación a través de un taller de evaluación participativa.

Los datos cualitativos obtenidos de la evaluación en tres etapas (es decir, los descriptores de la aplicación de los componentes/subcomponentes de la estrategia para cada distrito) se transformaron en datos cuantitativos (Creswell y Clarck 2011). Se tabularon las frecuencias de los subcomponentes no implementados, modificados e implementados para todos los distritos. El promedio se calculó para los cuatro componentes. Para explorar las relaciones entre los principales componentes, se asignó una puntuación de 0, 1 y 2 a los componentes omitidos, modificados y aplicados de la estrategia, respectivamente. Sus valores fueron sumados para cada cabalgata. Posteriormente, se calcularon los coeficientes de correlación de Spearman Rank entre los componentes utilizando el software SPSS V.15.

Además, se realizaron entrevistas semiestructuradas con trece informantes clave para determinar posibles explicaciones de las variaciones observadas en la aplicación. Se utilizó un muestreo intencional para la selección de los entrevistados a fin de identificar a los que desempeñaban un papel importante en la aplicación, en diferentes niveles y funciones. En la guía de entrevistas se analizó la comprensión de los entrevistados

de los componentes/subcomponentes de la estrategia, las dificultades que surgieron durante la aplicación y sus causas percibidas. Las entrevistas fueron grabadas, transcritas y analizadas con el apoyo del software QSR Nvivo 7. El análisis fue principalmente deductivo, pero permitió la aparición de temas adicionales (por ejemplo, dificultades).

Principales resultados

La vigilancia fue el componente más aplicado, seguido de la creación de capacidad. El trabajo comunitario y la organización y gestión fueron los componentes menos implementados o modificados (Cuadro 2) (Pérez et al. 2010). Incluso para los componentes más implementados, los subcomponentes no se implementaron. Cuanto más innovadores eran los componentes y subcomponentes, menos se aplicaban, por ejemplo, los que entrañaban actividades diferentes o perturbadoras de las prácticas actuales de lucha contra los vectores y que no podían lograrse fácilmente utilizando los conocimientos especializados, las experiencias anteriores o los conocimientos técnicos de los agentes de ejecución.

COMPONENTES DE LA ESTRATEGIA	El Consejo del Pueblo 1					El Consejo del Pueblo 2						El Consejo del Pueblo 3					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
MONITOREO																	
Identificación de los riesgos ambientales	I	I	I	I	I	I	I	I	I	N	N	I	I	I	N	N	
Identificación de los riesgos de la vivienda	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
Identificación de los comportamientos asociados	I	I	N	I	N	I	N	N	N	N	N	I	I	N	N	N	
DESARROLLO DE LA CAPACIDAD																	
Diagnóstico, trabajo en grupo y participación	I	I	I	I	I	I	I	I	I	I	I	M	N	N	N	N	
Vigilancia del riesgo y del comportamiento	I	I	I	I	I	I	I	I	I	M	N	M	N	M	M	M	
Estrategia de planificación y comunicación	I	I	I	I	M	I	I	I	I	M	N	M	N	M	N	N	
Evaluación participativa	I	I	I	I	N	I	I	I	I	N	N	N	N	N	N	N	
Talleres "auto-organizadores"	A					A											
TRABAJO COMUNITARIO																	
Cartografía de riesgos	I	I	I	I	M	I	I	I	I	I	N	I	I	I	N	N	
El diagnóstico de los problemas	I	I	I	M	M	I	I	I	I	I	N	I	I	I	N	N	
Plan de acción	I	I	N	N	N	I	I	I	N	N	N	I	M	I	N	N	
Acciones realizadas	I	I	N	N	N	I	I	I	N	N	N	I	I	N	N	N	
Estrategia de comunicación	I	I	M	M	M	I	N	N	N	N	N	N	N	N	N	N	
Desarrollo de materiales de comunicación	I	I	N	N	N	I	N	N	N	N	N	N	N	N	N	N	
Vigilancia y evaluación	I	I	I	I	I	M	M	M	M	M	N	N	N	N	N	N	
Evaluación de los puntos fuertes de la comunidad	A	A	A	A													
ORGANIZACIÓN Y GESTION																	
La presencia de los CTCs utilizando la estrategia	I	I	M	M	N	I	I	I	M	N	N	I	M	I	N	N	
El personal de AuCP en los RCT	I	I	M	M	N	I	I	I	M	N	N	M	M	M	N	N	
Identificación de los recursos de la comunidad	I	I	N	N	N	M	I	N	N	N	N	N	N	N	N	N	
Mobilización de recursos externos	I	I	N	N	N	M	N	N	N	N	N	N	N	N	N	N	
Se ha establecido un vínculo con la escuela primaria	A			A		A											
Participación de otros actores	A			A		A								A			

Tabla 2 : Componentes y subcomponentes de la estrategia implementada, modificada, no implementada o añadida por el Consejo del Pueblo 1 y las circunscripciones. Municipio de La Lisa, La Habana, 2004-2007

Aparte de la vigilancia y la creación de capacidad, todos los componentes estaban significativa y altamente correlacionados (coeficiente de correlación de Spearman Ranketgt; 0,70, P etlt;0,01). Si un componente se había implementado en una cabalgata, los otros componentes tenían la misma probabilidad de implementarse. La correlación entre la aplicación

de los componentes de Trabajo comunitario y Organización y gestión fue de 0,95, mientras que la correlación entre los componentes de Vigilancia y Creación de capacidad fue menor (coeficiente de correlación de 0,50).

Los subcomponentes que se aplicaron de manera deficiente fueron la movilización de recursos internos y externos por parte de los grupos de interés, el diseño de estrategias de comunicación y la elaboración de materiales de comunicación locales. Los principales cambios realizados han sido en la composición de los grupos de trabajo comunitarios, los cambios en el enfoque de la creación de capacidad, la participación individualizada y los cambios en el diseño de las actividades de capacitación con la utilización de técnicas de participación mejor adaptadas a las características de los participantes. Teniendo en cuenta la estrategia en su conjunto, se añadieron ciertas actividades, como la incorporación de actividades de la escuela primaria en la estrategia, la participación de actores no previstos inicialmente, talleres adicionales de capacitación de la comunidad y la evaluación de los puntos fuertes de la comunidad como parte del diagnóstico participativo inicial.

Las dificultades encontradas durante la aplicación se relacionaron con la capacitación y la movilización de las aptitudes adecuadas, el tiempo disponible, la falta de apoyo y compromiso del gobierno local y las autoridades sanitarias, la falta de motivación de los dirigentes locales, la falta de integración de los agentes y la falta de movilización de recursos. El estudio demostró una amplia variabilidad en la fidelidad de la aplicación. Esta variabilidad se explica en gran medida por la complejidad de la estrategia, que a veces perturbó las actuales prácticas de control de vectores y la falta de conocimientos de los encargados de aplicarla respecto de los principios básicos de la estrategia. Sin embargo, la variación en la fidelidad de la aplicación no se tradujo en una pérdida sustancial de eficacia (Castro y otros, 2012).

Análisis reflexivo

La tipología de Rebchook (et al. 2006) ha demostrado ser apropiada para evaluar la fidelidad y la adaptación, pero ha mostrado limitaciones. No nos permitió saber qué cambios introducidos en la estrategia contribuían o no al logro de resultados. También es cuestionable la afirmación de Rebchook

(et al. 2006) de que la adición de nuevos componentes o subcomponentes a un modelo inicial no amenaza a priori la fidelidad. De hecho, esta adición podría estar en contradicción con la teoría del cambio que subyace a la intervención. Además, la tipología no nos proporcionó una visión muy detallada de la fidelidad de la intervención en cuanto a contenido, dosis y cobertura. Por lo tanto, comparamos nuestra investigación empírica con la literatura existente sobre la fidelidad. Identificamos el marco para analizar la fidelidad en la aplicación de Carroll y colegas (2007) como uno que podría potencialmente abordar nuestra preocupación por evaluar la adaptación en el contexto de la fidelidad (Pérez et al. 2016). Sin embargo, necesitábamos adaptarlo para acomodar las intervenciones de salud global adaptables. Por lo tanto, proponemos algunas modificaciones a este marco con implicaciones prácticas para la evaluación de la fidelidad y la adaptación (Pérez et al. 2016).

Nuestro marco modificado para analizar la fidelidad de la aplicación conserva las ideas de Carroll (et al. 2007) con respecto a la evaluación de los efectos logrados y el análisis de los componentes para identificar los esenciales para una intervención. De hecho, la evaluación de la aplicación y el logro de un equilibrio adecuado entre la fidelidad y la adaptación sólo tiene sentido en el contexto de la evaluación de sus efectos (Backer 2002). Sin embargo, hemos mejorado la representación gráfica del papel de la fidelidad y la evaluación del impacto en la identificación de los componentes esenciales de una intervención (Figura 2).

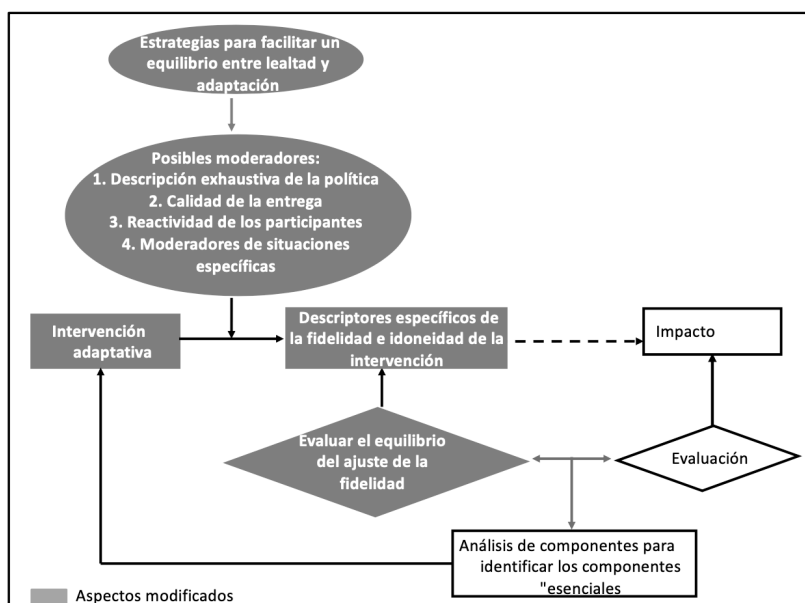


Figura 2. Marco conceptual modificado para la evaluación de la aplicación, de Carroll y otros, 2007

Carroll y sus colegas (2007) establecieron la adhesión como la única medida básica de fidelidad en la aplicación. Como hemos aprendido de nuestra investigación empírica sobre la fidelidad (Pérez et al. 2010), la naturaleza de las adaptaciones debe tenerse en cuenta conscientemente para evaluar su efecto en la eficacia. Además, es posible que las subcategorías de adhesión no siempre sean suficientes para evaluar cada intervención. Tal vez sea necesario considerar otros aspectos. Por lo tanto, sugerimos que se evalúen descriptorios específicos de la intervención tanto para la fidelidad como para la adaptación, en lugar de considerar sólo las subcategorías fijas de la adhesión.

Para avanzar en esta dirección, los diseñadores de la intervención necesitan 1) tener una idea clara de los resultados previstos, 2) hacer explícitos los principios operativos o la teoría del cambio de la intervención, 3) poder presentarlos como descriptorios específicos de la fidelidad, y 4) desarrollar preguntas para identificar las adaptaciones relacionadas con la descripción de la intervención. Al responder a estas preguntas se desarrollan

descriptores específicos de las adaptaciones de la intervención. Un paso adicional consiste en determinar en qué medida las adaptaciones identificadas afectan a los principios operativos de un componente y/o a la intervención en su conjunto.

Un ejemplo de este razonamiento puede proponerse utilizando el componente de creación de capacidad de nuestra estrategia. El resultado esperado de este componente era que los participantes adquirieran los conocimientos y aptitudes necesarios para modificar las relaciones de poder existentes entre ellos. Los principios de funcionamiento se establecieron utilizando el modelo pedagógico de la educación popular (Freire 1972). Los participantes en la capacitación deben adquirir conocimientos y aptitudes específicos mediante un proceso de aprendizaje participativo (grupo de aprendizaje). El grupo de aprendizaje debe incluir a las partes interesadas que necesitan cambiar sus relaciones de poder. Los descriptores específicos de la fidelidad para la creación de capacidad y las preguntas para identificar posibles adaptaciones se encuentran en el Cuadro 3.

Tabla 3 : Ejemplo de descriptores de fidelidad específicos y preguntas para identificar adaptaciones para el componente de creación de capacidad de la estrategia de creación de capacidad para el control del vector del dengue. Municipio de La Lisa, La Habana, 2004-2007

Descriptores específicos de la fidelidad	Preguntas para identificar las adaptaciones
<p>Qué: desarrollo de conocimientos y habilidades en cuatro áreas: 1) diagnóstico, trabajo en grupo y participación; 2) vigilancia del riesgo y del comportamiento; 3) planes de acción y estrategia de comunicación; y 4) evaluación participativa</p>	<p>¿Se ha cambiado el contenido de la capacitación? ¿Cómo se ha cambiado? ¿Se ha eliminado un tema? ¿Qué tema? ¿Por qué se quitó? ¿Se reemplazó un tema? ¿Por qué? ¿Por qué fue reemplazado? ¿Se ha añadido un tema? ¿Qué tema se añadió? ¿Por qué se añadió?</p>
<p>Cómo: a través de un taller basado en los principios del modelo pedagógico de educación popular: el objetivo es que los participantes reflexionen y, en consecuencia, transformen su realidad utilizando una lógica dialéctica entre la teoría y la práctica y métodos de aprendizaje participativos y basados en la experiencia.</p>	<p>¿Se ha adaptado algún principio del modelo pedagógico: objetivos, lógica, métodos de aprendizaje? ¿Cuál de ellos? ¿Cómo se hizo esto? ¿Cómo? ¿Por qué? ¿Ha sido reemplazado el modelo pedagógico por otro? ¿Cuál? ¿Por qué?</p>
<p>Con qué frecuencia: un taller de cuatro horas para cada sujeto durante un periodo de tres meses</p>	<p>¿Se ha hecho una adaptación en la frecuencia del entrenamiento: número de sesiones, número de horas por sesión? ¿Cómo se hizo esto? ¿Cómo? ¿Por qué? ¿Se han reducido las sesiones del taller con el tiempo? ¿Cómo sucedió esto? ¿Por qué ha ocurrido esto? ¿Se ha introducido algún ajuste en la duración y el calendario del entrenamiento? ¿Cómo sucedió esto? ¿Por qué fue este el caso?</p>
<p>A quién: Un grupo de aprendizaje compuesto por tres a cinco actores de al menos tres comunidades con diferentes relaciones de poder entre ellas, relacionadas con las actividades contra el vector del dengue.</p>	<p>¿Se ha adaptado el grupo de aprendizaje: el número de participantes, el papel de los interesados en relación con las actividades de control de los vectores del dengue? ¿Cómo se ha adaptado? ¿Por qué fue adaptado? ¿El grupo de aprendizaje ha sido reemplazado por otra estrategia de enseñanza? ¿Por cuál? ¿Por qué se hizo esto?</p>
<p>¿Por quién? Facilitadores previamente entrenados basados en los principios del modelo pedagógico de educación popular.</p>	<p>¿No estaba entrenado un facilitador? ¿Por qué no? ¿Se adaptó un principio del modelo pedagógico durante la formación de los facilitadores? ¿Cuál de ellos? ¿Cómo se hizo esto? ¿Por qué se hizo esto? ¿Ha sido reemplazado el modelo pedagógico por otro? ¿Por cuál? ¿Por qué fue reemplazado?</p>
<p>Especificaciones del contexto: Tres Consejos Populares (PC) participan en el proyecto. Hay de cinco a seis grupos seleccionados al azar por cada PC. Se proporciona apoyo material para la capacitación: por ejemplo, directrices escritas sobre la forma de realizar un taller de educación popular, asesoramiento metodológico a los facilitadores por parte de al menos un diseñador de estrategias.</p>	<p>¿Ha habido un cambio en el número de PCs? ¿Por qué ha habido un cambio en el número de PCs? ¿Ha habido un cambio en el número de grupos de interés involucrados? ¿Ha sido reemplazado un distrito electoral? ¿Cómo fue reemplazado uno? ¿Por qué se reemplazó un caballo? ¿Se han producido cambios en el apoyo metodológico: el suministro de directrices, el contenido de las directrices, el asesoramiento metodológico? ¿Cómo se hicieron los cambios? ¿Por qué se hicieron los cambios?</p>

Obsérvese que los descriptores específicos de la fidelidad ofrecen una descripción completa de la intervención tal como se pretende, con detalles sobre el contenido, los procesos (“qué”, “cómo”, “con qué frecuencia”, “a quién” y “por quién”) y las especificaciones relacionadas con el contexto de la aplicación. Supongamos que, al aplicar las preguntas a las pruebas empíricas, las adaptaciones identificadas en relación con el fomento de la capacidad se hubieran modificado de tres maneras diferentes: 1) se eliminó el tema “diagnóstico, trabajo en grupo y participación”, 2) se añadió un taller sobre estrategias de comunicación, y 3) se adaptaron los métodos de aprendizaje a las características de los participantes. Teniendo en cuenta nuestros principios operativos, debemos preguntarnos cuál de las tres adaptaciones podría influir en los efectos esperados. Desde nuestro punto de vista, no es posible proporcionar a los participantes la capacidad de participar sin enseñarles el tema eliminado. Por el contrario, no consideramos que ninguna de las otras dos adaptaciones amenace los principios operativos de la creación de capacidad y sus resultados.

Mediante este análisis de las adaptaciones que se producen en la realización de una intervención, se pueden identificar posibles soluciones para mejorar la aplicación en el futuro. Esto ayuda a los diseñadores de la intervención a identificar las adaptaciones no predefinidas que podrían mejorar el diseño de la intervención y, por lo tanto, su eficacia. Una vez que se identifique una adaptación positiva, la intervención podría pasar por un nuevo ciclo de diseño, aplicación y análisis. Esto requeriría mecanismos de retroalimentación.

Algunos autores han destacado la necesidad de seguir identificando las posibles fuentes de variabilidad en la aplicación (Arai y otros, 2005; Roen y otros, 2006; Belaid y Ridde, 2015). Carroll y sus colegas (2007) reconocen que el nivel de fidelidad alcanzado está influido por posibles factores moderadores que no son necesariamente independientes. Hasson y sus colegas (2012) también destacaron la importancia de otros mecanismos y factores que influyen en la fidelidad de la aplicación cuando experimentaron con el marco de Carroll y sus colegas (2007). Nuestro marco modificado mantiene la amplitud de la descripción de la política, la calidad de la aplicación y la capacidad de respuesta de los participantes como posibles moderadores principales. Sin embargo, permite la inclusión de otros específicos del contexto o situación.

Además, estamos de acuerdo con Carroll y sus colegas (2007) en que las estrategias de facilitación podrían influir en los posibles moderadores en cuanto al nivel de fidelidad alcanzado. Por ejemplo, el suministro de manuales y la capacitación de los encargados de la ejecución podrían mejorar la calidad de la ejecución de las intervenciones. Sin embargo, en nuestro marco modificado, estas estrategias se aplican no para lograr una adhesión estricta, sino para contribuir a un equilibrio adecuado entre la adaptación y la fidelidad. Una vez que se hayan identificado las adaptaciones, las estrategias de facilitación sólo se ocuparán de las que se consideren inadecuadas. En el caso de las intervenciones adaptables, esto también proporciona una visión global de las adaptaciones que se han producido.

En la práctica, el objetivo de lograr un equilibrio de fidelidad-ajuste implica su interdependencia. La necesidad de ese equilibrio, puesta de relieve por Backer (2002) y von Thiele Schwarz (et al. 2014), está firmemente respaldada por los resultados de nuestra investigación empírica sobre la fidelidad (Pérez et al. 2010) y nuestras investigaciones posteriores sobre la difusión de la estrategia de potenciación para la prevención del dengue (Pérez et al. 2013).

Referencias clave

Rebchook, M., Kegeles, S., Huebner, D. et TRIP Research Team. (2006). Translating research into practice: the dissemination and initial implementation of an evidence-based HIV Prevention Program. *AIDS Education and Prevention*, 18(Suppl. A), 119-136.

Este artículo presenta los resultados de la investigación empírica sobre la ampliación de un programa de prevención del VIH. Uno de sus objetivos es explorar la forma en que los interesados tratan de aplicar el programa. En el documento se analizan los tipos de reinversiones/adaptaciones realizadas y sus implicaciones para la fidelidad.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40-48.

El presente documento ofrece el marco conceptual más completo para evaluar la fidelidad de la aplicación. Este marco ha demostrado posteriormente ser un instrumento de evaluación útil para evaluar la fidelidad de las intervenciones complejas.

Rogers, E. (2003). *Diffusion of innovations* (5^e éd.). New York: New York Free Press.

La quinta edición del libro en la que el autor explica cómo las nuevas ideas y tecnologías se difunden entre los miembros de un sistema social y se ponen en práctica. El libro se basa en la observación de las regularidades en la difusión de las diferentes innovaciones entre las diferentes culturas y usuarios y contiene desarrollos teóricos desarrollados a lo largo de décadas.

Referencias

- Arai, L., Roen, K., Roberts, H. et Popay, J. (2005). It might work in Oklahoma but will it work in Sout? Oakhampton: Context and implementation in the effectiveness literature on domestic smoke detectors. *Injury Prevention*, 11(3),148-151.
- Backer, T. (2002). *Finding the balance: Program fidelity and adaptation in substance abuse prevention: A state-of-the-art review*. Rockville, M. (ed.). Center for Substance Abuse Prevention.
- Belaid, L. et Ridde, V. (2015). Contextual factors as a key to understanding the heterogeneity of effects of a maternal health policy in Burkina Faso?. *Health Policy and Planning*, 30(3), 309-321.
- Bellg, A., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D. et Czajkowski, S. (2004). Enhancing treatment fidelity in health behaviour change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, 23(5), 443-451.
- Borrelli, B. (2011). The assessment, monitoring, and enhancement of treatment fidelity in public health clinical trials. *Journal of Public Health Dentistry*, 71, 52-63.

- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. et Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2, 40-48.
<http://www.implementationscience.com/content/2/1/40>
- Castro, M., Pérez, D., Pérez, K., Polo, V., Lopez, M. et Sanchez, L. (2008). Contextualización de una estrategia comunitaria integrada para la prevención del dengue. *Revista Cubana de Medicina Tropical*, 60, 83-91.
- Castro, M., Sanchez, L., Pérez, D., Carbonell, N., Lefèvre, P., Vanlerberghe, V. et Van der Stuyft, P. (2012). A community empowerment strategy embedded in a routine dengue vector control programme: a cluster randomised controlled trial. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 106(5), 315-321.
- Collins, L., Murphy, S. et Bierman, K. (2004). A conceptual framework for adaptive preventive interventions. *Prevention Science*, 5(3), 185-196.
- Creswell, J. et Clark, V. P. (2011). *Design and conducting mixed methods research*. Washington, DC: Sage Publications.
- Dane, A. et Schneider, B. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control?. *Clinical Psychology Review*, 18(1), 23-45.
- Dusenbury, L., Brannigan, R., Falco, M. et Hansen, W. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256.
- Fixsen, D., Naoom, S., Blase, K., Friedman, R. et Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida.
- Freire, P. (1972). *Pedagogy of the Oppressed*. Londres: Sheed and Ward.
- Gagliardi, A., Straus, S., Shojania, K. et Urbach, D. (2014). Multiple interacting factors influence adherence and outcomes associated with surgical safety checklists: a qualitative study. *PLoS One*, 9(9): e108585.
<https://doi.org/10.1371/journal.pone.0108585>
- Guzman, M., Rodriguez-Roche, R. et Kouri, G. (2006). Dengue and dengue hemorrhagic fever. *Giornale Italiano Di Medicina Tropicale*, 11, 1-2.
- Hasson, H., Blomberg, S. et Dunér, A. (2012). Fidelity and moderating factors in complex interventions: a case study of a continuum of care program

- for frail elderly people in health and social care. *Implementation Science*, 7, 23-33.
<https://doi.org/10.1186/1748-5908-7-23>
- Hernandez, M. et Hodges, S. (2003). Building upon the theory of change for systems of care. *Journal of Emotional and Behavioral Disorders*, 11(1), 19-26.
- Kay, B. et Nam, V. (2005). New strategy against *Aedes aegypti* in Vietnam. *The Lancet*, 365(9459), 613-617.
- Kay, H., Thanh, T. T., Le, N., Quy, T., Nam, V., Hang, P., ... et Ryan, P. (2010). Sustainability and cost of a community-based strategy against *Aedes aegypti* in northern and central Vietnam. *American Journal of Tropical Medicine and Hygiene*, 82(5), 822-830.
- Kouri, G., Guzman, M., Bravo, J. et Triana, C. (1989). Dengue haemorrhagic fever/dengue shock syndrome: lessons from the Cuban epidemic, 1981. *Bulletin of World Health Organization*, 67(4), 375-380.
- Parks, W. et Lloyd, L. (2004). *Planning social mobilization and communication for dengue fever prevention and control: A step-by-step guide*. Genève: World Health Organization.
- Pérez, D., Lefèvre, P., Castro, M., Sanchez, L., Toledo, M., Vanlerberghe, V. et Van der Stuyft, P. (2010). Process-oriented fidelity research assists in evaluation, adjustment and scaling up of community-based interventions. *Health Policy and Planning*, 26(5), 413-422.
- Pérez, D., Lefèvre, P., Castro, M., Toledo, M., Zamora, G., Bonet, M. et Van der Stuyft, P. (2013). Diffusion of community empowerment strategies for *Aedes aegypti* control in Cuba: A muddling through experience. *Social Science and Medicine*, 84, 44-52.
- Pérez, D., Van der Stuyft, P., Zabala, M. et Lefèvre, P. (2016). A modified theoretical framework to assess implementation fidelity of adaptive public health interventions. *Implementation Science*, 11, 106.
<https://doi.org/10.1186/s13012-016-0457-8>
- Peters, D., Tran, N. et Adam, T. (2013). *Implementation Research in Health: A Practical Guide*. Genève: World Health Organization.
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R. et Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement, challenges and research agenda.

Administration and Policy in Mental Health and Mental Health Services Research, 38(2), 65-76.

<https://doi.org/10.1007/s10488-010-0319-7>

- Rebchook, M., Kegeles, S., Huebner, D. et TRIP Research Team. (2006). Translating research into practice: the dissemination and initial implementation of an evidence-based HIV Prevention Program. *AIDS Education and Prevention*, 18(Suppl. A), 119-136.
- Rifkin, S. (1996). Paradigms lost: Toward a new understanding of community participation in health programmes. *Acta Tropica*, 61(2), 79-92.
- Rifkin, S. (2014). Examining the links between community participation and health outcomes: a review of the literature. *Health Policy and Planning*, 29(Suppl. 2), 98-106.
- Roen, K., Arai, L., Roberts, H. et Popay, J. (2006). Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. *Social Science and Medicine*, 63(4), 1060-1071.
- Rogers, E. (2003). *Diffusion of innovations* (5^e éd.). New York: New York Free Press.
- Sanchez, L., Pérez, D., Cruz, G., Castro, M., Kouri, G., Shkedy, Z., Vanlerberghe, V. et Van der Stuyft, P. (2009). Intersectoral coordination, community empowerment and dengue prevention: six years of controlled interventions in Playa Municipality, Havana, Cuba. *Tropical Medicine and International Health*, 14(11), 1356-1364.
- Schwarz, U. V., Hasson, H. et Lindfors, P. (2014). Applying a fidelity framework to understand adaptations in an occupational health intervention. *Work*, 51(2), 195-203.
- Stoddard, S., Forshey, B., Morrison, A., Paz-Soldan, V., Vazquez-Prokopec, G., Astete, H., Reiner, R.C., Vilcarromero, S., Elder, J. P., Halsey, E. S., Kochel, T. J., Kitron, U. et Scott, T. (2013). House-to-house human movement drives dengue virus transmission. *Proceedings from the National Academy of Sciences of the United States of America*, 110(3), 994-999.
- Sundell, K., Beelmann, A. H et Schwarz, U. V. (2015). Novel programs, international adoptions, or contextual adaptations?: Meta-analytical

results from German and Swedish intervention research. *Journal of Clinical Child and Adolescent Psychology*, 45:6, 784-796.

<http://doi:10.1080/15374416.2015.1020540>

Toledo, M., Baly, A., Vanlerberghe, V., Rodriguez, M., Benitez, J., Duvergel, J. et Van der Stuyft, P. (2008). The unbearable lightness of technocratic efforts at dengue control. *Tropical Medicine and International Health*, 13(5), 728-736.

Toledo, M., Vanlerberghe, V., Baly, A., Ceballos, E., Valdes, L., Searret, M., Boelaert, M. et Van der Stuyft, P. (2007). Towards active community participation in dengue vector control: results from action research in Santiago de Cuba, Cuba. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 101, 56-63.

Vanlerberghe, V., Toledo, M., Rodriguez, M., Gomez, D., Baly, A., Benitez, J. et Van der Stuyft, P. (2009). Community involvement in dengue vector control: cluster randomised trial. *British Medical Journal*, 338 :b1959. <https://www.bmj.com/content/338/bmj.b1959>

Resumen / Résumé / Abstract

Este capítulo se centra en la evaluación de la fidelidad en la implementación y la adaptación de intervenciones en salud global. La fidelidad, o el grado en que una intervención se implementa según lo previsto por sus diseñadores, está especialmente destinada a garantizar que la intervención mantenga los efectos previstos. La adaptación, por el contrario, es la modificación del modelo original de una intervención por parte de los actores involucrados en su implementación. Existen cinco dimensiones para medir la fidelidad: adherencia, dosis, calidad del servicio, respuesta de los participantes y diferenciación del programa. Sin embargo, existe poca investigación o consejos prácticos sobre cómo adaptar una intervención manteniendo los mecanismos que garantizan su eficacia. La implementación de intervenciones puede ser una mezcla de fidelidad y adaptación, por lo que en la evaluación deben considerarse ambos aspectos. Comparamos nuestra investigación empírica con la literatura existente sobre fidelidad. Como resultado consideramos que el marco conceptual de Carroll y colaboradores permitiría evaluar adaptación en el contexto de fidelidad.

Por ello proponemos modificaciones a este marco teórico para adecuarlo a las intervenciones adaptativas en salud global con implicaciones prácticas para la evaluación de fidelidad y adaptación.

Ce chapitre porte sur l'évaluation de la fidélité dans la mise en œuvre et l'adaptation des interventions de santé mondiales. La fidélité, ou le degré avec lequel une intervention est mise en œuvre comme prévu par ses concepteurs. Elle est particulièrement destinée à assurer que l'intervention maintienne ses effets prévus. L'adaptation, au contraire, est la modification du modèle d'origine d'une intervention par les acteurs et actrices de sa mise en œuvre. Cinq dimensions ont été mises en avant pour mesurer la fidélité : l'adhérence, la dose, la qualité de la prestation, la réactivité des participant-e-s et la différenciation du programme. Cependant, peu de recherches ou de conseils pratiques portent sur la façon d'adapter une intervention, tout en maintenant les mécanismes essentiels à son efficacité. Les interventions peuvent être un mélange de fidélité et d'adaptation et l'évaluation doit examiner les deux aspects. Nous avons confronté notre recherche empirique avec la littérature existante sur la fidélité et en avons déduit que le cadre conceptuel proposé par Carroll permettait d'évaluer l'adaptation dans le contexte de la fidélité. Nous proposons des modifications à ce cadre pour accommoder les interventions en santé mondiales adaptables ainsi que des implications pratiques.

This chapter is about the evaluation of the implementation fidelity and adaptation while building evidence for global health interventions. This issue is particularly relevant for adaptive interventions. Fidelity or the degree to which an intervention is implemented as intended by its developers is an implementation outcome that is particularly meant to ensure that the intervention maintains its intended effects. Adaptation, on the opposite is the process of bringing changes to the original design of an intervention by its implementers or users. Five dimensions have been forwarded to measure fidelity: adherence, dose, quality of delivery, participant responsiveness, and program differentiation. However, there has been little research or practical advice on how to adapt an intervention to maintain its effective ingredients and mechanisms. Interventions can be a blend of both fidelity and adaptation; thus, evaluation must carefully examine both aspects. We

confronted our empirical research with existing literature on fidelity. As a result, we considered that the framework for implementation fidelity proposed by Carroll and colleagues responds to our concern to assess adaptation in the context of fidelity. We proposed modifications to the framework to accommodate adaptable global health interventions with practical implications for the evaluation of fidelity and adaptation.

Dennis Pérez es una socióloga cubana especializada en educación popular e investigación de acción participativa. Tiene una Maestría en Control de Enfermedades y Desarrollo Social, un Doctorado en Sociología (Universidad de La Habana) y en Ciencias de la Salud (Universidad de Gante). Desde 1997, trabaja como investigadora en la Unidad de Epidemiología del Instituto de Medicina Tropical Pedro Kouri de La Habana. Actualmente tiene una beca postdoctoral en el Institut de Recherche en Santé publique de l'Université de Montréal. Sus principales investigaciones se centran en la evaluación de los procesos de aplicación; la fidelidad en la reproducción y difusión de las estrategias de control del dengue basadas en la creación de capacidad; y los avances teóricos y metodológicos conexos. También se interesa por estudios cualitativos sobre las percepciones y la aceptabilidad de las intervenciones y servicios de salud por parte de las poblaciones.

Marta Castro es una médica especializada en psiquiatría pediátrica. También tiene una maestría y un doctorado en epidemiología. Actualmente trabaja como investigadora en la Unidad de Epidemiología del Instituto de Medicina Tropical Pedro Kouri de La Habana, principalmente en proyectos de epidemiología social relacionados con el dengue, el VIH y la tuberculosis. Sus proyectos recientes se centran en el diseño, la aplicación y la evaluación de programas de prevención y control del dengue de base comunitaria, prestando especial atención a la comprensión de los determinantes sociales de la salud y a la promoción de la participación. También ha realizado investigaciones cualitativas sobre el VIH en niños y adolescentes en Cuba.

Pierre Lefèvre es Doctor en Sociología, graduado en la Universidad Libre de Bruselas. Entre 1989 y 2017, trabajó como investigador en el Departamento de Salud Pública del Instituto de Medicina Tropical de Amberes (Bélgica). Primero en la Unidad de Nutrición y luego en la Unidad de Epidemiología y Control de Enfermedades. Actualmente es consultor independiente en salud pública internacional. Sus investigaciones se han

centrado en la evaluación pluralista (tesis doctoral), la planificación participativa, la participación de la comunidad, la transición nutricional, el crecimiento y el desarrollo del niño y las percepciones de la población sobre diversas enfermedades e intervenciones en muchos países. También ha apoyado metodológicamente (diseño, análisis) los componentes cualitativos de numerosos proyectos de investigación epidemiológica. Sus actuales esferas de interés son la institucionalización de la participación de la comunidad en los programas de control de vectores, los estudios de aceptabilidad y, más en general, la ciencia de la aplicación.

Cita

Dennis Pérez, Marta Castro y Pierre Lefèvre (2020). Evaluación de la fidelidad y la adaptación. Un ejemplo de la puesta en marcha de las intervenciones en salud global. Une intervention de financement basado en resultados en Mali. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 453-479. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

19. Evaluación realista

El ejemplo de la adopción de una política de salud pública en Benín

JEAN-PAUL DOSSOU Y BRUNO MARCHAL

Evaluación realista

Definición del método

Este enfoque de evaluación basado en la teoría tiene por objeto explicitar sistemáticamente los mecanismos causales de los resultados de la intervención para diferentes poblaciones y en diferentes contextos. No sólo responde a la clásica pregunta “¿Funciona o no funciona una intervención para producir un efecto?” sino, lo que es más importante, responde a importantes preguntas para la toma de decisiones como “¿Por qué funciona o no una intervención? ¿Cómo funciona? ¿Para quién y en qué contexto?”

Fortalezas del método

- Se trata de un enfoque emergente; se demanda cada vez más.
- Puede responder a preguntas de evaluación que no pueden abordarse con otros enfoques, a saber: ¿Cómo, para quién y en qué contexto funciona o no funciona una intervención, y por qué?
- Pone a los actores operativos en el centro de la evaluación. Este enfoque de empoderamiento y potenciación hace del proceso de evaluación una herramienta para una mejor redistribución del poder.
- Tiene el potencial de estimular y fortalecer la cultura local de aprendizaje promoviendo una explicación y confrontación sistemáticas de los supuestos normalmente implícitos que subyacen a las intervenciones y prácticas.

Desafíos del método

- Depende de los conocimientos del evaluador/a sobre el programa de evaluación, los agentes y los contextos de aplicación; requiere recursos relativamente importantes.
 - Se requiere un importante esfuerzo de síntesis y creatividad para comunicar eficazmente los resultados.
 - El criterio de calidad de los resultados de la evaluación realista es la verosimilitud. Los financiadores de la evaluación que se interesan por la atribución de resultados a un programa determinado podrían considerar esto como una limitación.
 - Sigue siendo un enfoque en construcción con un debate metodológico que todavía está vivo. A veces puede ser difícil encontrar puntos de referencia metodológicos claros, precisos y establecidos para el estudiante realista principiante.
-

Presentación del método, utilidad y retos de aplicación

La evaluación realista es un enfoque de la evaluación basado en la teoría (Stern y otros, 2012). Responde no sólo a la clásica pregunta “¿funciona o no una intervención para producir un efecto?” sino, más importante aún, responde a importantes preguntas de toma de decisiones como “¿Por qué funciona o no la intervención? ¿Cómo funciona? ¿Para quién y en qué contexto?” (Ridde y Dagenais 2012). Trata de explicitar de manera sistemática los mecanismos causales de los resultados de una intervención para diferentes poblaciones y en diferentes contextos y, por lo tanto, permite la adaptación óptima de los programas sociales a diferentes grupos de personas. De hecho, es un excelente enfoque para promover la equidad.

La evaluación realista es más apropiada para las intervenciones complejas implementadas en sistemas abiertos que cambian constantemente (Marchal et al. 2012). Es adecuado para evaluar nuevas iniciativas, intervenciones piloto e intervenciones con resultados mixtos, cuyos mecanismos de éxito o fracaso se desconocen para comprender las diferencias (Westhorp 2014). La evaluación realista también es adecuada para evaluar las intervenciones en las que hay poco consenso entre los interesados sobre el enfoque, en las que se conoce muy poco sobre las posibles opciones o en las que los conocimientos especializados son limitados. En este caso, la evaluación realista permite un enfoque de aprendizaje sistemático sobre cuestiones clave importantes para orientar la política en su desarrollo, aplicación y mejora progresiva.

Esta forma de evaluación rompe fundamentalmente con los enfoques convencionales porque se basa en los conceptos de causalidad generativa y en la corriente filosófica del realismo (Pawson y Tilley 1997). Según el modelo de causalidad generativa, una acción es causal si, y sólo si, sus efectos son desencadenados por un mecanismo que actúa en un contexto específico. Este mecanismo causal es el foco central del trabajo del evaluador realista. La intervención no se percibe como causal: proporciona recursos que, en interacción con recursos preexistentes en el contexto, pueden o no generar un mecanismo causal que desencadene un efecto. Así pues, la causalidad generativa difiere del modelo de causalidad sucesional, que establece que un efecto E puede atribuirse a una intervención I si se puede demostrar

la constante conjunción entre E e I. El matiz entre estos dos modelos de causalidad es esencial para definir y comprender la evaluación realista (Van Belle et al. 2016).

La evaluación realista forma parte de la corriente filosófica del realismo, que reconoce que la realidad existe independientemente de las interpretaciones humanas o del conocimiento de esa realidad (Audi 1999). El realismo filosófico es distinto del positivismo y el constructivismo, que están en la raíz de los enfoques de evaluación de programas más utilizados (Pawson 2006; Westhorp y otros 2011). Para comprender mejor la evaluación realista, es esencial ser consciente de estas corrientes y sus implicaciones para la evaluación. Un primer libro de Ridde y Dagenais discute ampliamente estos fundamentos teóricos y sus implicaciones prácticas (Ridde y Dagenais 2012).

En este capítulo presentamos la aplicación de este enfoque a la evaluación de la adopción de una política de salud pública (Goodnough et al. 1993). La adopción forma parte de la aplicación, una fase compleja del ciclo de políticas públicas que es una de las más cruciales, pero lamentablemente una de las menos estudiadas en el África subsahariana (Ridde y Olivier de Sardan 2017). Presentamos la metodología de este ejercicio siguiendo el ciclo de evaluación realista (Marchal et al. 2012). Presentamos las etapas del ciclo y la aplicación que hicimos de él. Parte de nuestros resultados se presentarán para ilustrar cómo se “construyeron” a través de las diferentes etapas del ciclo. Finalmente, discutimos algunas de las lecciones aprendidas.

Contexto y programa evaluado: la política de cesáreas gratuitas en Benin

En abril de 2009, Benin introdujo una política de exención de tasas para las mujeres que requieren cesáreas en 48 hospitales de todo el país. Estos hospitales deben ofrecer cesáreas sin cobrar directamente a los pacientes. A cambio, el gobierno reembolsa a los hospitales una suma global de 100.000 francos CFA (178 dólares) por cada cesárea realizada. Los gastos de remisión previa a la cesárea, la consulta sobre el ingreso en el hospital, el procedimiento de cesárea, los medicamentos y consumibles, la hospitalización y los cuidados postoperatorios se tienen teóricamente en cuenta según el decreto que instituyó la política (Centre de Recherche en

Reproduction Humaine et en Démographie 2013). Una evaluación preliminar había medido los resultados de la aplicación en siete hospitales y demostró que el número de elementos del paquete gratuito, que en realidad los hospitales ofrecen de forma totalmente gratuita, varía de 1 en 8 (Hospital 2) a 5 en 8 (Hospital 5) (Tabla 1). Esta variabilidad en la adopción de políticas es un desafío clásico que se observa en la aplicación de varias políticas sanitarias en el África subsahariana.

Tabla 1 : Adopción variable de la política de cesáreas gratuitas en 7 hospitales de Benin, 2012

	H1	H2 (Caso 1)	H3	H4	H5 (Caso 2)	H6	H7	Número de hospitales que ofrecen el artículo de forma gratuita
Traslado al distrito de salud								0/7
Enfoque venoso con solutos antes de la transferencia								0/7
Consulta								2/7
Acto quirúrgico								6/7
Medicamentos								1/7
Consumibles								2/7
Hospitalización								5/7
Control postoperatorio								6/7
Artículos gratuitos sobre el número total de artículos	3/8	1/8	2/8	3/8	5/8	4/8	4/8	

Fuente: Centro de Investigaciones sobre la Reproducción Humana y la Demografía 2014

- Caja negra: Totalmente pagada
- Caja de color gris: Parcialmente libre
- Caja blanca : Totalmente libre

Protocolo de evaluación: El ciclo de evaluación realista

La evaluación realista toma un enfoque cíclico que comienza con una teoría de programa inicial y termina con una teoría de programa refinada. Estas teorías son teorías de medio alcance (MRT), definidas por Merton (1968, p. 39) como teorías que

cuyo nivel se sitúa entre las hipótesis de trabajo menores pero necesarias que se desarrollan diariamente en la investigación y los esfuerzos sistemáticos de integración para desarrollar una teoría unitaria que explique todas las uniformidades observadas en el comportamiento social, la organización social y el cambio social”.

Así pues, el proceso realista tiene por objeto producir teorías de programas del tipo del TMP cada vez más refinadas, que se pueden aplicar cada vez mejor a contextos cada vez más variados y amplios, o explicar un fenómeno social determinado de manera más refinada y precisa (Marchal et al. 2012).

Etapas del Ciclo de Evaluación Realista: Formulación de la Cuestión de Investigación

El enfoque de la formulación de la pregunta de evaluación es similar al de todas las demás formas de evaluación, con la diferencia de que la perspectiva realista de evaluación amplía considerablemente la gama de preguntas que pueden seleccionarse. El valor añadido es mayor en preguntas como “¿Por qué, cómo, para quién y en qué contextos o circunstancias? ».

Etapas del ciclo de evaluación realista: Desarrollo de la teoría del programa inicial

La teoría del programa inicial puede ser desarrollada usando uno o más procesos (Figura 1).

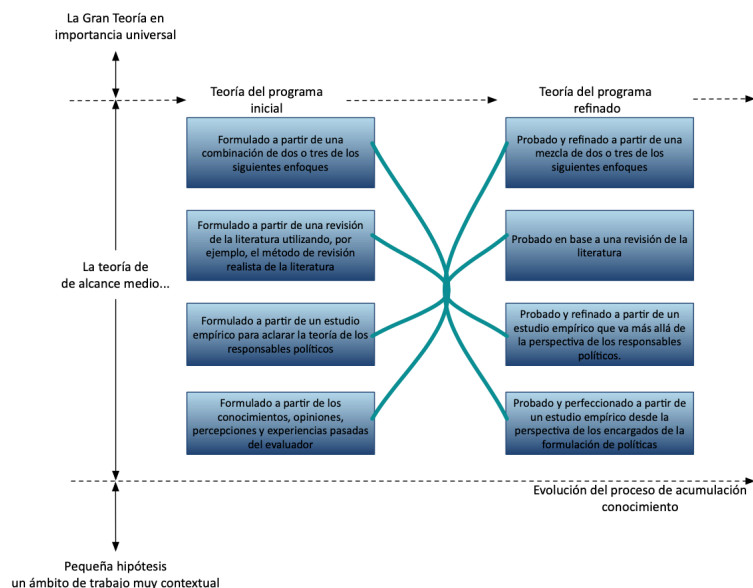


Figura 1. Ilustración de las teorías de rango medio y los diferentes enfoques para formular y refinar la teoría de los programas (adaptado de Merton (1968) y Marchal (et al. 2013))

La elección de la estrategia para formular la teoría del programa inicial depende de la naturaleza de la intervención, la amplitud de los conocimientos del evaluador sobre el tema, los objetivos de la evaluación, los recursos disponibles y la naturaleza de los productos prioritarios que se esperan de la evaluación. Usar una combinación de los diferentes enfoques es el modelo ideal.

Etapas 3 del Ciclo de Evaluación Realista: Selección del Diseño del Estudio

La evaluación realista no impone ninguna estimación de estudio. Esta elección depende principalmente de las cuestiones que se están considerando y de los supuestos que conforman la teoría del programa original. Sin embargo, es esencial utilizar especificaciones que permitan modelos de inferencia configurables, no modelos de inferencia basados en variables (Van Belle et al. 2016). En relación con las preguntas de evaluación,

la evaluación realista utiliza con mayor frecuencia metodologías mixtas, combinando métodos cuantitativos y cualitativos en diversos grados para responder a las preguntas de estudio (Ridde y Dagenais 2012).

Etapla 4 del ciclo de evaluación realista: Recopilación de datos

La recopilación de datos se hace de acuerdo con los métodos elegidos para cualquier evaluación. Hay una forma especial de entrevista cualitativa llamada “entrevista realista”. En este enfoque, el investigador comparte su teoría con los interesados y les pide su opinión sobre las diversas hipótesis así enunciadas. Al hacerlo, el investigador da su teoría a los entrevistados que comparten con él los elementos de su teoría según el principio “Dame tu teoría y yo te daré la mía”. Esta técnica no es una prescripción absoluta.

Etapla 5 del ciclo de evaluación realista: Análisis de los datos

El analista realista identifica los mecanismos de los datos empíricos mediante la argumentación inductiva o deductiva y formula por retroducción los vínculos implícitos entre las condiciones contextuales y los mecanismos por los que es probable que se expliquen los resultados observados.

La retroducción (o abducción) es un modo de razonamiento para el descubrimiento, distinto de la inducción y la deducción. Se parte de los resultados y de las pruebas contextuales observables para formular explicaciones plausibles. Por consiguiente, el proceso de análisis realista puede adoptar la siguiente forma: en lo que respecta a los resultados R, en presencia de los agentes A y los elementos contextuales C1, C2 y C3, estos resultados sólo pueden observarse si existe un razonamiento o una reacción de los agentes (mecanismo subyacente M) acerca de los recursos proporcionados por la intervención I. La identificación de estos agregados de Intervención-Contexto-Actor-Mecanismo de Resultados (ICAMO) (Marchal et al., 2018) es el trabajo esencial del analista realista.

La teoría del programa guía al analista realista que también se asegura de estar abierto a nuevas explicaciones que surgen de los datos. Los pasos clave del proceso de análisis son: 1) identificación de los resultados, 2) descripción de la intervención en términos de contenido, lugar, tiempo e intensidad, 3) identificación de los actores involucrados o no en el diseño y la implementación, 4) identificación de los vínculos implícitos entre las prácticas en torno a la intervención y los resultados observados (los mecanismos), 5) identificación de las características del contexto y su

influencia diferencial en el desencadenamiento de los mecanismos y, por lo tanto, en los resultados, y 6) búsqueda activa de explicaciones alternativas. Estos pasos deberían permitir identificar los elementos de las configuraciones de ICAMO.

Para el análisis de los datos cualitativos deben respetarse criterios de calidad como el análisis iterativo, acumulativo, adaptativo y reflexivo (Teddle y Tashakkori, 2009).

Etapas del Ciclo de Evaluación Realista: Síntesis

Las configuraciones se sintetizan entonces de forma narrativa o gráfica. La formulación narrativa puede adoptar la estructura básica: en un contexto C, se observa un resultado R si un mecanismo M es activado en los actores A por la intervención I.

En la síntesis, los mecanismos y sus condiciones de ejercicio destacados se confrontan con la teoría del programa inicial, que se reformula así sumando las hipótesis pertinentes que faltan. Las hipótesis invalidadas pueden ser reformuladas negativamente o eliminadas. En el enfoque metodológico, sugerimos en esta etapa presentar en la primera columna de un cuadro las hipótesis de la teoría del programa inicial y en la segunda columna la nueva formulación de esas hipótesis poniendo de relieve las alteraciones de perfeccionamiento realizadas. Se trata de una buena práctica recomendada en el Informe de Evaluación Realista (Wong et al. 2014). La síntesis de este ciclo se puede informar en una figura como la que se muestra en la figura 2 al final de este capítulo.

Aplicando el enfoque de evaluación realista

Para evaluar de manera realista la adopción de la política de cesáreas gratuitas en los hospitales de Benin pueden ser necesarios varios ciclos, en los que cada ciclo explica con mayor detalle uno o más aspectos de las variaciones en la adopción de la política de un hospital a otro. Los ciclos subsiguientes también pueden adaptarse a medida que surjan cuestiones de los ciclos anteriores. El ejemplo que se presenta aquí se centra en un ciclo.

Etapas del Ciclo de Evaluación Realista: La cuestión de la investigación

Para la evaluación de la aplicación de la política de cesáreas gratuitas en Benin, la pregunta de investigación fue la siguiente: ¿Cómo, por quién y en qué circunstancias se adoptó la política de cesáreas gratuitas en Benin en los hospitales y por qué?

Etapas 2 del Ciclo de Evaluación Realista: Teoría del Programa Inicial

Para desarrollar la teoría inicial del programa para la política en estudio, realizamos una revisión bibliográfica no sistemática que se centró en las teorías, los marcos analíticos y los modelos con más probabilidades de responder a la pregunta de investigación de forma aislada o de forma complementaria y coherente. La selección de las publicaciones siguió la estrategia de “recolección de bayas” (Booth 2008). Este enfoque iterativo y flexible comienza con una pregunta de búsqueda de literatura. El investigador realiza una búsqueda bibliográfica, selecciona un primer conjunto de artículos pertinentes a la cuestión de la investigación, analiza los conocimientos de esas publicaciones, formula otra pregunta más refinada sobre el mismo tema, realiza otra recopilación de datos, y así sucesivamente, hasta que se identifica una cuestión de investigación definitiva, coherente y complementaria (Booth 2008).

Este enfoque iterativo y flexible comienza con una pregunta de búsqueda de literatura. El investigador realiza una búsqueda bibliográfica, selecciona una primera serie de artículos pertinentes a su pregunta, analiza los conocimientos derivados de esas publicaciones, formula otra pregunta más refinada sobre el mismo tema, procede a otra recopilación de datos, y así sucesivamente, hasta llegar a un conocimiento sintético que satisfaga pragmáticamente sus necesidades de conocimiento.

Por lo tanto, comenzamos a seleccionar publicaciones en la literatura sobre políticas de atención de la salud materna exentas de pago en el África subsahariana. Encontramos una relativa falta de publicaciones que desarrollen teorías, marcos analíticos o modelos relevantes (Ridde y Olivier de Sardan 2016). Por lo tanto, decidimos ampliar nuestra investigación al campo más amplio de la aplicación y gestión de políticas públicas.

Cabe señalar que esta evaluación se llevó a cabo en el marco del estudio FEMHealth, que es una evaluación multinacional en la que los interesados se encuentran en diferentes niveles de la pirámide de la salud. Pensamos que una buena teoría de programa inicial en tal contexto de evaluación debería:

- tener una arquitectura básica lo suficientemente amplia como para permitir la recolección estructurada de los diferentes niveles de la pirámide de la salud y de los diferentes grupos de interesados
- ser fácil de entender y comunicar, de modo que pueda estructurar fácilmente el pensamiento, los debates y los conocimientos generados con múltiples interesados a nivel nacional.

Este proceso dio lugar a la selección de tres publicaciones clave (Berman 1978, Matland 1995 y Elmore 1979) que sirvieron de base para el desarrollo de la teoría del programa inicial. Esas publicaciones tenían por objeto explícitamente la elaboración de modelos para la aplicación de políticas públicas y cumplían los requisitos establecidos anteriormente en relación con el contexto de la evaluación. La teoría del programa inicial resultante fue formulada narrativamente como sigue (Recuadro 1).

Recuadro 1. Teoría del programa inicial que explica las variaciones en la adopción de políticas de salud de una estructura de implementación a otra

Un programa de salud es adoptado a nivel operacional por los administradores locales que se encargan de aplicarlo y que pueden hacerlo de manera variable. Los factores personales, así como los factores contextuales locales, desempeñan un papel, pero también son importantes otros elementos a nivel de políticas, a nivel de su traducción en un programa y a nivel de la comunidad.

Hay tres escenarios posibles en la adopción de un programa por parte de los administradores locales. El primer escenario es la decisión de adoptar y aplicar el programa sin modificaciones. Esto es más probable si el programa se considera una extensión de los objetivos locales (alineación), es decir, si satisface las necesidades locales y proporciona los recursos necesarios. El segundo escenario es la adaptación (positiva) al contexto local; es probable que ocurra cuando los administradores locales estén convencidos de que el programa puede mejorarse. Esto también requiere que tengan la capacidad y la libertad de ajustar el programa a la situación local. Ese ajuste puede dar lugar a mejores resultados en la aplicación de políticas. La tercera hipótesis es la desviación de programas; esto puede ocurrir si el programa se considera un medio de crear oportunidades personales u organizativas y si hay poco "apoyo" para su ejecución y poca contrapresión local. El escenario real final depende de la presión local, pública y política (lo que a su vez requiere que los representantes de la comunidad conozcan la política y dispongan de canales efectivos para ser escuchados), y del compromiso personal de los gestores locales. La aplicación sostenible requiere la institucionalización a nivel local.

En el plano normativo, son factores influyentes la participación de los administradores y los beneficiarios en el proceso de adopción de decisiones y la medida en que se definen claramente los objetivos de las políticas y los grupos destinatarios.

Entre los factores a nivel de los programas figura la existencia de instrucciones operacionales claras (incluida la determinación de los objetivos de política, los grupos destinatarios, las disposiciones prácticas para la aplicación del reembolso, etc.). Si se abordan adecuadamente las cuestiones relativas a los trabajadores de la salud, como la compensación por la pérdida de ingresos, la capacitación, la (re)distribución, etc., es más probable que se adopte y aplique el programa.

Pasar de un nivel del sistema a otro puede cambiar la política:

- Un buen consenso sobre los objetivos, los grupos destinatarios y las modalidades de aplicación (bajo nivel de ambigüedad y conflicto) puede facilitar la transición de un nivel del sistema a otro si cada nivel cuenta con la capacidad y los recursos necesarios.
- Una débil capacidad de vigilancia y aplicación puede conducir a una aplicación simbólica si existe un fuerte sentido de profesionalismo (dependiendo de la voluntad de los profesionales locales).
- Si no hay consenso sobre los objetivos, los grupos destinatarios y las modalidades de aplicación, la política puede enredarse en conflictos de poder o transformarse a nivel local.

Etapas 3 del Ciclo de Evaluación Realista: El Diseño del Estudio

Elegimos la cita del estudio de caso integrado (Yin 2003) que se adapta bien a situaciones complejas (Marchal 2013). El objetivo de la fase de elaboración de instrumentos y de recogida de datos era reunir de manera óptima datos empíricos que permitieran formular conclusiones de configuración. Adoptamos un método mixto simultáneo y anidado (concurrente anidado) con un componente cuantitativo más pequeño integrado en un estudio predominantemente cualitativo. Los componentes de la teoría del programa guiaron el desarrollo de esta herramienta.

Se consideró que la aplicación de la política de cesáreas gratuitas a nivel nacional era el estudio de caso. La elección de las unidades de análisis se hizo de manera razonada para poder “probar” las hipótesis expresadas por la teoría del programa inicial. Se seleccionaron como unidades de análisis dos hospitales con resultados contrastados en la aplicación de políticas (Caso 1 y Caso 2 de la Figura 1).

Etapas 4 del Ciclo de Evaluación Realista: Recolección de datos

Combinamos una revisión de la literatura con la extracción de datos cuantitativos, entrevistas semiestructuradas, un estudio de costos y el seguimiento de los flujos financieros para responder a una serie de preguntas intermedias que pretendían contribuir a la respuesta de la pregunta central. En el cuadro siguiente se presentan las fuentes de datos y los instrumentos para los estudios de casos.

Tabla 2 : Fuentes de datos, técnicas de recopilación y muestra por unidad de análisis

Fuentes de datos	Técnicas	Caso 1: Hos pital privado de base religiosa	Caso 2: Hos pital público
Gestores de salud locales	Entrevistas semiestructuradas	4	4
Los administradores de hospitales y los proveedores de atención médica	Entrevistas semiestructuradas	15	8
Representantes de la comunidad en los comités locales de salud	Entrevistas semiestructuradas	6	3
Mujeres embarazadas a las que se dirige la política	Entrevistas semiestructuradas	6	6
	Encuestas de producción	28	17
Informes sobre la política y diversos documentos contables en el Organismo Nacional de Cesáreas Libres y en las estructuras de ejecución	Vigilancia de los estados financieros	1	1
Datos del Sistema de Información de Salud, proveedores de atención de salud y administradores locales de salud	Revisión de la literatura con extracción de datos para estudios de costos	30	30
Implementación hospitalario Entorno	Observación no participante de las prácticas en torno a la política	1	1

La recogida de datos fue llevada a cabo por un equipo multidisciplinario de socio-antropólogos, economistas de la salud y médicos. Duró desde marzo de 2012 hasta septiembre de 2012.

Un elemento clave fue el estudio de las percepciones, experiencias y opiniones de los diversos interesados a todos los niveles: integró las observaciones preliminares realizadas con todos los demás instrumentos de manera iterativa, acumulativa, adaptable y flexible.

Las frecuentes interacciones dentro del equipo de recopilación permitieron integrar los datos pertinentes producidos por el equipo multidisciplinario con las diferentes técnicas de recopilación y adaptar la colección en consecuencia sobre la marcha.

Etapas 5 y 6: Análisis y síntesis

Los datos cualitativos fueron gestionados por el software NVIVO 10. Se importaron memorandos de recopilación, informes de análisis preliminares de otras técnicas, documentos de política, etc. en el programa informático para formar una única base de datos utilizada para el análisis. Los diversos componentes de la teoría inicial guiaron la codificación. El análisis se llevó a cabo de tal manera que también permitió la adición de nuevos códigos de manera inductiva.

Los análisis preliminares llevados a cabo en las diferentes disciplinas que participan en el estudio fueron importantes para el análisis realista en sí.

En la síntesis, los mecanismos e hipótesis así destacados se enfrentaron a la teoría inicial, que se reformuló añadiendo las hipótesis pertinentes.

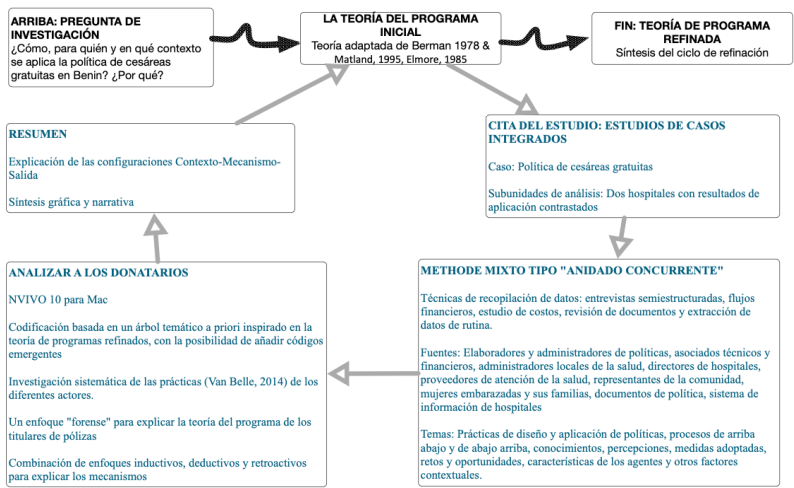


Figura 3. Ciclo de la evaluación realista de la aplicación de la política de cesáreas gratuitas en Benín

Algunos resultados

En esta sección presentamos, de manera ilustrativa, algunos resultados de las diferentes etapas metodológicas. No se trata de resultados exhaustivos sobre el tema que se examina, sino de ejemplos para ilustrar el proceso y la forma en que podrían formularse los resultados de una evaluación realista.

Prácticas en torno a la adopción de la política por parte de los administradores de los hospitales

En el hospital privado religioso

En este hospital, de los ocho elementos del paquete de políticas, sólo el examen postoperatorio estaba totalmente exento. La tarifa media que pagaban los usuarios por una cesárea después de la puesta en marcha de la política era de 48 dólares [mín. 45 dólares; máx. 135 dólares], en lugar de ser completamente gratuita como se prevé en la política promovida por el gobierno.

Este hospital es propiedad de una organización religiosa que es propietaria de la infraestructura, nombra a los administradores, contrata al personal, valida el presupuesto, moviliza recursos externos para el funcionamiento del hospital cuando es necesario y controla la gestión. El hospital funciona principalmente sobre la base de los recursos internos movilizados mediante la recuperación de los costos de los pagos directos. La intervención del Estado en el funcionamiento del hospital es mínima. El Comité de Salud, el órgano local para la participación de la comunidad en el gobierno del sistema de salud local, casi nunca interviene en este hospital. La población está mal informada sobre la política de este hospital y no tiene un canal oficial para la información o las quejas.

El Decreto N° 2008-730, de 22 de diciembre de 2008, por el que se estableció la Política, acreditó a este hospital para la aplicación de la Política tan pronto como se puso en marcha en abril de 2009. Sin embargo, los directores de este hospital no adoptaron y efectivamente comenzaron a ofrecer cesáreas gratuitas hasta septiembre de 2009. Los administradores indicaron que no confiaban en que el gobierno proporcionara reembolsos oportunos como se indica en los documentos oficiales de política. Los gestores justificaron esta desconfianza por sus experiencias anteriores con el gobierno, que les sigue debiendo, desde hace varios años, los recursos financieros exentos a los beneficiarios de la cobertura sanitaria parcial concedida a los funcionarios públicos. Los administradores comenzaron a aplicar la exención del pago de las cesáreas cuando se dieron cuenta de que los demás hospitales privados que aplicaban la política en realidad comenzaban a recibir reembolsos con regularidad y sin demasiada demora.

El estudio de costos demostró que el costo de producción medio (incluidos todos los costos directos) de una cesárea en este hospital religioso

era de 44 dólares [mín. 37 dólares; máx. 69 dólares]. Las tarifas de los usuarios anteriores a la política eran de 163 dólares. Nuestra herramienta de seguimiento del flujo de efectivo mostró que hasta diciembre de 2011, este hospital fue reembolsado por todas las cesáreas realizadas. Los administradores consideraron que la suma fija de 178 dólares era insuficiente e injusta, pues creían que la tasa de reembolso de los hospitales no estatales debía ser más elevada que la de los hospitales estatales en los que las subvenciones públicas cubrían muchos gastos de funcionamiento. Por lo tanto, los gestores decidieron cobrar sistemáticamente a los usuarios 35 dólares por cesárea, además del reembolso recibido del gobierno.

En el hospital público

De los ocho artículos incluidos en el paquete de cesáreas gratuitas, cinco estaban totalmente exentos del costo de la atención en este hospital. El resto de la tarifa que se cobró a los usuarios fue de US\$6,22 [US\$0,1 mín; máx. US\$53,3].

Este hospital es propiedad del Estado, que lo construyó y lo equipó con recursos públicos. El gobierno paga una parte importante del personal y nombra a los directivos. La población está mejor informada sobre la política. Los usuarios utilizan varios canales, como las emisoras de radio comunitarias, para informarse sobre la política o para quejarse si es necesario.

Los gestores de este hospital adoptaron la política rápidamente y comenzaron a implementarla a la semana de su lanzamiento oficial. El director del hospital participó en el taller técnico en el que se elaboraron opciones de política en la fase de elaboración de políticas. Consideró que la política apoyaba su visión de aumentar el acceso financiero a los servicios para los usuarios de su hospital. Tomó medidas activas para facilitar su aplicación con éxito.

En este hospital, se consideraba que la tasa de reembolso cubría en gran medida el costo de la cesárea para el hospital. Los gestores que son funcionarios públicos se sintieron obligados a empezar a aplicar la política, como explicó un gestor:

El público ha sido ampliamente informado de que la cesárea es gratuita y es difícil decir ahora, que no voy a empezar a aplicar

la política. Una mujer no puede venir y tú dices, hoy no hacemos cesáreas gratis, porque el director provincial de salud sabe que las cesáreas son gratis. Si una mujer se queja de que el centro le ha pedido que pague, como gestor, tendrá que explicar por qué. Gestor del Hospital Público, 2012

Algunas configuraciones causales en torno a la adopción se derivan de los resultados

Nuestro análisis de las prácticas declaradas en el hospital privado de base religiosa indica que, en este contexto, los gestores no tienen una relación administrativa jerárquica con el Estado, ya que las presiones de la población son demasiado débiles o están ausentes. En este contexto, la administración pública es percibida como lenta, poco creíble en el cumplimiento de sus compromisos financieros y despierta una actitud de desconfianza, a pesar de que la movilización de recursos de la prestación de cuidados es esencial para el funcionamiento del hospital. El pago directo es sistemático, está fuertemente arraigado en la cultura de la organización y se considera una medida crucial de la supervivencia de la organización. En este contexto, la política de exención del pago de los derechos de atención de salud (Intervención) es percibida por los administradores de los hospitales (Actores) como una amenaza para la supervivencia de los hospitales (Mecanismo) y, de hecho, un escenario no óptimo en la adopción de la política (Resultado).

El éxito de la adopción (Resultado) en este contexto requiere que se perciba que la política protege los intereses financieros de los hospitales o, en el mejor de los casos, que aporta un valor añadido financiero a los hospitales (Mecanismo), mediante el suministro de recursos más fácilmente recuperables, más abundantes o más regulares (Intervención).

En el contexto de los hospitales públicos, nuestro análisis muestra que es relativamente más probable que los administradores cumplan con las órdenes administrativas jerárquicas del Estado. La participación de la comunidad en la gestión de los hospitales es relativamente más importante, ya que las poblaciones están bien informadas sobre la política y tienen canales para expresar sus quejas. Esta doble presión, administrativa de arriba

hacia abajo y comunitaria de abajo hacia arriba sobre los administradores de los hospitales (Actores) genera coerción (Mecanismo) que lleva a una mejor adopción de la política (Resultado).

Las inversiones estructurales del Estado antes de la iniciación de la política en el hospital público son más importantes que en el hospital privado (Contexto). En este contexto, el reembolso de la misma suma global por acto (Intervención) genera un sentimiento de injusticia (Mecanismo) entre las estructuras no estatales (Actores) que tienden a compensar esta injusticia mediante el cobro de tasas adicionales a los usuarios, ofreciendo así sólo una parte de los servicios de forma gratuita (Resultado).

Síntesis y debate: teorías sobre la adopción de políticas

A través de las configuraciones causales formuladas anteriormente, podemos ver que la adopción depende de la confianza entre las partes interesadas y el promotor de la política de exención, por ejemplo, el Estado. Esta confianza es tanto más necesaria si se pone en duda la credibilidad del promotor sobre la base de la experiencia previa. Entre las medidas que deben adoptarse para tranquilizar a los administradores de los hospitales figura el suministro oportuno de recursos apropiados para la prestación del servicio, por lo menos como compensación suficiente por la pérdida de recursos generada por la exención del pago.

Esta explicación es coherente con la teoría del comportamiento planificado, que afirma que es poco probable que los individuos desarrollen una fuerte intención de actuar y comportarse de cierta manera si sienten que carecen de los recursos y las oportunidades para hacerlo, incluso si tienen una actitud favorable hacia el comportamiento en cuestión e incluso si sienten que los miembros de la familia o los compañeros humanos aprobarían el comportamiento (normas subjetivas) (Ajzen y Fishbein 1980). Así pues, aun cuando la voluntad política y los mandatos administrativos infundan la formación de nuevas normas y estructuras de referencia, es poco probable que la adopción de la política sea una realidad si no tiene las características requeridas y si no proporciona suficientes recursos garantizados presentes y futuros para fortalecer el control que se percibe de los agentes sobre su aplicación.

La percepción de justicia en la distribución de los recursos es crucial en los casos en que las estructuras de ejecución están sujetas a diferentes derechos de propiedad y a una financiación variable. Es probable que se produzca una sensación de injusticia si la política no define claramente los costos reales que entraña la producción de cada acto por tipo de establecimiento sanitario y si no define explícitamente el margen excedente concedido a cada unidad de ejecución. Esta necesidad percibida de equidad en la adopción no está presente en la teoría del comportamiento planificado (Ajzen y Fishbein 1980), en el modelo de difusión de innovaciones de Rogers (1995), ni en el modelo de aceptación de la tecnología (Davis 1989). La influencia de esta sensación de injusticia en la adopción de políticas puede analizarse como una presión isomórfica mimética (Zucker 1988), en la que los actores de las estructuras sanitarias no estatales aprenden sobre los beneficios derivados de la política de las estructuras sanitarias estatales y buscan beneficios similares a través de su comportamiento.

Este análisis nos llevó a reformular la teoría original del programa (Recuadro 2).

Recuadro 2. Teoría de programa refinada

Un programa de salud es adoptado a nivel operacional por los administradores locales de salud que se encargan de su aplicación y que pueden hacerlo en diversos grados. Los factores personales, así como los factores contextuales locales, desempeñan un papel, pero también son importantes otros elementos a nivel de políticas, su traducción en un programa y a nivel de la comunidad.

Hay cuatro escenarios posibles en la adopción de un programa por parte de los administradores locales. El primer escenario es la decisión de adoptar y aplicar el programa sin modificaciones. Esto es más probable si el programa se considera una extensión de los objetivos locales (alineación), es decir, si satisface las necesidades locales y proporciona los recursos necesarios. El segundo escenario es la adaptación (positiva) al contexto local. Es probable que ocurra cuando los administradores locales estén convencidos de que el programa puede ser mejorado. También requiere que tengan la capacidad y la libertad de ajustar el programa a la situación local. Ese ajuste puede dar lugar a mejores resultados en la aplicación de políticas. La tercera hipótesis es la desviación de programas; esto puede ocurrir si el programa se percibe como un medio de crear oportunidades personales u organizativas y si hay poco "apoyo" para su ejecución y poca contrapresión local. Un cuarto escenario posible es la decisión de adoptar y poner en práctica una versión reducida del programa en la que sólo se ofrezca gratuitamente una parte de los servicios. Es probable que esta situación se produzca incluso si los recursos proporcionados por la política son suficientes. Es más probable que los administradores se sientan en desventaja en la asignación de los recursos proporcionados por la política en comparación con otras estructuras de ejecución. El escenario real final depende de la presión local, pública y política (lo que a su vez requiere representantes de la comunidad que conozcan la política y dispongan de canales eficaces para ser escuchados), la confianza de los administradores de las estructuras de ejecución en el compromiso del proponente de la política y el compromiso personal de los administradores locales con el éxito de la política. La aplicación sostenible requiere la institucionalización a nivel local.

En el plano normativo, son factores influyentes la participación de los administradores y los beneficiarios en el proceso de adopción de decisiones y la medida en que se definen claramente los objetivos de las políticas y los grupos destinatarios.

Entre los factores a nivel de los programas figura la existencia de instrucciones operacionales claras (incluida la determinación de los objetivos de política, los grupos destinatarios, las disposiciones prácticas para la aplicación del reembolso, etc.). Si se abordan adecuadamente las cuestiones relativas a los trabajadores de la salud, como la compensación por la pérdida de ingresos, la capacitación, la (re)distribución, etc., es más probable que se adopte y aplique el programa.

Pasar de una fase a otra puede cambiar la política:

- Un buen consenso sobre los objetivos, los grupos destinatarios y las modalidades de aplicación (bajo nivel de ambigüedad y conflicto) puede facilitar la transición de una fase a otra si cada nivel cuenta con la capacidad y los recursos necesarios.
 - Una débil capacidad de vigilancia y aplicación puede conducir a una aplicación simbólica si existe un fuerte sentido de profesionalismo (dependiendo de la voluntad de los profesionales locales).
 - Si no hay consenso sobre los objetivos, los grupos destinatarios y las modalidades de aplicación, la política puede enredarse en conflictos de poder o transformarse a nivel local.
-

En esta refinada teoría de los programas, se añaden nuevos supuestos a la teoría original de los programas, por ejemplo, sobre el papel de los sentimientos de injusticia y confianza en la adopción de la política. El perfeccionamiento también implicó una modificación de una hipótesis existente. Por ejemplo, el número de posibles escenarios de adopción se redujo de tres a cuatro.

Esta refinada teoría del programa proporciona una respuesta plausible a la pregunta de evaluación. Los encargados de la formulación de políticas pueden utilizarla de manera práctica para mejorar el diseño y la aplicación de las políticas. Esta teoría del programa también puede ser utilizada para mejorar el diseño de otras políticas similares.

Sin embargo, esta sigue siendo una teoría programática que de ninguna manera pretende ser universal o transferible a todos los contextos o tipos de intervenciones. Es inherentemente susceptible de mejorar en ciclos de evaluación posteriores.

Lecciones aprendidas y debate sobre cuestiones de aplicación de la evaluación

- Requiere mucha preparación. Aunque esta fase no tiene necesariamente aspectos fundamentalmente diferentes de la preparación de otras formas de evaluación, debe reunir toda la información posible sobre los asociados locales en la investigación, la intervención que se va a evaluar, los interesados en torno a la intervención y los contextos de desarrollo y aplicación. Sugerimos en esta etapa distinguir claramente entre las dos situaciones siguientes:
 - se elige desarrollar la teoría del programa inicial a partir de la de los diseñadores de la intervención. En este caso, no es esencial que el evaluador finalice una teoría inicial del programa antes de iniciar las actividades sobre el terreno. Sin embargo, el evaluador debe tener suficientes conocimientos teóricos y metodológicos para explicar la teoría del programa inicial de los diseñadores de manera realista.
 - se elige desarrollar la teoría del programa inicial de forma independiente o como complemento de la de los diseñadores de la intervención. En este caso, es esencial dedicar tiempo a la revisión de la literatura.
- Al comienzo de la evaluación debe dominarse una definición clara del concepto de “mecanismo causal”. Este es uno de los principales desafíos para comprender la evaluación realista.
- La distinción entre el mecanismo, el contexto y los resultados puede a veces convertirse en un rompecabezas y una fuente de confusión. Hemos aprendido que la definición más operativa es la de “el

razonamiento y las reacciones de los actores sobre los recursos aportados por la intervención”. En otras palabras, los mecanismos se ocultan en las percepciones, opiniones y lógicas implícitas o explícitas de los actores que subyacen a sus actitudes y acciones en relación con los recursos proporcionados por la intervención, lo que determina los resultados (Pawson y Tilley 1997; Lacouture y otros 2015). Estos mecanismos pueden formularse explícitamente en el discurso de los actores o ser objeto de una retroalimentación que debe identificarse.

- Es esencial centrar la recogida de datos en el enfoque de la configuración: a veces se necesitan instrumentos específicos para documentar el contexto, la aplicación de la intervención, los resultados de la intervención y los mecanismos, todo ello basado en la teoría del programa inicial. Durante el análisis es esencial mantener una visión integrada de esos componentes mediante la triangulación de los datos para identificar las configuraciones en las que interactúan la intervención, los agentes, los mecanismos, los elementos contextuales y los resultados.
- Refinar la teoría del programa es un proceso interminable. El objetivo final de un ciclo no es producir una teoría de programa universal, sino útil para generar conocimientos que sean relevantes para los interesados o los investigadores. Esta mirada pragmática al propósito de un ciclo de evaluación realista es esencial para orientar los esfuerzos y evitar quedar atrapado en un proceso interminable.

Referencias clave

Pawson, R. et Tilley, N. (1997). *Realistic evaluation*. Londres : Sage Publications.

Los autores de este libro son los padres fundadores de la evaluación realista. En este libro se explican detalladamente los fundamentos de la evaluación realista y se ilustran los conceptos clave con ejemplos de programas sociales en diferentes campos de los países desarrollados.

Marchal, B., Van Belle, S., Van Olmen, J., Hoerée, T. et Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192–212.

En esta publicación se examina el uso de la evaluación realista en la investigación de los sistemas de salud. Describe, entre otras cosas, los diferentes modelos de uso, las confusiones conceptuales y los desafíos metodológicos observados. En él se propone el ciclo de evaluación realista y se describen las diversas medidas para reducir la confusión y facilitar la estructuración del enfoque metodológico.

Lacouture, A., Breton, E., Guichard, A. et Ridde, V. (2015). The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science*, 10, 153-162.

En la presente publicación se ofrece una síntesis de las diferentes definiciones operacionales del concepto de “mecanismo” en la evaluación de las políticas de salud pública. Es útil para una mejor comprensión de este concepto clave y puede ayudar a evitar posibles confusiones entre los conceptos de “mecanismo”, “contexto” y “resultados”.

Sitio <http://www.ramesesproject.org/>

Este sitio presenta los productos de los proyectos Realist And Meta-narrative Evidence Syntheses: Evolving Standards (RAMESES) I y II. RAMESES I ha elaborado normas de calidad y publicación y materiales de capacitación para revistas realistas, así como enfoques conexos para revistas metanarrativas. RAMESES II ha elaborado normas de calidad y publicación y materiales de capacitación para una evaluación realista. Este sitio está en inglés.

Referencias

- Ajzen, I. et Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ : Prentice-Hall.
- Audi, R. (1999). *The Cambridge dictionary of philosophy* (2e éd.). Cambridge, NY: Cambridge University Press.
- Berman, P. (1978). *The study of macro and micro implementation of social policy*. Santa Monica, CA : RAND Corporation.
- Booth, A. (2008). Unpacking your literature search toolbox: on search styles and tactics. *Health Information et Libraries Journal*, 25(4), 313-317.

- Centre de Recherche en Reproduction humaine et en Démographie. (2014). *Rapport de synthèse : l'évaluation de la politique de gratuité de la césarienne dans cinq zones sanitaires au Bénin*. Cotonou, Bénin: University of Aberdeen.
- Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B. et Lhussier, M. (2015). What's in a mechanism?: Development of a key concept in realist evaluation. *Implementation Science*, 10, 49-55.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *Management Information Systems Quarterly*, 13, 319-340.
- Elmore, F. R. (1979). Backward Mapping : Implementation Research and Policy Decisions. *Polit Sci Q*, 94, 601-16.
- Goodnough, L. T., Riddell, J., Lazarus, H., Chafel, T. L., Prince, G., Hendrix, D., et Yomtovian, R. (1993). Prevalence of platelet transfusion reactions before and after implementation of leukocyte-depleted platelet concentrates by filtration. *Vox Sanguinis*, 65(2), 103-107.
- Lacouture, A., Breton, E., Guichard, A. et Ridde, V. (2015). The concept of mechanism from a realist approach: a scoping review to facilitate its operationalization in public health program evaluation. *Implementation Science*, 10, 153-162.
- Marchal, B., Kegels, G. et Van Belle, S. (2018). Realist evaluation in health policy and systems research: theory incarnate. Dans N. Emmel, J. Greenhalgh, A. Manzano, M. Monaghan et S. Dalkin (dir.), *Doing Realist Research* (édition 2018, volume 1). Londres, Angleterre: Sage Publications.
- Marchal, B., Van Belle, S., De Brouwere, V. et Witter, S. (2013). Studying complex interventions: reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco. *BMC Health Services Research*, 13, 469-477.
- Marchal, B., Van Belle, S., Van Olmen, J., HoerÇe, T. et Kegels, G. (2012). Is realist evaluation keeping its promise?: A review of published empirical studies in the field of health systems research. *Evaluation*, 18(2), 192-212.
- Marchal, B., Van der Veken, K., Essolbi, A., Dossou, J. P., Richard, F. et Van Belle, S. (2013). *Methodological reflections on using realist evaluation in a study of fee exemption policies in West Africa and Morocco*. Aberdeen : University of Aberdeen.

- Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, 5(2), 145-174.
- Merton, R. K. (1968). *Social theory and social culture*. New York : The Free Press.
- Pawson, R. (2006). *Evidence-based policy: a realist perspective*. London; Thousand Oaks, Calif. : SAGE Publications.
- Pawson, R. et Tilley, N. (1997). *Realistic evaluation*. Londres : SAGE Publications.
- Ridde, V. et Dagenais, C. (2012). *Approches et pratiques en évaluation de programmes*. Montréal : Presses de l'Université de Montréal.
- Ridde, V. et Olivier de Sardan J. P. (2016). The implementation of public health interventions in Africa: a neglected strategic theme. *Médecine et Santé Tropicales*, 27(1), 6-9.
- Ridde, V. et Olivier de Sardan J. P. (2017). The implementation of public health interventions in Africa: a neglected strategic theme. *Médecine et Santé Tropicales*, 27(1), 6-9.
- Rogers, E. M. (1995). *Diffusion of innovations* (4e éd.). New York : Free Press.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. et Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. Londres : Department for International Development.
- Teddlie, C. et Tashakkori, A. (2009). *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Londres : SAGE Publications
- Van Belle, S., Wong, G., Westhorp, G., Pearson, M., Emmel, N., Manzano, A. et Marchal, B. (2016). Can 'realist' randomised controlled trials be genuinely realist?. *Trials*, 17, 313-318.
- Westhorp, G. (2014). *Realist impact evaluation: An introduction*. Londres : A method lab.
- Westhorp, G., Prins, E., Kusters, C., Hultink, M., Guijt, I. et Brouwers, J. (2011). *Realist evaluation: an overview*. Report from an expert seminar with Dr. Gill Westhorp. Wageningen: Wageningen UR Centre for Development Innovation.

- Wong, G., Greenhalgh, T., Westhorp, G. et Pawson, R. (2014). *Development of methodological guidance, publication standards and training materials for realist and meta-narrative reviews: the RAMESES (Realist And Meta-narrative Evidence Syntheses – Evolving Standards) project*. Health Services and Delivery Research, No. 2.30. Southampton, UK : NIHR Journals Library.
- Yin, R. (2003). *Case study research. Design and methods*. Londres : Sage Publications.
- Zucker, L. G. (1988). *Institutional patterns and organizations: culture and environment*. Cambridge, MA : Ballinger Publishing.

Resumen / Résumé / Overzicht / Abstract

La evaluación realista (ER) explica los mecanismos causales de los resultados de la intervención en diferentes contextos. Basado en la evaluación de la implementación de políticas de salud en África Occidental, este capítulo ilustra las diferentes etapas del ciclo de la ER. En el primer paso, se formulan las preguntas de la investigación. El valor añadido de la evaluación realista es óptimo para responder a preguntas como “¿Por qué? ¿Cómo? ¿Para quién? y ¿en qué contextos funciona o no una intervención? ». El segundo paso es la formulación de la teoría del programa inicial, que depende de la naturaleza de la intervención, los objetivos de la evaluación, la naturaleza de los productos prioritarios principales esperados de la evaluación, los recursos disponibles y la familiaridad del evaluador con el tema de la evaluación. La recolección de datos puede basarse en una forma particular de entrevista cualitativa llamada “entrevista realista”. En la práctica, la recolección de datos debe adaptarse a los informantes clave, a sus percepciones y expectativas de los evaluadores para evitar la tendencia de los entrevistados a confirmar sistemáticamente la teoría del evaluador. El analista realista identifica activamente las configuraciones “Intervención-Contexto-Actor-Mecanismo-Resultado”. En la práctica, el razonamiento retroductivo se utiliza para explicar los mecanismos. Una tabla para el seguimiento de los cambios en la teoría original del programa facilita el paso final de la síntesis y el informe de la ER.

L'évaluation réaliste (ER) explicite les mécanismes causaux des résultats d'une intervention dans différents contextes. En s'appuyant sur l'évaluation de la mise en œuvre d'une politique de santé en Afrique de l'Ouest, ce chapitre illustre les différentes étapes du cycle de l'ER. À la première étape, les questions de recherche sont formulées. La valeur ajoutée de l'approche de l'ER est optimale sur des questions de type « Pourquoi? Comment? Pour qui? et dans quels contextes une intervention marche ou non? ». La seconde étape est la formulation de la théorie de programme initiale qui dépend de la nature de l'intervention, les objectifs d'évaluation, la nature des produits prioritaires attendus de l'évaluation, les ressources disponibles et la familiarité de l'évaluateur avec l'objet d'évaluation. La collecte des données peut se baser sur une forme particulière d'entretien qualitatif appelé « entretien réaliste ». En pratique, la collecte de données doit s'adapter aux informateurs clés, à leurs perceptions et attentes vis-à-vis des évaluateurs et évaluatrices pour éviter la tendance à la confirmation systématique de la théorie de l'évaluateur par les interviewés. L'analyste réaliste identifie activement les configurations « Intervention-Contexte-Acteur-Mécanisme-Résultat ». En pratique le raisonnement par rétroduction est utilisé pour expliciter les mécanismes. Un tableau permettant de suivre les changements apportés à la théorie de programme initiale facilite la dernière étape de la synthèse et du rapport de l'ER.

Realistische evaluatie (RE) verklaart de causale mechanismen die de resultaten van een interventie verklaren in verschillende contexten. Op basis van de evaluatie van de implementatie van een gezondheidsbeleid in West-Afrika illustreert dit hoofdstuk de verschillende stadia van de RE-cyclus. In de eerste fase worden de onderzoeksvragen geformuleerd. De toegevoegde waarde van de RE-benadering is optimaal voor vragen als “waarom, hoe, voor wie, en in welke context werkt een interventie of niet?”. De tweede stap is de formulering van de initiële programmatheorie, die afhangt van de natuur van de interventie, de evaluatiedoelstellingen, de aard van de outputs die van de evaluatie worden verwacht, de beschikbare middelen en de bekendheid van de beoordelaar met het doel van de evaluatie. Voor het verzamelen van gegevens kan elke methode gebruikt worden? Een specifieke vorm van interviews, genaamd “realistisch interview” wordt vaak gebruikt. In de praktijk moet het verzamelen van gegevens zich aanpassen aan de informanten, hun percepties en verwachtingen ten opzichte van de

beoordelaars om de neiging tot systematische bevestiging van de initiële theorie door de geïnterviewden te vermijden. De realistische analist identificeert vervolgens de configuraties ‘Interventie-context-actor-mechanisme-resultaat’. In de praktijk wordt retroductie gebruikt om de mechanismen te indetificeren. Een tabel om wijzigingen in de oorspronkelijke programmatheorie bij te houden, vergemakkelijkt de laatste stap van de synthese en rapportage van de RE.

Realist Evaluation (RE) identifies the causal mechanisms explaining the results of a given intervention in different contexts. Based on the evaluation of the implementation of a health policy in West Africa, this chapter illustrates the different stages of the RE cycle. In the first stage, the research question is formulated. The added value of the RE approach is optimal on questions such as “why, how, for whom, and in what contexts does an intervention work or not?”. The second step is the formulation of the initial program theory, which depends on the nature of the intervention being studied, the objectives, and the nature of the main outputs expected from the evaluation, the resources available and the evaluator’s familiarity with the object of evaluation. The Data collection is method neutral and can be based on a particular form of qualitative interview called “realist interview”. In practice, data collection must adapt to key informants, their perceptions and expectations vis-à-vis the evaluators to avoid the tendency for systematic confirmation of the evaluator theory by the interviewees. The realist analyst actively identifies the “Intervention-Context-Actor-Mechanism-Outcome” configurations. In practice, retroductive analysis is used to explicit mechanisms that are implicit by nature. A table to track changes to the initial program theory facilitates the final step of the synthesis and the reporting of the RE.

Jean-Paul Dossou es médico de salud pública e investigador de políticas y sistemas de salud en cuestiones de salud sexual y reproductiva y cobertura sanitaria universal en el África occidental y oriental. Ha estado forjando apasionadamente sus habilidades teóricas y prácticas en la complejidad y las metodologías realistas desde 2012, trabajando estrechamente con investigadores realistas de clase mundial, incluyendo al Pr Bruno Marchal.

Bruno Marchal es jefe de la Dependencia de Organización de Servicios de Salud del Departamento de Salud Pública del Instituto de Medicina Tropical de Amberes (Bélgica). Su labor de investigación y docencia se centra en la gestión estratégica de los servicios de salud, así como en las metodologías de investigación adaptadas a cuestiones sanitarias complejas. Es el autor de las publicaciones más citadas sobre la evaluación realista y está a la vanguardia del desarrollo de la evaluación realista en la salud mundial.

Cita

Jean-Paul Dossou y Bruno Marchal (2020). Evaluación realista. El ejemplo de la adopción de una política de salud pública en Benín. En *Evaluación de las intervenciones sanitarias en salud global. Métodos avanzados*. Bajo la dirección de Valéry Ridde y Christian Dagenais, pp. 481-510. Québec: Éditions science et bien commun y Marseille: IRD Éditions.

Acerca de la casa editorial

Las Éditions science et bien commun son una rama de la Association science et bien commun (ASBC), una organización sin fines de lucro registrada en Quebec desde julio de 2011.

La Association science et bien commun

La misión del ASBC es estimular la vigilancia y la acción de la ciencia abierta para el bien común. Con este fin, se esfuerza por..:

- Defender y promover una visión de la ciencia al servicio del bien común;
- Recopilar, analizar, producir y difundir información sobre la ciencia y su relación con la sociedad;
- Apoyar, promover u organizar experimentos de democratización de la ciencia;
- Organizar experimentos para el debate público sobre diversas facetas de la ciencia;
- Establecer experiencias de encuentro entre el mundo científico y otras esferas sociales (por ejemplo, el mundo artístico, el mundo político, etc.);
- Ofrecer un servicio de orientación para grupos de la sociedad civil en el mundo académico;
- Ofrecer, con sujeción a la Ley de Educación Privada (R.S.Q., c. E-9.1) y su reglamento, formación sobre responsabilidad social, ciencia con los ciudadanos y ética de la ciencia.

En su sitio web encontrará una gran cantidad de información sobre sus actividades y publicaciones. Es posible hacerse miembro de la Asociación de Ciencia y Bienestar Común pagando una modesta cuota.

Para más información, escriba a info@scienceetbiencommun.org, suscríbase a su cuenta de Twitter [@ScienceBienComm](https://twitter.com/ScienceBienComm) o a su página de Facebook: <https://www.facebook.com/scienceetbiencommun>.

Las *Éditions science et bien commun*

Un proyecto editorial innovador cuyos principales valores son :

- la publicación digital de acceso abierto, además de otros formatos
- la multidisciplinariedad, en la medida de lo posible
- el plurilingüismo, que fomenta la publicación en varios idiomas, incluidos los idiomas nacionales africanos o el criollo, además del francés
- la internacionalización, lo que lleva al deseo de reunir a autores de diferentes países o de escribir pensando en un público de diferentes países y culturas
- pero especialmente la justicia cognitiva :
 - cada libro colectivo, aunque sea el acta de una conferencia, debe aspirar a la paridad entre mujeres y hombres, entre jóvenes y mayores, entre autores del Norte y del Sur; en todo caso, todos los libros deben evitar un desequilibrio flagrante entre estos puntos de vista;
 - cada libro, aunque haya sido escrito por una sola persona, debe tratar de incluir referencias tanto a los países del Norte como a los del Sur, ya sea en sus temas o en su bibliografía;
 - cada libro debe tener como objetivo la accesibilidad y la “legibilidad”, minimizando la jerga, incluso si está orientado científicamente y revisado por pares.

El catálogo

El catálogo de las *Éditions science et bien commun* (ESBC) está compuesto por libros que respetan los valores y principios de la ESBC como se ha indicado anteriormente:

- Obras científicas (libros colectivos de todo tipo o monografías), que pueden ser manuscritos originales inéditos de tesis, disertaciones, coloquios, seminarios o proyectos de investigación, reediciones digitales o libros de texto universitarios. Los manuscritos no publicados serán revisados por pares de manera abierta, a menos que los autores no deseen hacerlo (véase el punto de evaluación anterior).
- Obras de ciencia ciudadana o participativa, de divulgación científica o

de presentación del conocimiento y el patrimonio local, cuyo objetivo es hacer accesible el conocimiento al mayor número posible de personas.

- Ensayos sobre ciencia y política científica (por ejemplo, en estudios sociales de la ciencia o en ética de la ciencia).
- Antologías de textos ya publicados pero no accesibles en la red, en un idioma distinto del francés, o que no son de libre acceso, pero de probado interés científico, intelectual o patrimonial.
- Libros de texto o libros educativos para niños.

Para el acceso libre y universal, a través de medios digitales, a los libros científicos publicados por autores de países del Sur y del Norte

Para más información: escriba a info@editionscienceetbiencommun.org

Les Éditions science et bien commun sont une branche de l'Association science et bien commun (ASBC), un organisme sans but lucratif enregistré au Québec depuis juillet 2011.

L'Association science et bien commun

L'ASBC a comme mission de stimuler la vigilance et l'action pour une science ouverte au service du bien commun. À cette fin, elle s'emploie à :

- Défendre et promouvoir une vision des sciences au service du bien commun;
- Colliger, analyser, produire et diffuser de l'information sur la science et sur ses rapports avec la société;
- Soutenir, promouvoir ou organiser des expériences de démocratisation des sciences;
- Organiser des expériences de débat public sur diverses facettes des sciences;
- Mettre en place des expériences de rencontre entre le monde scientifique et d'autres sphères sociales (ex. le milieu artistique, le milieu politique, etc.);
- Offrir un service d'orientation des groupes de la société civile dans le

monde universitaire;

- Offrir, sous réserve de la Loi sur l'enseignement privé (L. R. Q., c. E-9.1) et de ses règlements, des formations sur la responsabilité sociale, la science avec les citoyens et l'éthique des sciences.

Sur son site Web se trouvent de nombreuses informations sur ses activités et ses publications. Il est possible de devenir membre de l'Association science et bien commun en payant un tarif modeste.

Pour plus d'information, écrire à [info @ scienceetbiencommun.org](mailto:info@scienceetbiencommun.org), s'abonner à son compte Twitter @ScienceBienComm ou à sa page Facebook : <https://www.facebook.com/scienceetbiencommun>

Les Éditions science et bien commun

Un projet éditorial novateur dont les principales valeurs sont :

- la publication numérique en libre accès, en plus des autres formats
- la pluridisciplinarité, dans la mesure du possible
- le plurilinguisme qui encourage à publier en plusieurs langues, notamment dans des langues nationales africaines ou en créole, en plus du français
- l'internationalisation, qui conduit à vouloir rassembler des auteurs et auteures de différents pays ou à écrire en ayant à l'esprit un public issu de différents pays, de différentes cultures
- mais surtout la justice cognitive :
 - chaque livre collectif, même s'il s'agit des actes d'un colloque, devrait aspirer à la parité entre femmes et hommes, entre juniors et seniors, entre auteurs et auteures issues du Nord et issues du Sud (des Suds); en tout cas, tous les livres devront éviter un déséquilibre flagrant entre ces points de vue;
 - chaque livre, même rédigé par une seule personne, devrait s'efforcer d'inclure des références à la fois aux pays du Nord et aux pays des Suds, dans ses thèmes ou dans sa bibliographie;
 - chaque livre devrait viser l'accessibilité et la « lisibilité », réduisant au maximum le jargon, même s'il est à vocation scientifique et évalué par les pairs.

Le catalogue

Le catalogue des Éditions science et bien commun (ESBC) est composé de livres qui respectent les valeurs et principes des ÉSBC énoncés ci-dessus :

- Des ouvrages scientifiques (livres collectifs de toutes sortes ou monographies) qui peuvent être des manuscrits inédits originaux, issus de thèses, de mémoires, de colloques, de séminaires ou de projets de recherche, des rééditions numériques ou des manuels universitaires. Les manuscrits inédits seront évalués par les pairs de manière ouverte, sauf si les auteurs ne le souhaitent pas (voir le point de l'évaluation ci-dessus).
- Des ouvrages de science citoyenne ou participative, de vulgarisation scientifique ou qui présentent des savoirs locaux et patrimoniaux, dont le but est de rendre des savoirs accessibles au plus grand nombre.
- Des essais portant sur les sciences et les politiques scientifiques (en études sociales des sciences ou en éthique des sciences, par exemple).
- Des anthologies de textes déjà publiés, mais non accessibles sur le web, dans une langue autre que le français ou qui ne sont pas en libre accès, mais d'un intérêt scientifique, intellectuel ou patrimonial démontré.
- Des manuels scolaires ou des livres éducatifs pour enfants

Pour l'accès libre et universel, par le biais du numérique, à des livres scientifiques publiés par des auteures et auteurs de pays des Suds et du Nord

Pour plus d'information : écrire
à info@editionscienceetbiencommun.org