

Guía de evaluación
para programas de
prevención de homicidios en
América Latina y el Caribe

Ignacio Cano
Emiliano Rojido
Doriam Borges



LABORATÓRIO DE
ANÁLISE DA VIOLÊNCIA

G uía de evaluación para programas de prevención de homicidios en América Latina y el Caribe

Ignacio Cano
Emiliano Rojido
Doriam Borges

Laboratório de Análise da Violência
Universidade do Estado do Rio de Janeiro



Apoya



Guía de evaluación para programas de prevención de homicidios en América Latina y el Caribe

Autores:

Ignacio Cano (ignaciocano62@gmail.com)

Emiliano Rojido (emilianorojido@gmail.com)

Doriam Borges (doriamb@gmail.com)

Laboratório de Análise da Violência
Universidade do Estado do Rio de Janeiro

Consultores (*):

Lucía Dammert (USACH-Chile)

Claudio Beato (UFMG, Minas Gerais, Brasil)

Raquel Willadino (Observatório de Favelas-Río de Janeiro)

David Ramírez (México Evalúa-México)

Apoya: Open Society Foundations

Producción editorial: Susana Aliano Casales

ISBN digital: 978-9974-8697-6-9

Montevideo, Uruguay

Junio de 2020

() Los consultores ayudaron a definir el contenido general y el índice de esta Guía. Lucía Dammert y Claudio Beato también ofrecieron su opinión técnica sobre la primera versión del texto. Los autores quieren agradecer la contribución significativa de los consultores y aclarar que ellos no son responsables por el texto publicado ni por sus eventuales errores u omisiones.*

CONTENIDO

Introducción	7
Capítulo 1. El fenómeno del homicidio.....	11
1.1 Definición.....	11
1.2 Especificidad del homicidio: concentración social y territorial..	12
1.3 Etiologías de homicidio.....	17
1.4 Datos sobre homicidio.....	20
Capítulo 2. Programas de prevención del homicidio.....	29
2.1 Introducción.....	29
2.2 Tipos de prevención.....	30
2.3 Tipos de programa.....	31
Capítulo 3. Por qué es necesario evaluar.....	47
3.1 Utilidades de la evaluación.....	47
3.2 Qué riesgos existen si no se evalúa.....	49
Capítulo 4. ¿Qué es una evaluación de programas?	53
4.1 Concepto: qué es una evaluación de programas	53
4.2 Tipos de evaluación: de implementación, de impacto y de costo-beneficio	54
4.3 Monitoreo y evaluación	60
Capítulo 5. Evaluabilidad	63
5.1 ¿Cuáles son los requisitos mínimos para evaluar?.....	63
5.2 Objetivos, indicadores y metas del programa.....	63
5.3 Diagnóstico y teoría de cambio	65
5.4 Temporalidad de los efectos y temporalidad de la evaluación ..	68
5.5 Otros elementos de la evaluabilidad	69
Capítulo 6. Diseños de evaluación	71
6.1 Diseños experimentales	73
6.2 Diseños cuasiexperimentales	78
6.3 Midiendo el rigor metodológico: la escala de Maryland.....	84
Capítulo 7. Desafíos éticos de una evaluación.....	89
7.1 Introducción	89
7.2 Imparcialidad.....	90
7.3 Consentimiento informado	91
7.4 Confidencialidad	91
7.5 No causar daño	92
7.6 Publicación.....	95
Capítulo 8. ¿Quién hace la evaluación?.....	99
8.1 Evaluación interna vs. externa.....	99
8.2 Elementos a ser tomados en cuenta a la hora de contratar una evaluación externa.....	102
8.3 Participación social en la evaluación	105
Capítulo 9. Resultados y usos de la evaluación.....	109
Capítulo 10. Costo y financiamiento de una evaluación	115
10.1 ¿Son caras las evaluaciones?	115
10.2 ¿Quién debe financiar las evaluaciones?.....	118

Capítulo 11. Dificultades para evaluar programas de prevención de homicidios	121
11.1 Ausencia o fragilidad de la lógica del modelo o teoría de cambio	121
11.2 Pequeño número de homicidios en el momento inicial	122
11.3 Desplazamiento criminal	123
11.4 Regresión a la media.....	125
11.5 Cambios poblacionales.....	126
11.6 Alteraciones en las categorías o en los sistemas de registro de homicidios.....	126
11.7 Cambios que afectan a todo el territorio e inviabilizan a un grupo de control	127
11.8 Intervenciones con múltiples estrategias simultáneas	127
11.9 Distinta temporalidad de los efectos	129
Capítulo 12. Elementos que condicionan el diseño de la evaluación ..	131
12.1 Diseños evaluativos con grupos de control	133
12.2 Diseños evaluativos sin grupo de control	134
Capítulo 13. Paso a paso de una evaluación de prevención de homicidios	137
Capítulo 14. Ejemplos de evaluaciones de programas de reducción de homicidios	139
14.1 Impacto en los homicidios de la creación del Metrocable en Medellín (2004). Evaluación realizada por Cerdá, Morenoff, Hansen, Hicks, Duque, Restrepo y Diez-Roux.....	139
14.2 Impacto en los homicidios de la Ley ‘Maria da Penha’ para inhibir la violencia contra las mujeres en Brasil (2006). Evaluación realizada por Cerqueira, Matos, Martins y Junior	142
14.3 Impacto de la prohibición temporal de porte de armas en Cali y Bogotá (1993-1997). Evaluación realizada por Villaveces, Cummings, Espitia, Koepsell, McKnight y Kellermann	145
14.4 Impacto de un nuevo procedimiento de investigación de homicidios introducido en Bogotá (2016). Evaluación realizada por Collazos, Fergusson, La Rota, Mejía y Ortega	149
Referencias bibliográficas	155

Esquemas, gráficos y tablas

Esquema 1. Clasificación del homicidio intencional.....	18
Esquema 2. Indicadores relacionados a las fases del sistema de justicia criminal.....	25
Esquema 3. Programas de prevención de homicidios en América Latina y el Caribe según estrategia y tipo de intervención	45
Esquema 4. Evaluación como instrumento de gestión.....	47
Esquema 5. Elementos que condicionan el diseño de la evaluación....	136

Gráfico 1. Tasas de homicidio por cada 100.000 habitantes, según región. Año 2017 o último disponible	13
---	----

Gráfico 2. Tasa de homicidio cada 100.000 habitantes por país. Año 2017 o último disponible	13
Gráfico 3. Distribución espacial de los homicidios en Ciudad de México (2009-2016)	14
Gráfico 4. Tasa global de homicidio, por sexo y tramo de edad. Año 2017 o último disponible	16
Gráfico 5. Homicidios según sexo y relación entre víctima y victimario.....	19
Gráfico 6. Homicidios de mujeres según relación entre víctima y victimario.....	19
Gráfico 7. Ejemplos de posibles efectos reflejados en una serie temporal.....	81
Gráfico 8. Ejemplo hipotético de discontinuidad en la regresión	84
Tabla 1. Tipos de evaluación.....	59
Tabla 2. Virtudes y riesgos de las evaluaciones externas.....	102

Cuadros de programas de prevención mencionados

El femicidio/feminicidio.....	18
Programa Nacional de Entrega Voluntaria de Armas de Fuego (PEVAF), Argentina.....	33
La “Lei Seca” en Diadema, Brasil	33
Plan de Recuperación y Mantenimiento de Espacios de Encuentro para el Fortalecimiento y Campaña de Sensibilización para la Valoración de la Vida, Venezuela.....	34
Campaña de Sensibilización para la Valoración de la Vida, Venezuela	35
Centros Emergencia Mujer (CEM), Perú	36
Fica Vivo!, Brasil.....	36
Plan Nacional de Vigilancia por Cuadrantes, Colombia	37
Premio Paraíba Unida por la Paz, Brasil.....	38
Paz y Justicia, Honduras	38
Índice de Aptitud para el Uso de la Fuerza Policial, Brasil	39
Proyecto Alcatraz, Venezuela.....	40
Peace Management Initiative, Jamaica.....	41
La Tregua entre Maras, El Salvador.....	42
Citizen Security Programme, Trinidad y Tobago	43
Estrategia Nacional contra el Homicidio, Colombia	44
Los efectos de los programas de prevención de violencia doméstica en el asesinato de parejas íntimas en 48 ciudades de los Estados Unidos	50
La “Donbernabilidad” en Medellín.....	56
El experimento de patrullaje de Kansas City (1972-1973)	75
Todos Somos Juárez, en Ciudad Juárez, México.....	128

INTRODUCCIÓN

Contexto de los homicidios en América Latina y Caribe

América Latina y el Caribe gozan del triste privilegio de ser la región con la mayor incidencia de homicidios en el mundo. Esta región no solo tiene la incidencia más elevada, sino también una tendencia creciente desde el año 2000, que contrasta fuertemente con la tendencia de disminución en todos los otros continentes.

Los efectos de la violencia letal son generalizados en todas las sociedades de la región. Además de la pérdida de las vidas y de la renta que las víctimas, en mayoría jóvenes, hubieran generado a lo largo de su vida, el conjunto de la sociedad siente temor y se retira de los espacios públicos, minando la confianza en las instituciones y debilitando la cohesión social.

En consecuencia, el costo económico y social de la violencia letal en América Latina y Caribe es extraordinario. Por ello, es urgente revertir este escenario y, para lograrlo, es preciso contar con programas de prevención y de reducción de homicidios basados en evidencias.

¿Por qué esta Guía es importante?

A pesar de la gravedad del problema, las intervenciones enfocadas en reducir homicidios son relativamente escasas en América Latina y el Caribe. Son mucho más comunes los programas generales de prevención de la violencia, el crimen y la inseguridad, en los que el homicidio es tratado transversalmente como una manifestación extrema de esos fenómenos. Es frecuente que la tasa de homicidio se utilice como una métrica de la violencia en general o como un criterio para seleccionar los territorios donde se implementan los programas, pero eso no significa que estos tengan el objetivo de reducir la violencia letal.

Los programas contra los homicidios no solo son escasos, sino que no suelen estar sometidos a evaluaciones rigurosas. El [Mapeo de programas de prevención homicidios en América Latina y el Caribe](#) (Cano y Rojido, 2016) reveló que menos del 20% de las iniciativas poseían una evaluación de impacto. Lo que en algunos casos se presentaban como evaluaciones no pasaban de auditorías o informes de implementación, sin ninguna estrategia para medir impactos.

En otras ocasiones, las evidencias que apuntan a la efectividad de un programa son originarias de otros países, generalmente de América del Norte y de Europa, y de contextos muy diferentes a los enfrentados en América Latina y el Caribe. Por ello, no hay garantía de que lo que

funcionó en Copenhague o en Chicago lo haga de la misma forma en Medellín o en San Paulo.

El resultado obvio de todo esto es que no sabemos en qué medida los programas funcionarán en la región y, por tanto, es muy difícil adoptar intervenciones eficaces y eficientes que consigan disminuir de forma significativa los niveles de homicidio. Esta guía pretende contribuir a cambiar este escenario.

Objetivos

La *Guía de evaluación para programas de prevención de homicidios en América Latina y el Caribe* tiene el propósito central de incentivar la realización de evaluaciones de calidad de los programas de prevención de homicidios, de forma que esto pueda contribuir a iniciativas más exitosas para que la región deje de ser un día un *hotspot* gigante de homicidios en el mundo. En ese sentido, es necesario, en primer lugar, sensibilizar a la sociedad sobre la importancia de la evaluación en este campo. En el futuro, sería conveniente promover una cultura de evaluación en esta área, de forma que todos los proyectos sean adecuadamente evaluados y que nuevas iniciativas se basen, siempre que sea posible, en estas evidencias. La creación de esta cultura evaluadora implica convencer a todos los actores sociales de la utilidad de la evaluación para que los programas sean más racionales y efectivos. Después de todo, la evaluación es una herramienta de aprendizaje, que proporciona conocimiento para poder mejorar el desempeño.

Dentro del objetivo general de promover evaluaciones metodológicamente rigurosas, podemos citar algunos objetivos específicos.

Entre ellos:

- Sensibilizar a todos los sectores sobre el papel de la evaluación.
- Fomentar la evaluabilidad de los programas.
- Capacitar, en alguna medida, a los gestores y las gestoras para pensar las evaluaciones.
- Fortalecer la capacidad de contratar evaluadores y evaluadoras competentes.
- Ayudar a entender los informes de evaluación.
- Fomentar un uso adecuado de las evaluaciones.

Sin ser un manual técnico de evaluación ni un texto académico, la *Guía* aspira a facilitar la comprensión de los principales aspectos que

componen la evaluación de programas de prevención al homicidio y a orientar a todos los actores relevantes (gestores y gestoras, sociedad civil, equipos técnicos, agencias financiadoras) sobre qué medidas deben tomarse durante y, preferencialmente, antes de comenzar los proyectos para garantizar una evaluación cualificada y transparente.

Aunque serán mencionados diversos tipos de evaluaciones, el énfasis será dado a las evaluaciones de impacto, que pueden determinar si las acciones adoptadas tuvieron o no el efecto de disminuir los homicidios y, en caso afirmativo, en qué medida. El diferencial de esta *Guía*, en relación con otras existentes, radica en el foco temático específico sobre los homicidios, subrayando los aspectos propios de estas intervenciones y las dificultades que enfrenta su evaluación. Desde el punto de vista geográfico, el foco será, como ya fue mencionado, América Latina y el Caribe.¹

Nuestra intención fue crear un documento versátil que cumpliera, entre otras, las siguientes condiciones:

- Ser de fácil manejo.
- Permitir que el lector o la lectora conozca diversas estrategias para evaluar programas de prevención al homicidio y que sea consciente de los desafíos metodológicos y éticos que involucran.
- Ayudar al lector o la lectora a pensar procesos de evaluación cuyos costes sean razonables para las diferentes instituciones y organizaciones.

Por otro lado, vale la pena recordar que este texto no discute la etiología de la violencia letal ni está pensado para ayudar a formular programas de reducción de homicidios, pues se dedica específicamente a responder a la cuestión de cómo evaluarlos.

¿A quién se dirige esta *Guía*?

El texto está pensado para cuatro tipos de público. El primero son los formuladores y ejecutores de proyectos de prevención de homicidios, que deben tener en cuenta la evaluación desde el inicio del diseño de los programas. Aunque solemos pensar en intervenciones gubernamentales contra la violencia, la *Guía* está elaborada tanto para programas públicos como para iniciativas desarrolladas desde la sociedad civil, incluso para aquellas que son locales y cuentan con presupuestos limitados.

De forma general, una evaluación será mejor cuanto antes sea formulada y las evaluaciones que deben ser pensadas una vez que el proyecto ya comenzó están casi siempre condenadas a tener una validez

¹ Algunas guías de evaluación están dirigidas a proyectos para prevenir la violencia en general, pero no se centran en el tema de los homicidios, como por ejemplo el *Manual de Avaliação de Programas de Prevenção a Violência*, elaborado por Siveira y Peixoto (2010). Otros ejemplos de guías de evaluación no son específicos del contexto latinoamericano, como el *Evaluation Handbook: Guidance for designing independent evaluation at UNODC*, elaborado y editado en el marco de la Oficina de las Naciones Unidas contra la Droga y el Delito (2017).

menor y tienen muchas menos opciones metodológicas disponibles. Por eso, es esencial que quien vaya a diseñar un proyecto para reducir homicidios tenga en cuenta una serie de elementos que serán explicados aquí.

La segunda audiencia de esta *Guía* son las personas en cargos públicos y electivos que, sin ser quienes dirigen personalmente los proyectos de prevención, deben tomar decisiones sobre aprobación de proyectos y distribución de recursos. Decisiones para las cuales las evaluaciones son de extrema importancia, en un doble sentido. Por un lado, los resultados de evaluaciones anteriores en este campo deben orientar estas decisiones. Por otro, cualquier decisión sobre proyectos a ser implementados en el futuro debe tomar en cuenta el componente evaluativo.

El tercer público relevante son las entidades financiadoras de programas de prevención de homicidios, que no son necesariamente cargos públicos, como por ejemplo, los y las integrantes de fundaciones privadas o de instituciones internacionales. Ellas tienen una gran responsabilidad en exigir que las intervenciones sean siempre evaluadas de forma competente y que las evaluaciones sean publicadas y utilizadas.

Por último, esta *Guía* está dirigida también a la sociedad civil en general, cuya exigencia y presión es esencial para que sean adoptados programas que reduzcan la violencia letal y que consigan que América Latina y el Caribe dejen de ser universalmente reconocidos como lugares de extrema violencia. En estas exigencias, la sociedad civil debe incluir que los programas sean basados en evidencias, que sean adecuadamente evaluados y que el resultado de estas evaluaciones sea ampliamente difundido y debidamente aprovechado para decisiones futuras.

CAPÍTULO 1. EL FENÓMENO DEL HOMICIDIO

1.1 Definición

Por atentar contra el derecho más básico a la vida de los seres humanos, el homicidio es considerado uno de los delitos más graves, a lo largo y ancho del planeta. Esta característica y ciertos atributos metodológicos² hacen de la tasa de homicidios el indicador más robusto de la violencia que ocurre en las sociedades. Con todo, a pesar de ser repudiado de manera extendida, no existe un concepto único o universal de homicidio. Los países establecen definiciones y categorías a partir de sus propias legislaciones nacionales, de forma que algunas muertes se consideran homicidio en algunos lugares, pero no en otros (Smit, de Jong & Bijleveld, 2012).

Para facilitar la comparabilidad de este fenómeno las agencias internacionales promueven definiciones que buscan ir más allá de las legislaciones nacionales. Sin embargo, tampoco hay un concepto consensuado en este ámbito.

La Oficina de las Naciones Unidas para las Drogas y el Crimen (UNODC), a través de su Clasificación Internacional de Delitos con Fines Estadísticos (ICCS), define el homicidio como la “muerte ilegal infligida contra una persona con la intención de causar la muerte o lesiones graves” (UNODC, 2015).³

Este concepto supone tres elementos constitutivos:

- 1) la muerte de una persona por otra persona (componente objetivo);
- 2) la intención de matar o lesionar seriamente a la víctima (componente subjetivo);
- 3) que la muerte es contraria a la ley (componente legal).

A partir de lo expuesto quedan excluidas las muertes producidas por legítima defensa o por la actuación de las fuerzas de seguridad, las que, en principio, son legales y no suponen la total responsabilidad del autor. Las muertes en el marco de guerras también son distinguidas de los homicidios, dado que en estos casos el agresor está autorizado por el Estado para infligir un trauma letal.⁴

La Organización Mundial de la Salud (OMS), por su parte, define el homicidio a partir de la Clasificación Internacional de Enfermedades (CIE) como la “muerte debida a lesiones infligidas por otra persona con la intención de lesionar o matar, por cualquier medio” (OMS, 2009).⁵ En este caso, no es importante la legalidad del hecho para considerarlo o no

2 [Ver apartado 1.4.](#)

3 La UNODC recopila y publica periódicamente los datos de homicidio en su sitio web: <https://dataunodc.un.org/>.

4 Una discusión sobre los conflictos armados y los registros de homicidio puede encontrarse en UNODC, 2015: 18.

5 La OMS publica datos de homicidio en su Observatorio Global de la Salud. Disponible en: <http://apps.who.int/gho/data/node.imr> y http://apps.who.int/healthinfo/statistics/mortality/causeofdeath_query/.

un homicidio, aunque también se contabilizan por separado las muertes en guerras⁶ y por intervenciones de los cuerpos de seguridad pública,⁷ dado su carácter episódico y etiología particular.

Una definición más amplia es proporcionada por el Protocolo de Bogotá, que entiende el homicidio como “la muerte de una persona provocada por la agresión intencional de otra(s)”.⁸ En este caso, no importa el carácter legal de la muerte y quedarían incluidas tanto las muertes en manos de las fuerzas de seguridad como las ocurridas en el contexto de guerras, guerras civiles y conflictos internos.

La diferencia en la cifra de homicidios que se obtiene a partir de la utilización de los conceptos mencionados puede ser importante, al menos en un contexto de alta incidencia y fragilidad institucional como el latinoamericano (Ribeiro, Borges y Cano, 2015). En particular, la exclusión de las muertes producidas por intervenciones de las fuerzas de seguridad, frecuentes en la región, puede llevar a subestimar el nivel de violencia letal en algunos países y a comprometer la comparabilidad internacional de los datos de manera general.

1.2 Especificidad del homicidio: concentración social y territorial

El fenómeno del homicidio presenta regularidades relativamente universales, entre las que se destacan su alta concentración geográfica y social. Mientras que el primer patrón se refiere al modo en que los hechos suelen distribuirse en el espacio, el segundo informa sobre las características más frecuentes de las personas involucradas, sea en el papel de víctimas o de victimarios.

1.2.1 Concentración geográfica

Una de las principales características del homicidio es su concentración geográfica. La tasa global es 6,1 homicidios por cada 100.000 habitantes, pero esa cifra esconde una importante variación en los niveles de violencia. De hecho, mientras que casi el 50% de la población mundial vive en sociedades de baja incidencia, responsables de alrededor del 10% de los homicidios, el 50% de los homicidios ocurre en sociedades que abarcan poco más del 10% de la población del planeta (Eisner & Nivette, 2012).

Es posible observar estas variaciones analizando diferencias entre las regiones. El continente con mayor tasa de homicidios por cada 100.000 habitantes es América (17,2); seguido de África (13,0), Europa (3,0), Oceanía (2,8) y Asia (2,3). De este modo, mientras que en América y en África la incidencia de homicidios más que duplica la tasa global, en Europa, Oceanía y Asia es inferior a la mitad. En términos de distribución

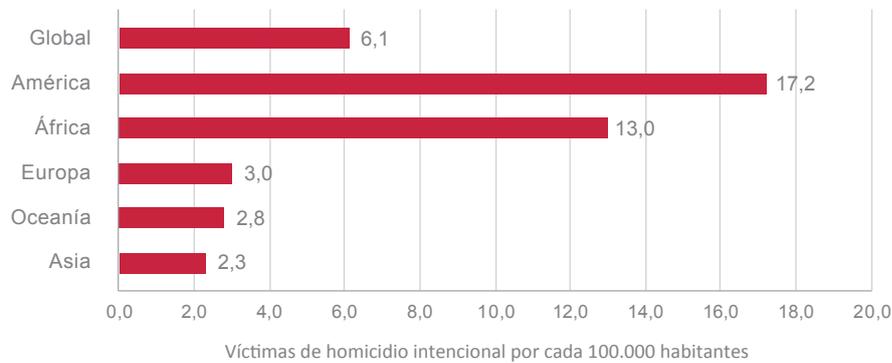
6 Lesiones a militares o civiles causadas por la guerra o la insurrección civil, incluidas aquellas que ocurren durante el tiempo de guerra o insurrección y después del cese de las hostilidades.

7 Lesiones infligidas por la policía u otros agentes de la ley, incluido el personal militar en servicio, durante el arresto o intento de arrestar a los infractores de la ley, suprimir disturbios, mantener el orden y otras acciones legales. Excluye lesiones causadas por insurrecciones civiles.

8 Disponible en: <http://homicidioslatam.org/wp-content/uploads/2020/05/Protocolo-de-Bogota_esp.pdf>.

porcentual, el continente americano concentra 37,4% de los homicidios con apenas el 8% de la población mundial, lo que proporciona una imagen nítida de la concentración regional de la violencia (UNODC, 2019).

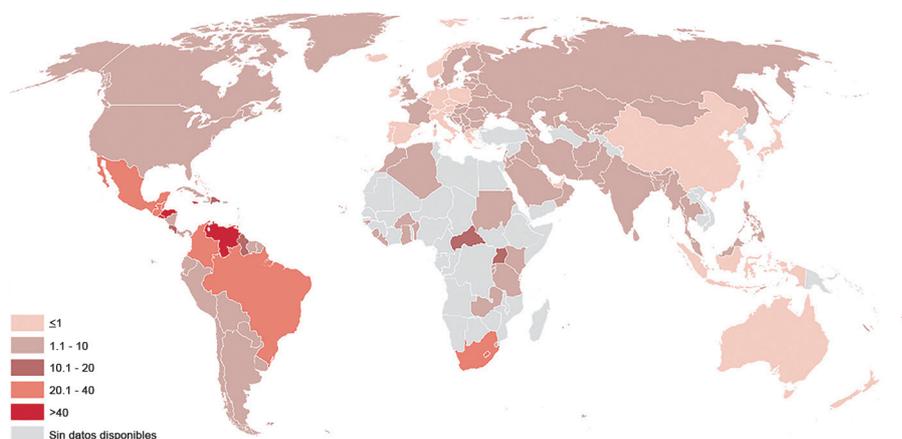
Gráfico 1. Tasas de homicidio por cada 100.000 habitantes, según región. Año 2017 o último disponible



Fuente: UNODC, *Global Study on Homicide: Executive Summary*, 2019: 11.

Variaciones similares se constatan en unidades de menor agregación. América, por ejemplo, se divide en cuatro subregiones con importantes disparidades: Norteamérica (5,1), Centroamérica (25,9), Sudamérica (24,2) y el Caribe (15,1). Mientras que Centroamérica es la subregión más violenta del planeta, la incidencia en Norteamérica se ubica por debajo del promedio global. También al observar una subregión pueden identificarse contrastes en su interior. En Sudamérica, por ejemplo, las tasas del Cono Sur (y principalmente en Chile) se aproximan a los valores de Europa, mientras que las del norte (Venezuela y Colombia) se asemejan a los altos valores de Centroamérica (UNODC, 2019).

Gráfico 2. Tasa de homicidio cada 100.000 habitantes por país. Año 2017 o último disponible

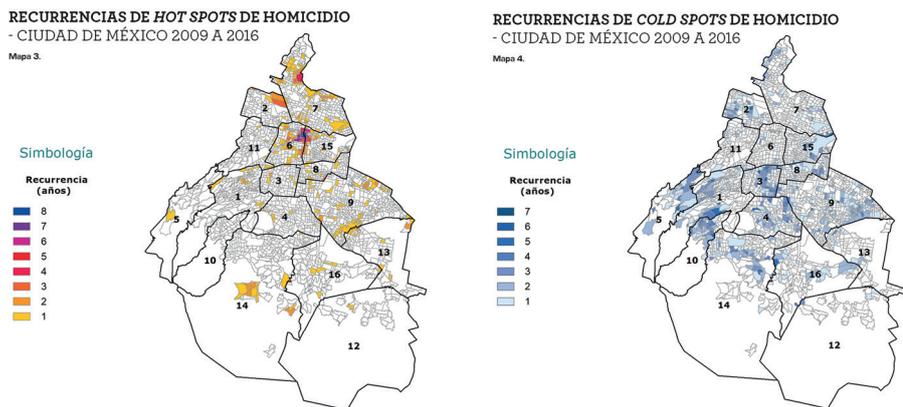


Fuente: UNODC, *Global Study on Homicide: Executive Summary*, 2019: 14.

Esta concentración geográfica de los homicidios tiende a repetirse al comparar divisiones subnacionales, las ciudades de un país, los barrios de una ciudad y las calles dentro de un barrio. Por ejemplo, un estudio sobre la violencia armada en Boston entre 1980 y 2008 descubrió que menos del 3% de las cuadras y esquinas de la ciudad dieron lugar a más de la mitad de los incidentes registrados. En contrapartida, el 89% de los segmentos de calle e intersecciones nunca habían experimentado un solo incidente durante el período analizado (Braga, Papachristos & Hureau, 2010).

Un estudio sobre la distribución espacial de los homicidios en la ciudad de México entre 2009 y 2016, también permite visualizar con claridad la existencia de microáreas que concentran un gran volumen de homicidios (*hotspots* o puntos calientes), así como de otras que, por el contrario, parecerían tener inmunidad ante este problema social (*coldspots* o puntos fríos).

Gráfico 3. Distribución espacial de los homicidios en Ciudad de México (2009-2016)



Fuente: México Evalúa, 2018: 57 y 61.⁹

Lo anterior sugiere que probablemente hay características espaciales, inherentes a las localidades, que expliquen la distribución de la letalidad violenta. En este sentido, es importante considerar que el fenómeno de los homicidios puede ser el resultado de una compleja interacción de factores individuales, relacionales, sociales, culturales y ambientales. Dado que los patrones espaciales del homicidio son relativamente consistentes,¹⁰ el análisis de *hotspots* y *coldspots* se ha tornado un insumo importante para gestionar los recursos destinados a la prevención. Desde el punto de vista de la intervención, focalizar los esfuerzos para reducir la violencia letal en las áreas donde se concentra parece ser una estrategia apropiada incluso si se pretende lograr un impacto más generalizado.

9 Una herramienta interactiva sobre la distribución espacial de los homicidios en la ciudad de México puede consultarse en: <<https://www.mexicoevalua.org/homicidioscdmx/>>.

10 Existen, además, regularidades temporales moderadas. Es frecuente, por ejemplo, que los homicidios aumenten los fines de semana, en relación con los días laborables, y en horas de la noche, en relación con las del día (Ceccato, 2005). Esto se vincula a distintos aspectos, como la mayor exposición a interacciones violentas en los momentos de ocio, el incremento en el consumo de alcohol y sustancias psicoactivas, el menor control social formal (por ejemplo, de policías) e informal (por ejemplo, de padres o familiares), etc.

Las tasas agregadas tienen una capacidad informativa limitada, no solo porque los homicidios no siguen una distribución aleatoria, sino porque los factores de protección y riesgo se combinan en grado variable en el territorio. No importa cuán pequeña sea la unidad de agregación, el riesgo probablemente se distribuirá de forma desigual entre sus habitantes. Además, cuando las unidades agregadas son muy pequeñas, el fenómeno se torna muy infrecuente —hay muchos ceros— y se hace difícil de analizar. Por ese motivo, la georreferenciación de los homicidios¹¹ suele ser importante como primer paso para comprender la dinámica local de la violencia y para abordarla en su especificidad allí donde se concentra.

1.2.2 Concentración social

Los homicidios no solo tienen regularidades según los lugares donde ocurren, sino también en función del perfil de las personas que involucran, en general, pertenecientes a sectores sociales con baja escolaridad, dificultades de inserción en el mercado de trabajo y bajo nivel de ingresos. En algunos países también existe un sesgo racial en el perfil de estas personas, lo que generalmente va acompañado de una acumulación de desventajas como las antedichas. En Brasil, por ejemplo, el 75,5% de las víctimas de homicidio en 2017 fueron negras, lo que supone una tasa para las personas negras de 43,1 por cada 100.000 habitantes, en comparación con 16 para las no negras (IPEA & FBSP, 2019).

Una de las razones de la poca prioridad otorgada al problema del homicidio en América Latina y el Caribe tiene que ver justamente con la baja visibilidad social de las víctimas y su escasa capacidad de estimular reacciones por parte del Estado. Por el contrario, las políticas en el área de seguridad han estado tradicionalmente dominadas por la lucha contra crímenes de menor gravedad, como los robos, los secuestros o el tráfico de drogas, crímenes que tienden a afectar a víctimas con un perfil socioeconómico más elevado o a encajar en agendas sociopolíticas de mayor repercusión (Cano y Rojido, 2017).

De modo relativamente uniforme a lo largo del tiempo y en las distintas culturas, es más probable la participación de hombres que de mujeres en los homicidios.¹² De hecho, los hombres constituyen una amplia mayoría tanto entre los perpetradores (90%) como entre las víctimas (81%), siendo que aproximadamente dos tercios de todos los homicidios del mundo consisten en hombres que matan a otros hombres. En términos de tasas de victimización, los hombres poseen un promedio casi 4 veces superior al de las mujeres: mientras que la tasa para los hombres es de 9,9, para las mujeres es 2,3, una cifra evidentemente

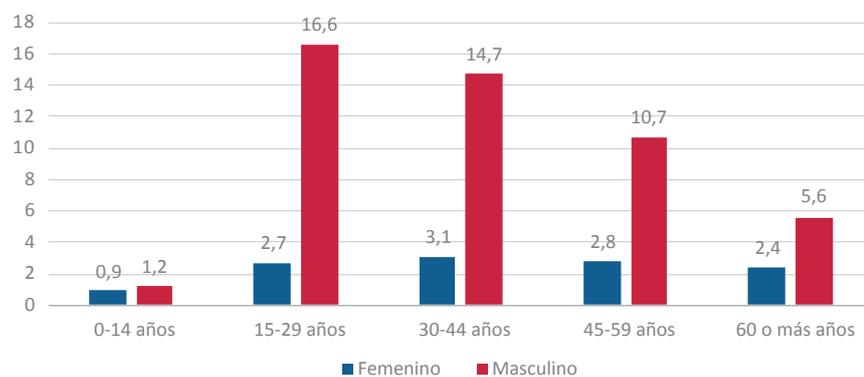
11 El geoprocesamiento denota el uso de técnicas matemáticas y computacionales para el tratamiento de información geográfica. Las herramientas computacionales, llamadas Sistemas de Información Geográfica (SIG), realizan análisis complejos integrando datos de diferentes fuentes y creando bases de datos georreferenciadas que hacen posible la producción automática de documentos cartográficos.

12 Estudios sobre la distribución de los homicidios por sexo revelan que hay menos diferencias entre las víctimas, donde los hombres representan entre el 59% y 78% de los casos, que entre los perpetradores, donde pueden oscilar entre el 82% y el 95% (Eckhardt & Pridemore, 2009). Aunque la mayoría de los estudios históricos no permiten una desagregación por sexo, datos de Europa a partir del siglo XIII señalan que la proporción de mujeres que cometieron homicidio difícilmente sobrepasaba el 15% y es frecuente el rango del 5% a 12%. En cuanto a la victimización de mujeres, las investigaciones disponibles sugieren que se ubicó en el orden del 0% al 20% del total de casos, con una media aproximada del 7% (Eisner, 2003).

inferior a la media global (6,1) (UNODC, 2019). Esta brecha de sexo tiende a ser mayor en sociedades con alta incidencia de homicidios, donde la proporción de víctimas masculinas es superior.

La edad de los individuos es otro de los aspectos a destacar. El involucramiento de las personas en actividades violentas no es constante a lo largo de la vida, sino que, mayoritariamente, tiene un incremento a partir de la adolescencia y decae progresivamente a medida que se ingresa en la vida adulta. El riesgo de perpetración y victimización por homicidio es mayor para los jóvenes y es en el tramo de 15 a 29 años de edad cuando se alcanza la mayor vulnerabilidad. De hecho, casi la mitad de las víctimas de homicidio en el mundo pertenecen a ese grupo etario (UNODC, 2019).

Gráfico 4. Tasa global de homicidio, por sexo y tramo de edad. Año 2017 o último disponible



Fuente: [UNODC \(2019\) - Database](#).

La concentración social por sexo y edad es más marcada en América que en las otras regiones, ya que las víctimas masculinas y jóvenes están sobrerrepresentadas. De este modo, si la tasa global es más de cuatro veces superior para los hombres (9,9) que para las mujeres (2,3), en América esta proporción es aproximadamente ocho veces mayor para los primeros (31,2) que para las segundas (3,6). Algo similar sucede con los homicidios de personas entre 15 y 29 años, que en América es casi cuatro veces superior a la tasa global para esas mismas edades. Por ejemplo, la tasa de homicidio para hombres jóvenes en países como Honduras, El Salvador o Jamaica excede los 100 cada 100.000 habitantes (UNODC, 2019). Esta singularidad se debe a un tipo de homicidio preponderante en la región, asociado a la presencia de grupos juveniles (*gangs*, *bandas*, *padillas*, etc.) y organizaciones criminales en disputa.

El realce de este perfil es, por un lado, una tragedia para dichos grupos poblacionales, pero representa, al mismo tiempo, una oportunidad para la adopción de políticas focalizadas de prevención.

1.3 Etiologías de homicidio

El homicidio es habitualmente tratado como un fenómeno unitario y homogéneo, pero en verdad abarca un conjunto diverso de manifestaciones extremas de violencia. Estos crímenes pueden suceder bajo circunstancias variables, como relaciones violentas de pareja, altercados entre vecinos, el robo de un transeúnte, un ataque sexual, el enfrentamiento entre grupos criminales, una represión policial excesiva, un ataque terrorista, manifestaciones de odio hacia ciertos grupos poblacionales, una contienda política, etc. Aunque todos pueden ser entendidos como homicidios, estos hechos tienen una naturaleza claramente diferenciada.

Cada vez son más los investigadores y las investigadoras que defienden la necesidad de agrupar los homicidios en categorías relevantes para mejorar el análisis y la prevención (Ioannou & Hammond, 2015). En algunos casos, explícitamente se indica que la tasa de homicidios no sería más que un promedio de fenómenos heterogéneos, cada uno con sus causas, tendencias y poblaciones afectadas.

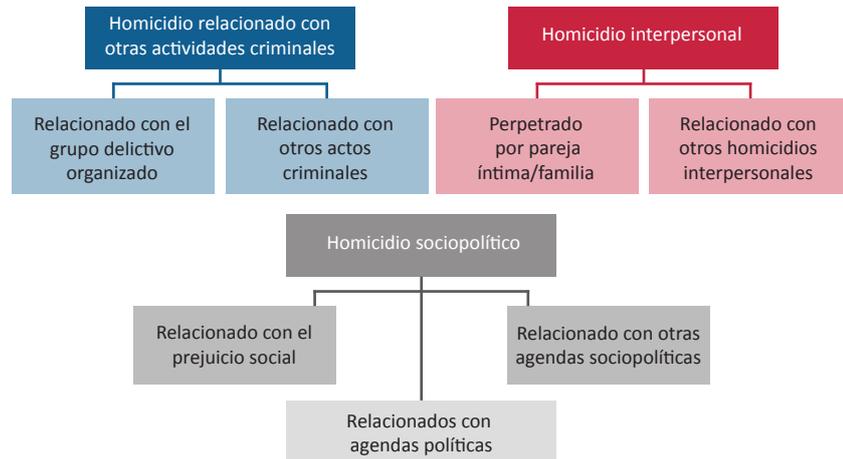
La literatura especializada presenta tipologías de homicidio en función de diferentes criterios, como los grupos poblacionales (hombres y mujeres, blancos y negros, jóvenes y adultos), la cantidad de víctimas u ofensores (simples y múltiples), la relación entre víctima y victimario (extraños, conocidos, parejas íntimas y familiares), la motivación de la ofensa (instrumental y expresiva),¹³ el mecanismo utilizado (armas de fuego, cuchillos, etc.) y el lugar del hecho (ambiente doméstico, laboral, etc.) (Kivivuori, Suonpää & Lehti, 2014). En última instancia, todas estas clasificaciones buscan regularidades que permitan desagregar los homicidios de acuerdo a su etiología particular.

La UNODC propone una tipología basada en una combinación de elementos como las circunstancias, las motivaciones y las relaciones entre la víctima y el victimario. Se identifican así tres tipos principales de homicidio, cada uno de los cuales es dividido en subcategorías: el homicidio relacionado a actividades criminales, el homicidio interpersonal y el homicidio sociopolítico. El *homicidio relacionado con actividades criminales* es el que se perpetra en vinculación directa o indirecta con otro hecho ilícito y puede estar vinculado al crimen organizado o a actividades delictivas comunes, como los robos. El *homicidio interpersonal* es menos instrumental que el anterior y sucede a consecuencia de conflictos entre las personas. Dependiendo

13 La violencia expresiva se asocia a confrontaciones personales en las que descargar ira o frustración y agredir a otra persona es un objetivo en sí mismo del perpetrador. En la violencia instrumental, por su parte, la agresión constituye un medio para lograr otro objetivo, como dinero.

de la relación entre la víctima y el perpetrador, estos se dividen en homicidios de pareja íntima o familiares y homicidios perpetrados por otras personas. Por último, se halla el *homicidio sociopolítico*, que se caracteriza por suceder en el curso de actividades realizadas para promover reivindicaciones sociales y políticas.

Esquema 1. Clasificación del homicidio intencional



Fuente: UNODC (2019): Global Study on Homicide: Understanding Homicide. Vienna.

Aunque los datos desagregados por tipo son insuficientes en cantidad y calidad, lo que dificulta la realización de análisis rigurosos, sugieren que existe una correlación entre la incidencia general de homicidios y los tipos predominantes en las sociedades. De este modo, en los contextos relativamente violentos el homicidio tendería a asumir un carácter más instrumental y asociativo que en los contextos pacíficos, donde sería más expresivo e individual (Eisner, 2015).¹⁴

El femicidio/feminicidio¹⁵

Dado que la mayoría de los homicidios son cometidos y sufridos por hombres, estos casos representan el principal sustrato del fenómeno, mientras que otras categorías minoritarias quedan subsumidas e invisibilizadas (Shilan, 2017).

El homicidio de mujeres tiene características distintivas en relación al de hombres, entre otras cosas, porque las mujeres tienen una probabilidad mucho mayor de morir en su propio hogar y a manos de personas próximas a su círculo íntimo, como parejas o familiares. En este sentido, a pesar de representar una proporción pequeña del total de víctimas de homicidio, las mujeres constituyen una amplia mayoría entre las víctimas de homicidio por pareja íntima.

14 El homicidio relacionado al crimen organizado es un fenómeno destacado en las Américas, continente con alta incidencia, donde aproximadamente el 30% de los asesinatos son de este tipo; a diferencia de Europa, Asia y Oceanía, donde es inferior al 1%. En contraste, la proporción de homicidios totales cometidos por parejas íntimas es sensiblemente menor en América que en las otras regiones menos violentas (UNODC, 2014).

15 El concepto femicidio (*femicide* en inglés) surge a mediados de los años setenta en la academia inglesa, a partir de los trabajos de Diana Russell que hacen referencia al “asesinato de mujeres realizado por hombres, motivado por odio, desprecio, placer o un sentido de propiedad de la mujer” (ONU-UNETE, 2013: 15). Posteriormente, en los años noventa, Marcela Lagarde crea el neologismo *feminicidio* para dar cuenta de la violencia extrema que, bajo la complicidad del Estado, existía contra las mujeres en México y en varios países de América Central (Lagarde, 2008). En general, sin embargo, ambos términos son usados de forma indistinta en la literatura sobre el tema.

Gráfico 5. Homicidios según sexo y relación entre víctima y victimario

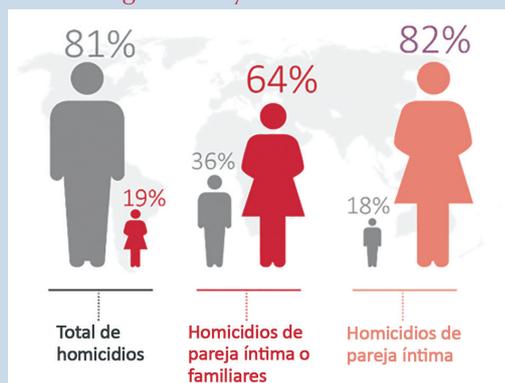
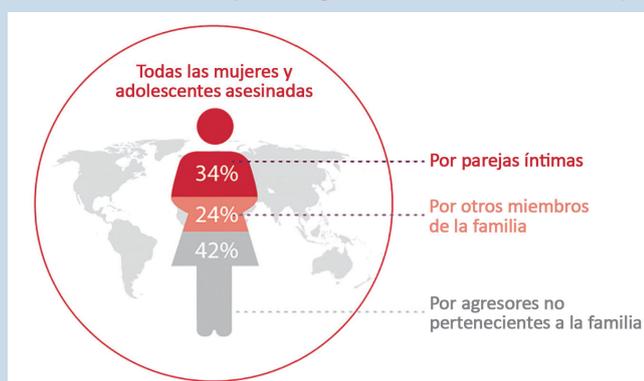


Gráfico 6. Homicidios de mujeres según relación entre víctima y victimario



Fuente: UNODC, 2019: 15 y 24.

La perspectiva feminista ha promovido el uso del término de femicidio/femicidio para dar cuenta del homicidio de mujeres basado en razones de género. Este fenómeno es entendido como la manifestación más extrema de la violencia de género, un recurso utilizado por los hombres para reproducir el orden social que los privilegia (Radford & Russell, 1992). Así, el femicidio/femicidio resignificó el homicidio de mujeres (en rigor, una parte de ellos) como violencia estructural, cuestionando la relevancia de los factores de riesgo individuales y de la neutralidad de género del “homicidio” y destacando el papel explicativo del patriarcado.

El femicidio/femicidio ha adquirido protagonismo social y político en los últimos años, aunque su definición todavía es objeto de debate entre los especialistas (Corradi *et al.* 2016). Adicionalmente, las limitaciones en la información disponible dificultan la operacionalización y medición del femicidio/femicidio, comprometiendo las posibilidades de dimensionar e intervenir sobre el problema (Cepal, 2015: 48).¹⁶

A pesar de ello, varios países de América Latina y el Caribe han incorporado el femicidio/femicidio como nuevo delito o como agravante del homicidio en sus códigos penales. En general, las modificaciones legislativas buscan otorgar visibilidad a la violencia letal contra las mujeres, postulando simultáneamente penas más severas para los agresores (Deus y González,

16 El Observatorio de Igualdad de Género de la Cepal es el único portal web que divulga información específica sobre la incidencia del femicidio/femicidio en América Latina y el Caribe. Más información disponible en: <<https://oig.cepal.org/es/autonomias/autonomia-fisica>>.

2018). Es esperable que estos cambios promuevan la producción de datos de calidad en el futuro, lo que es fundamental para el desarrollo de políticas basadas en evidencias que puedan reducir la incidencia de los femicidios/feminicidios en una región caracterizada, a la vez, por la violencia y la desigualdad desde una perspectiva de género.

La desagregación de los homicidios por etiología plantea dificultades derivadas, entre otras cosas, de la superposición de tipos (un homicidio podría implicar, por ejemplo, un femicidio en el marco de otras actividades criminales) y de una frecuente falta de informaciones mínimas para asignar los casos a las categorías. Sin embargo, avanzar en este abordaje es fundamental para desarrollar medidas preventivas adecuadas a cada situación. De hecho, reconocer que el homicidio es un fenómeno con múltiples etiologías supone abandonar cualquier pretensión de solución única y reconocer la necesidad de diagnósticos precisos para desarrollar iniciativas de prevención.

1.4 Datos sobre homicidio

Datos precisos y confiables son esenciales para mejorar la comprensión de los mecanismos que generan y perpetúan la violencia en la región. Asimismo, son una precondition para la evaluación de las políticas públicas en la materia, lo que permite conocer si las iniciativas funcionan o no y en qué medida. Por último, contar con datos de calidad es fundamental en términos de transparencia y rendición de cuentas a la ciudadanía.

1.4.1 Fuentes de información

Las informaciones sobre homicidio suelen consistir en registros administrativos, es decir, datos producidos por el funcionamiento regular de las instituciones públicas. Estos datos se derivan de dos fuentes principales: los registros criminales y los certificados de defunción.

Los registros criminales más utilizados son las denuncias policiales, aunque en esta categoría también se incluyen los datos del Ministerio Público (acusaciones, denuncias, etc.), el Poder Judicial (procesos, sentencias, etc.) y el sistema penitenciario (reclusos, etc.). Cada fuente del sistema de justicia criminal presenta ventajas y desventajas, siendo en general recomendable asumirlas como partes complementarias de un todo (Cepal, 2014). Los registros policiales se aproximan más al número de homicidios y, en ese sentido, puede decirse que tienen mayor validez.¹⁷ Esto sucede porque, en muchos casos, no es posible identificar, encontrar y procesar al autor de los homicidios. Por otra

parte, las estadísticas judiciales y penitenciarias tienden a ser más fiables que las policiales, ya que mientras una denuncia podría o no vincularse a un delito real, un procesamiento o una condena se basa en decisiones más cuidadosas e informadas por parte de las autoridades (Aebi, 2008). Adicionalmente, es importante subrayar que mientras que los datos policiales de homicidio están referidos en parte a las víctimas, los datos de procesados y condenados se refieren exclusivamente a los perpetradores.¹⁸

En este sentido, estos últimos suelen ser criticados por ofrecer una imagen incompleta y sesgada de la criminalidad, ya que apenas una proporción de los delitos son denunciados y procesados por el sistema de justicia. Sin embargo, en comparación a delitos menos graves, los homicidios suelen tener menor subregistro,¹⁹ dado que son más denunciados por la ciudadanía, concitan mayor atención de las autoridades y resultan en un cadáver humano que sirve de evidencia para que esos hechos sean descubiertos e investigados. En la misma dirección, los homicidios tienen un mayor nivel de esclarecimiento y condena²⁰ en comparación con otros delitos, por lo que sus datos se consideran de calidad superior.

Por su parte, los certificados de defunción son documentos producidos por profesionales de la salud en los que se indica la causa de muerte de un individuo. Estas causas de muerte son establecidas con fines epidemiológicos por la CIE de la OMS. Mediante este instrumento se elaboran las estadísticas de mortalidad y se registran en particular los homicidios (códigos X85 a Y09 de la CIE-10) y las otras muertes por causa externa, como los accidentes y los suicidios. Como fue señalado antes, otras categorías de interés, como las muertes por intervenciones legales (Y35) y las muertes en guerras (Y36), son contabilizadas por fuera de los homicidios (OMS, 2015).

En relación con los registros criminales, los datos derivados de certificados de defunción son más comparables internacionalmente al no estar sujetos a variaciones jurídicas. En contrapartida, ofrecen un conjunto más acotado de informaciones, dado que se centran apenas en las características de las víctimas. Los detalles sobre las circunstancias de la muerte y el perfil del presunto agresor otorgan un mayor potencial de análisis a los registros criminales. Aunque, en teoría, cada fuente presenta ventajas y desventajas, la calidad de los datos de una u otra es variable según el caso concreto. Con todo, contar con dos fuentes es en sí una ventaja, ya que una de las mejores formas de examinar la calidad de los datos es verificar el grado de convergencia que existe entre ellas.²¹ Además de los certificados de defunción y de los registros criminales existen otras fuentes de información sobre los homicidios. En algunas circunstancias estas fuentes representan un insumo valioso

17 La validez de un indicador corresponde a su capacidad de medir aquello que pretende.

18 La unidad de recuento de los homicidios varía a lo largo de proceso criminal. En general, la policía se centra en el registro del número de víctimas, el Ministerio Público en el número de presuntos criminales, el Poder Judicial en el número de procesos penales o hechos delictivos, y el sistema penitenciario en el número de condenados. Al contabilizar fenómenos distintos, la comparabilidad de los registros entre las diversas instituciones puede llegar a ser problemática.

19 El subregistro delictivo consiste en la diferencia entre los hechos criminales denunciados y los ocurridos efectivamente en la sociedad.

20 El esclarecimiento y la condena se refieren a la capacidad de las autoridades competentes de identificar y castigar a los y las responsables de los delitos.

para revisar y mejorar la cobertura de las cifras oficiales, mientras que en otras ocasiones pueden ser el único modo de medir la violencia letal. La fuente de información no oficial más utilizada es la prensa escrita, donde se publican periódicamente noticias sobre homicidios que luego pueden ser sistematizadas por los investigadores y las investigadoras (*press clipping*). Este es un abordaje especialmente interesante para relevar hechos de alta repercusión pública, como los femicidios/feminicidios, y es más limitado para contabilizar homicidios comunes en contextos de alta incidencia o en lugares donde la libertad de prensa no está garantizada.

En países como Venezuela, donde en los últimos años no siempre hubo disponibilidad pública de datos oficiales y donde la incidencia es muy elevada, se están empleando encuestas de victimización a familiares de las víctimas para estimar la tasa de homicidios,²² técnica que carecería de sentido en circunstancias convencionales.

Aunque los registros de homicidio tienden a ser mejores que los de otros delitos, contar con datos de calidad es todavía un reto para varios países de América Latina y el Caribe. En la región se producen estadísticas que a menudo son insuficientes, inconsistentes y poco confiables (Dammert *et al.*, 2010). En algunos casos, las autoridades ni siquiera reportan cifras sobre criminalidad, mientras que en otros los datos son difundidos de manera desactualizada y agregada, lo que limita enormemente su utilidad para la prevención y la evaluación.²³ En particular, los países con alta incidencia de violencia suelen presentar estadísticas oficiales de menor validez y confiabilidad, en parte como consecuencia de la fragilidad de las instituciones responsables de la seguridad. Se plantean dos dificultades principales. La primera es que en tales condiciones las instituciones a cargo de producir los registros dejan de funcionar correctamente, lo que afecta la calidad y completitud de los datos. La segunda es que, en algunas circunstancias, la distinción entre violencia criminal y violencia política puede tornarse difusa, lo que afectaría la clasificación de esos casos como homicidios (Eisner, 2013).

Recientemente, la UNODC ha desarrollado una metodología para estimar la calidad de los datos de homicidio en diferentes países.²⁴ Específicamente para América Latina y el Caribe también existen otros instrumentos idóneos, como el Protocolo de Bogotá.²⁵ De modo general se recomienda triangular las fuentes de información disponibles para analizar eventuales inconsistencias y, fundamentalmente, identificar cambios repentinos en los registros que pudieran afectar los resultados de la evaluación. Es decir, más allá del grado de calidad de los datos de homicidio, a efectos de la evaluación lo más importante es saber si hay variaciones durante el período considerado (antes, durante y después de la intervención) que pudieran llevar a confundir un cambio objetivo de

21 Un estudio reciente indica que los países de América Latina y el Caribe tienen, en promedio, 60% más homicidios cuando se contabiliza este fenómeno con base en registros criminales que cuando se hace con certificados de defunción (Ribeiro, Borges y Cano, 2015).

22 Tradicionalmente, las encuestas de victimización se emplean para estimar la incidencia de diferentes delitos, pero no para los homicidios. Cuando se trata de homicidios, la pregunta es del tipo siguiente: ¿En los últimos 12 meses algún miembro de su hogar ha sido víctima de homicidio intencional? Este abordaje metodológico solo es viable con muestras muy grandes y en lugares donde las tasas de homicidio son elevadas. Por otra parte, esta técnica no contabiliza las víctimas institucionalizadas ni las personas que vivían solas antes de ser asesinadas.

23 Como se desprende del apartado 1.2.1, datos puntuales y con coordenadas geográficas que permitan el análisis espacial siempre serán mejores que estadísticas agregadas sobre unidades más amplias. Estos datos también son importantes para la precisión, la evaluación y el control de los efectos de desplazamiento criminal (ver apartado 11.3).

24 Para mayor información, ver: <https://www.unodc.org/documents/data-and-analysis/gsh/Meth_Annex_GHS.pdf>.

25 Para más información, consulte: <<http://homicidioslatam.org/>>.

incidencia del fenómeno con simples modificaciones en los criterios de registro o medición ([ver apartado 11.6](#)).

1.4.2 Indicadores más usados

Como se mencionó antes, los homicidios suelen tener menor subregistro y mayor esclarecimiento y condena en comparación a otros delitos menos graves. La posibilidad de triangular información proveniente del sector de la salud (certificados de defunción) y del sistema de justicia criminal (delito de homicidio) es otro elemento que puede contribuir a mejorar los datos. Estas características, sumadas al hecho de que se trata de una medida simple y habitualmente disponible en períodos de larga temporalidad, hacen de la tasa de homicidios el principal indicador para los estudios sobre violencia.

La tasa de homicidio mide el riesgo de ser víctima de homicidio que existe en determinado lugar (país, ciudad, etc.) y momento del tiempo. En general, estas tasas son anuales, pero pueden estar referidas a un período diferente. Consiste en un cociente que tiene como numerador la cantidad de homicidios ocurridos en un territorio y en el denominador la población total de ese mismo territorio. Dado que el homicidio es un fenómeno estadísticamente raro o infrecuente, el cociente es escalado a 100.000 para evitar datos con números decimales que sean difíciles de manejar.²⁶ De ese modo, la tasa de homicidios representa el número de homicidios ocurridos anualmente en un determinado territorio por cada 100.000 habitantes, y es una medida útil para comparar la incidencia a lo largo del tiempo o en distintos territorios o sociedades, independientemente de su tamaño. No obstante, cuando el territorio o sociedad es menor a la escala de medida (100.000 personas) las tasas se vuelven inestables y pueden brindar una imagen distorsionada de la situación.²⁷ Para enfrentar este problema existen dos alternativas simples: calcular tasas plurianuales considerando la población y los homicidios correspondientes a varios años, o calcular tasas considerando poblaciones mayores, incluyendo, por ejemplo, los territorios limítrofes o que son parte de la región. Una posibilidad más compleja es el cálculo de estimaciones espaciales bayesianas, que calculan las tasas de cada territorio como una media entre su propio valor y el de los territorios circundantes. Aunque es más común que se utilice la tasa de victimización por homicidio, también es posible calcular la tasa de perpetración, llevando en cuenta el número de victimarios en lugar del de las víctimas.

Una limitación metodológica de la tasa de homicidios es su sesgo en función de la población flotante, fenómeno que incide en el numerador (ocurrencia de homicidios), pero no en el denominador (población residente). La población flotante está compuesta por personas que pasan

26 Cálculo de tasas de homicidio: $[(\text{Homicidios ocurridos en un territorio o sociedad} / \text{Población total del territorio o sociedad}) \times 100.000]$.

27 Por ejemplo, un departamento de un pequeño país del Caribe con una población de 2000 personas que no ha registrado ningún homicidio en los últimos 9 años, pasaría de 0 a tener una tasa de 50 cada 100.000 con la ocurrencia de un único caso en el décimo año.

regularmente horas o días en un área, pero no residen en ella. Si esas personas son víctimas de homicidio en ese local serán contadas en el numerador, pero no en el denominador, puesto que no son residentes, sesgando al alza el valor de la tasa.²⁸

Para evitar este sesgo se sugiere calcular dos tasas diferentes: la tasa de homicidio por local de ocurrencia y la tasa de homicidio por local de residencia. La primera considera en el numerador los homicidios ocurridos en un territorio y en el denominador la población residente más la población flotante de ese territorio.²⁹ La segunda pone en relación la población residente de un territorio que fue víctima de homicidio con la población total residente en ese territorio.³⁰ Nótese que la tasa de homicidios por local de ocurrencia considera las muertes ocasionadas en un lugar, independientemente de dónde residan las personas involucradas; en tanto que la tasa de homicidios por local de residencia considera las muertes de personas residentes en un lugar, independientemente de dónde hayan sucedido estos hechos. La conveniencia de una u otra de estas tasas dependerá del tipo de uso que se pretenda darles. Por ejemplo, la tasa por local de ocurrencia sería más útil si estamos pensando en implementar un programa de refuerzo de la presencia policial en las áreas más violentas de la ciudad. Por otro lado, la tasa por local de residencia sería más adecuada si queremos intervenir en la prevención social temprana de personas con riesgo de victimización en determinada comunidad.

Una forma de medir con más precisión el riesgo de homicidio es considerar poblaciones específicas en vez de la población en general. Las tasas específicas se construyen calculando la tasa de homicidios apenas para una parte de la población. Un ejemplo es la tasa de homicidio juvenil, entendida como el número de homicidios de personas entre 15 y 29 años por cada 100.000 habitantes de ese mismo tramo etario en determinado territorio.³¹

Adicionalmente, las tasas específicas de diferentes subpoblaciones pueden compararse entre sí mediante el cálculo de riesgos relativos. El riesgo relativo indica cuántas veces es mayor o menor el riesgo de una población en comparación con otra. Un ejemplo es el riesgo relativo de homicidio por sexo, calculado como el cociente entre la tasa específica de homicidio de hombres y la tasa específica de homicidio de mujeres.³² Dicho cociente asume valor 1 cuando el riesgo de hombres y mujeres es igual, es mayor que 1 cuando el riesgo de los hombres es superior y menor a 1 cuando el de las mujeres es superior. Un riesgo relativo de 10, siguiendo con el ejemplo, significaría que los hombres tienen diez veces más riesgo que las mujeres de ser víctimas de homicidio.

28 La tasa de homicidios en el centro de las ciudades tiende a ser sobrestimada, ya que la población flotante es numerosa, pero la residente es mucho menor. Lo opuesto puede observarse en las “ciudades dormitorio”.

29 Tasa de homicidio por local de ocurrencia: $[(\text{Homicidios ocurridos en un territorio} / \text{Población residente más población flotante del territorio}) \times 100.000]$.

30 Tasa de homicidio por local de residencia: $[(\text{Residentes del territorio víctimas de homicidio} / \text{Población residente del territorio}) \times 100.000]$.

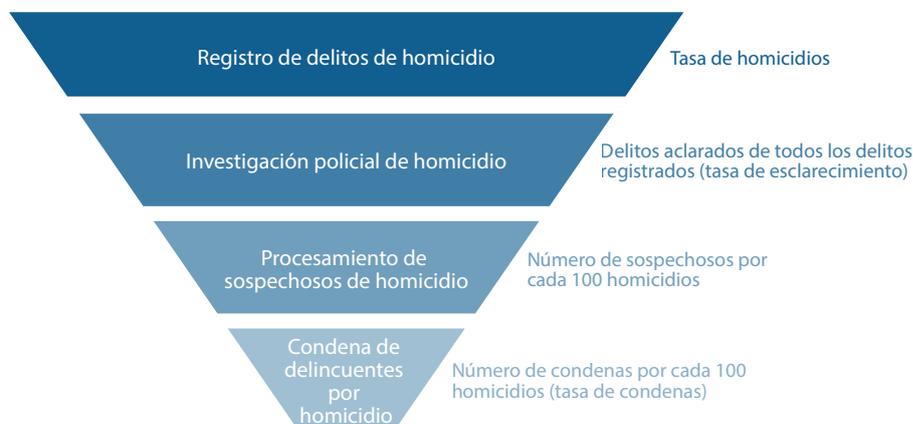
31 Tasa específica de homicidio juvenil: $[(\text{Víctimas de homicidios entre 15 y 29 años ocurridos en un territorio o sociedad} / \text{Población total residente de personas entre 15 y 29 años en un territorio o sociedad}) \times 100.000]$.

32 Riesgo relativo de homicidio por sexo: $\text{tasa específica de homicidio de hombres} / \text{tasa específica de homicidio de mujeres}$.

Otros indicadores a mencionar son los vinculados al proceso penal, que, por su naturaleza, informan sobre la eficacia del sistema de justicia criminal en dar respuesta a estos hechos. Las medidas más utilizadas son la tasa de esclarecimiento y la tasa de condena. La primera representa el porcentaje de homicidios esclarecidos entre todos los homicidios denunciados o registrados por la policía. Frecuentemente, esto significa que se ha identificado a uno o varios sospechosos y que el caso ha sido entregado al Ministerio Público, pero el hecho también puede esclarecerse por otras causas, como cuando se constata la muerte del ofensor. Con todo, el esclarecimiento es un concepto ambiguo y coexisten varias definiciones. Por ejemplo, en algunos casos se exige que el sospechoso sea preso para considerar el caso esclarecido, en otros no.

El número de sospechosos podría ser mayor o menor que el número de víctimas, dependiendo de las características de estos crímenes: será mayor en presencia de homicidios con autores múltiples y menor cuando existan hechos con varias víctimas. Algo similar ocurre con los condenados, aunque en general este número sí suele ser notablemente inferior al de las víctimas, debido a que en muchos casos las pruebas incriminatorias resultan insuficientes.

Esquema 2. Indicadores relacionados a las fases del sistema de justicia criminal



Fuente: UNODC (2019). Global Study on Homicide: Homicide Trends, Patterns and Criminal Justice Response. Vienna.

Lo anterior remite al segundo indicador procesal: la tasa de condena, medida como la cantidad de condenas por homicidio por cada 100 homicidios ocurridos el mismo año. Esta tasa es una de las principales medidas de la impunidad que existe en una sociedad, aunque tiene limitaciones. La principal es la temporalidad, ya que se comparan los homicidios perpetrados y condenados en un mismo año, siendo que las investigaciones criminales suelen extenderse más allá de ese período. Por tanto, los homicidios del numerador son en buena parte diferentes

de los del denominador. De hecho, los homicidios recientes no tuvieron tiempo aún de ser investigados y, sin embargo, acaban contando como si no hubieran sido resueltos. Paralelamente, la tasa incluye condenas que corresponden a homicidios que sucedieron hace años. Alternativamente, la solución sería calcular la proporción de homicidios registrados en un año que han llegado a una condena judicial después de un cierto tiempo, digamos un período de 5 años. Pero para eso es necesario un sistema de informaciones que acompañe los casos longitudinalmente.

1.4.3 Sistemas de información

La información es un recurso crítico para cualquier programa y sirve para fundamentar las operaciones y las decisiones. Permite nada menos que conocer la realidad, aunque es siempre parcial en dos sentidos: 1) es selectiva, pues se refiere a una parte de la realidad, ya que no es posible incorporarlo todo; 2) parte de una observación basada en determinados marcos conceptuales y en propósitos específicos.

Un sistema de información es un mecanismo que contempla la recolección periódica, el almacenamiento, el procesamiento, el análisis y la difusión de las informaciones.³³ En general, comprende y vincula datos de distinta naturaleza, como las personas atendidas, los insumos que el programa demanda (recursos humanos, económicos, etc.), los servicios que se realizan (actividades, productos, etc.) y una serie de eventos de interés. Entre los eventos de interés ciertamente constarán los homicidios, pero también pueden figurar, dependiendo del caso, factores de riesgo o de protección, como la presencia de servicios públicos en las comunidades.

Estos sistemas suelen contemplar datos primarios, que son producidos en el marco del funcionamiento del programa, y secundarios, generados por terceros. Los datos primarios pueden ser recogidos de diferentes formas, como encuestas, análisis de prensa, observación directa o entrevistas a beneficiarios y beneficiarias; y con los registros administrativos del programa, como en el caso de los servicios ofrecidos. Los datos secundarios son a menudo producidos por organismos públicos, ya sea con fines estadísticos (como los censos de población) o administrativos (como los registros criminales), aunque también pueden incluirse en esta categoría las informaciones provenientes de estudios realizados por centros de investigación, empresas de opinión pública u organizaciones de la sociedad civil. De este modo, diversas fuentes son integradas para complementar las estadísticas oficiales básicas y permitir un abordaje más instruido sobre la realidad a intervenir, que será posteriormente evaluada.

Idealmente, el sistema de información para la evaluación de un programa de prevención de homicidios debe contar con al menos dos bases de datos:

33 Entre los elementos que componen un sistema de información, se destacan: 1) los equipos informáticos (*hardware*), 2) los programas (*software*) y 3) el personal (recursos humanos).

- Base de datos sobre los homicidios, abarcando el territorio en el que se realizó la intervención y otros lugares de características físicas y sociales similares que podrían funcionar como grupo de control. Esta información, idealmente, debería incluir un período de 20 unidades temporales (semanas, meses, etc.) anteriores y 20 unidades temporales posteriores al inicio del programa, para una evaluación más precisa. A menudo el uso de registros oficiales es suficiente para atender esta demanda de información, pero en caso de que esos datos no estén disponibles o no sean de calidad, el programa deberá producir sus propios registros de homicidio para la evaluación.
- Bases de datos sobre el funcionamiento del programa, conteniendo información sobre los recursos humanos y financieros, la cobertura, los servicios prestados (según tipo, intensidad, tiempos, etc.), y beneficiarios y beneficiarias individuales. Esta información es clave para caracterizar la intervención.

Además de los aspectos vinculados a la gestión operativa (distribución de recursos, priorización de casos a ser atendidos, supervisión, etc.), los sistemas de información suelen ser empleados para realizar análisis y diagnósticos, evaluar y tomar decisiones estratégicas basadas en evidencias.

El fenómeno del homicidio

Aunque existen diferentes definiciones de homicidio, en general nos referimos a la muerte de una persona provocada por la agresión intencional de otra(s).

Este fenómeno posee dos regularidades muy destacadas: la concentración geográfica y social.

El homicidio agrupa distintas manifestaciones extremas de violencia, cada una con sus causas, tendencias y poblaciones afectadas.

Existen dos fuentes principales de información sobre el homicidio: los registros criminales y los certificados de defunción.

Dados ciertos atributos metodológicos, la tasa de homicidios es el principal indicador para los estudios sobre violencia. La fórmula más común de cálculo es: [(Homicidios ocurridos en un territorio / Población total residente en el territorio) x 100.000]

Para calcular la tasa de homicidios de forma más precisa, pueden calcularse de forma separada tasas según el local de ocurrencia y según el local de residencia. También es posible calcular

RESUMEN

tasas específicas que midan el riesgo de homicidio apenas para determinadas subpoblaciones (jóvenes, mujeres, etc.).

Otros indicadores importantes que están asociados al funcionamiento del sistema de justicia criminal son la tasa de esclarecimiento y la tasa de condenación.

Contar con sistemas de información adecuados es fundamental para la evaluación de los programas de prevención de homicidios. Se requieren al menos dos bases de datos: una sobre los homicidios y otra sobre el funcionamiento del programa.

CAPÍTULO 2. PROGRAMAS DE PREVENCIÓN DEL HOMICIDIO

2.1 Introducción

La violencia está generando un interés creciente por parte de la comunidad internacional. En 2014, la Asamblea Mundial de la Salud solicitó a la OMS la elaboración de un plan de acción para prevenir y responder a las situaciones de violencia (OMS, 2014); y en 2015, la Asamblea General de las Naciones Unidas adoptó como uno de los objetivos de la Agenda 2030 para el Desarrollo Sostenible la “promoción de sociedades pacíficas e inclusivas” (Objetivo de Desarrollo Sostenible 16). Aunque la importancia de prevenir la violencia ha sido destacada internacionalmente, las estrategias sugeridas para hacerlo tienden a seguir un abordaje generalista, poco focalizado en los homicidios. Esta perspectiva ha sido estimulada en América Latina y el Caribe por varias agencias internacionales, lideradas por el Banco Interamericano de Desarrollo (BID), como una alternativa a las políticas represivas de mano dura, tan populares en varios países de la región. A partir del enfoque de seguridad ciudadana se postula un amplio abanico de intervenciones con la expectativa de prevenir la criminalidad en general y, eventualmente, reducir la violencia letal (Aguirre y Muggah, 2017).

Frente a la gravedad y singularidad de los homicidios en América Latina y el Caribe³⁴ cabría esperar una proliferación de políticas públicas específicamente destinadas a enfrentar el problema. Sin embargo, no es esto lo que ha sucedido hasta ahora. Las iniciativas para contener los homicidios son escasas y, lamentablemente, solo una pequeña proporción de ellas ha sido objeto de evaluaciones de impacto que permitan concluir qué programas han tenido éxito y en qué medida (Cano y Rojido, 2016). Con todo, la región puede revertir esta situación y tal vez haya comenzado a hacerlo. Las iniciativas para reducir los homicidios son recientes en su mayoría, presentan intervenciones de naturaleza variada y son promovidas por diversos actores tanto del poder público como de la sociedad civil organizada. Adicionalmente, el [Plan de Acción Hemisférico para Orientar la Elaboración de Políticas Públicas de Prevención y Reducción del Homicidio Intencional](#), impulsado por la Secretaría de Seguridad Multidimensional de la Organización de Estados Americanos (OEA) en 2019, representa un hito sin precedentes tanto en el reconocimiento regional del problema como en el fomento de intervenciones focalizadas para enfrentarlo.

34 Para mayor información, [ver Capítulo 1](#).

2.2 Tipos de prevención

Prevenir es reconocer el riesgo de que un evento acontezca y realizar alguna acción para eliminarlo o reducirlo. Por lo tanto, prevenir los homicidios implica anticiparse y evitar estos hechos. Diferente de la prevención es la represión, que consiste en perseguir y castigar a los autores de un hecho ilícito una vez que estos fueron perpetrados. Con todo, la distinción entre prevención y represión no siempre es clara y unívoca. Por ejemplo, tener un sistema de justicia criminal eficaz y reducir los niveles de impunidad puede disuadir a terceros de cometer delitos. A su vez, construir comunidades sanas e integradas no solo promueve el control social y la prevención, sino que fortalece la contribución activa de la población (denunciando, atestiguando, etc.) en el esclarecimiento de los delitos por parte de las autoridades.

Las iniciativas de prevención pueden ser categorizadas **en función del tipo de población** a las que se destinan. La *prevención primaria o universal* se destina a todos los individuos de una sociedad. Es amplia y exhaustiva, pero también poco específica y menos intensiva. Un ejemplo en este sentido podría ser una campaña de sensibilización contra la violencia en medios masivos de comunicación, como la radio o la televisión.

La *prevención secundaria o selectiva* está centrada en grupos con alto riesgo de convertirse en víctimas o perpetradores de la violencia. El número de beneficiarios y beneficiarias es menor y las intervenciones son más intensivas que en el caso anterior. Uno de los aspectos fundamentales para este tipo de prevención es la precisión de la focalización, que suele realizarse en función de criterios sociales (por ejemplo, jóvenes entre 15 y 29 años) o territoriales (residentes de los barrios con mayor tasa de homicidios). Un ejemplo de prevención secundaria son los programas de protección a testigos y funcionarios que participan en el proceso de justicia criminal, personas que están expuestas a un riesgo elevado de sufrir represalias.

La *prevención terciaria o señalada* se concentra en individuos que ya son víctimas o autores de la violencia para evitar la revictimización o la reincidencia. Este tipo de abordaje requiere de medidas muy intensivas, destinadas a un número reducido de beneficiarios y beneficiarias que presentan múltiples factores de riesgo. Los programas que buscan la rehabilitación de agresores combinando distintas intervenciones (cognitivo-conductuales, familiares, laborales y educativas, etc.) podrían ser un ejemplo (Tilley, 2009).

Otro criterio utilizado para clasificar las iniciativas de prevención es su **naturaleza**. La *prevención social* busca intervenir sobre las causas sociales de la violencia y la criminalidad (pobreza, bajo nivel educativo, dificultades de empleo, etc.), mejorando las condiciones de vida de

las personas para disminuir la probabilidad de que se envuelvan en comportamientos violentos o delictivos. Estas iniciativas pueden estar dirigidas a individuos, grupos o comunidades en su conjunto, y suelen tener objetivos que trascienden la prevención de la violencia, por lo que se confunden a menudo con las políticas sociales. En este sentido, un ejemplo de prevención social podría ser la capacitación laboral y la creación de oportunidades de empleo juvenil en barrios en situación de vulnerabilidad social.

La *prevención situacional*, por su parte, se compone de medidas de reducción de oportunidades delictivas que: 1) están dirigidas a formas específicas de delincuencia; 2) involucran la manipulación del entorno inmediato; 3) hacen que el delito sea más difícil y arriesgado, o menos gratificante. A diferencia de la prevención social, que intenta reducir la motivación para el comportamiento criminal, la prevención situacional busca influir principalmente sobre las circunstancias ambientales (físicas y sociales) que aumentan la probabilidad de ocurrencia de esos delitos.³⁵ El uso de sistemas de videovigilancia o la mejora de la iluminación son medidas que pueden ser entendidas dentro de este marco.

Por último, la *prevención policial* consiste en estrategias desarrolladas por las instituciones policiales para reducir la violencia criminal. Las posibilidades en este sentido son varias, pudiendo incluir la disuasión a partir de la simple presencia de las autoridades en el espacio público o la interrupción de futuras actividades delictivas en curso a partir de operaciones de inteligencia. Un ejemplo de prevención policial es la vigilancia por cuadrantes, que asigna responsabilidades a policías que prestan servicios en una pequeña área o territorio.

Aunque la distinción entre prevención primaria, secundaria y terciaria, por un lado, y social, situacional y policial, por el otro, es analíticamente útil, estas categorías se combinan entre sí (derivan, por ejemplo, en iniciativas de prevención social y terciaria), siendo además bastante común en la práctica, que los programas integren componentes de diverso tipo, tanto en función del público al que se destinan (por ejemplo, primaria y secundaria) como del abordaje que proponen (situacional y policial).³⁶

2.3 Tipos de programa³⁷

Un estudio reciente se propuso identificar y analizar los programas de prevención de homicidios existentes en América Latina y el Caribe, para contribuir a una mejor comprensión de estas iniciativas y ofrecer una gama de opciones realistas a los políticos y miembros de la sociedad civil que deseen emprender o mejorar sus intervenciones en la materia. Fueron sistematizados programas que tenían una meta explícita de reducción

35 Esta perspectiva se asocia a la teoría de actividades rutinarias, que sostiene que los delitos se producen cuando coinciden en el espacio y en el tiempo un ofensor motivado y una víctima potencial en ausencia de un guardián (Cohen & Felson, 1979). Prevenir los delitos supondrá, por tanto, intervenir sobre las situaciones para evitar que esa coincidencia se concrete y genere oportunidades delictivas.

36 Para introducción a la literatura sobre prevención del crimen, ver Tilley (2009), Welsh & Farrington (2007) y Weisburd, Farrington & Gill (2016).

37 Este apartado está basado en Cano y Rojido (2016): *Mapeo de programas de prevención de homicidios en América Latina y el Caribe*. Una actualización posterior de los programas de prevención de homicidios también puede ser consultada en: <http://homicidioslatam.org>. Este sitio web ofrece información ampliatoria sobre todos y cada uno de los programas aquí mencionados.

de homicidios y, adicionalmente, iniciativas en el área de seguridad con un impacto comprobado en la disminución de los homicidios, incluso cuando esa no era su intención original (Cano y Rojido, 2016).

A pesar del número reducido (109),³⁸ estas iniciativas destacan por su variedad. Los programas fueron clasificados en 15 tipos agrupados en 6 estrategias o áreas temáticas: a) control de factores de protección o de riesgo; b) promoción de cambios culturales; c) protección a grupos de riesgo; d) mejoras en el funcionamiento del sistema de justicia criminal; e) programas de reinserción, mediación o negociación dirigidos a victimarios; y f) estrategias integradas de reducción de la violencia. Se elaboró así una tipología empírica, es decir, basada en la existencia de casos concretos desarrollados en la región y no en una reflexión teórica sobre todas las posibilidades existentes. En este sentido, no se trata de una tipología exhaustiva, ya que no se incluyen programas que podrían ser aplicados en el futuro ni tampoco experiencias que hayan sido aplicadas afuera de América Latina y el Caribe.

Estrategia I. Control de factores de riesgo o de protección. Esta estrategia intenta incidir sobre factores que han demostrado estar asociados a un menor o mayor riesgo de homicidios. De esa forma, no se trata de un abordaje sistémico sobre todos los posibles elementos que podrían incidir en la violencia letal, sino de intervenciones puntuales sobre factores específicos.

Tipo 1: Control de armas de fuego (9 programas)

Las armas de fuego son el principal vector de la violencia letal en el mundo, pero su peso es aún mayor en el continente americano.

El control de armas de fuego puede ser ejercido de diversas formas:

- a) *Disminución del stock de armas de fuego* en circulación, por ejemplo, mediante programas de entrega o canje de armas. Este tipo de programas posee larga tradición en la región, aunque en general no es posible determinar su impacto en el stock de armamento. El Programa Nacional de Entrega Voluntaria de Armas de Fuego, implementado desde 2007 en Argentina, no habría logrado una reducción de los homicidios ni de los suicidios, aunque sí de las muertes accidentales con armas de fuego (Lenis, Ronconi y Schargrodsky, 2010).
- b) *Restricción del porte de armas*. En este caso, el objetivo no es eliminar las armas, sino evitar que sean portadas en ámbitos públicos y, con ello, disminuir la probabilidad de su utilización. Algunas de estas iniciativas, como las que se llevaron a cabo en Bogotá y en Cali, cuentan con evaluaciones rigurosas que constatan un impacto moderado pero significativo en la reducción de homicidios (Villaveces *et al.*, 2000).³⁹
- c) *Control de incidentes armados* en determinadas áreas o dentro de instituciones. En Costa Rica, por ejemplo, la iniciativa Escuela Libre de Armas⁴⁰ intenta controlar los incidentes armados en los centros educativos y ofrece orientaciones sobre cómo actuar en caso de que se produzcan.

38 El número total de programas sistematizados fue 109. Aunque no es frecuente, algunas iniciativas corresponden a más de un tipo simultáneamente, por lo que la suma de programas presentados de cada tipo no coincide con el total.

39 [Ver apartado 14.3.](#)

40 Ver <https://www.unicef.org/PROTOCOLO-ACTUACION-ARMAS.pdf>.

Programa Nacional de Entrega Voluntaria de Armas de Fuego (PEVAF), Argentina

El PEVAF fue creado por la Ley 26216 (conocida como “Ley de Desarme”) y comenzó a implementarse en julio de 2007 a cargo del Ministerio de Justicia y Derechos Humanos, con los siguientes objetivos: 1) la disminución del uso y proliferación de armas de fuego y municiones; 2) la reducción de accidentes y hechos de violencia ocasionados por el acceso y uso de armas de fuego; 3) la sensibilización acerca de los riesgos de la tenencia y uso de armas; y 4) la promoción de una cultura de no violencia y resolución pacífica de conflictos que desaliente la tenencia y el uso de armas de fuego.

El principal componente de esta iniciativa fue el ofrecimiento de una compensación monetaria a los ciudadanos que entregaran de forma voluntaria, municiones y armas de cualquier tipo y calibre. El programa garantizó en todo momento el anonimato de los participantes y procedió a destruir las armas y municiones recolectadas.

Tipo 2: Control del consumo de alcohol (3 programas)

El alcohol está asociado a la violencia interpersonal en los conflictos cotidianos, que puede provocar resultados fatales. Así, en muchos países la incidencia de homicidios aumenta en días y horarios festivos, en los que el consumo de alcohol también es elevado.

El control del consumo de alcohol puede traducirse en campañas de concientización, así como también en restricciones a los días y horarios en que los establecimientos pueden dispensar este tipo de bebidas. Bogotá (“La Hora Zanahoria”) y Diadema (Brasil) están entre los casos más conocidos en la región de ciudades que limitaron el horario de venta de alcohol en el marco de sus políticas de seguridad.

La “Lei Seca” en Diadema, Brasil

Diadema es un municipio de la zona metropolitana de San Pablo que era, a fines de los años noventa, uno de los más violentos de Brasil. Un diagnóstico de los homicidios ocurridos en Diadema entre 1997 y 1999 reveló que esas muertes sucedían principalmente los fines de semana, en horas de la noche y en las inmediaciones de bares, posiblemente a causa de conflictos interpersonales que escalaban en intensidad. En función de ese análisis, una de las principales medidas del Plan Municipal de Seguridad de Diadema sería la aprobación de una normativa, popularmente conocida como “Lei Seca”. La ley municipal n.º 2107 del año 2002 estableció el horario de 06 a 23 horas para el funcionamiento de bares y creó un sistema progresivo de sanciones por incumplimiento. Sin embargo, quizás lo más destacable de esa iniciativa fue el operativo de fiscalización, basado en equipos integrados por fiscales municipales, miembros de la Guardia Civil municipal y de la Policía Militar. La ciudadanía también desempeñó un papel decisivo apoyando la iniciativa y colaborando con las denuncias.

Tipo 3: Prevención situacional en espacios públicos (3 programas)

La prevención situacional cuenta con una larga tradición en América Latina y el Caribe, principalmente por medio de la recuperación de espacios públicos deteriorados que son asociados a la presencia de violencia o, más comúnmente, a una percepción de inseguridad por parte de los ciudadanos. Estas iniciativas suelen estar encuadradas dentro de planes de prevención de la violencia en un sentido amplio, pero excepcionalmente pueden inscribirse en una estrategia de reducción de los homicidios o tener un efecto comprobado en este sentido. Ese sería el caso, por ejemplo, del Plan de Desarrollo Territorial de Medellín (Cerdá *et al.*, 2012).⁴¹

Plan de Recuperación y Mantenimiento de Espacios de Encuentro para el Fortalecimiento y Embellecimiento de Caracas, Venezuela

Este programa, ejecutado por el gobierno municipal de Sucre, busca la recuperación y creación de espacios públicos. Las acciones realizadas incluyen:

- 1) la recuperación de la Redoma de Petare, a partir de devolverle el orden y la seguridad;
- 2) llenar Sucre de verde, incrementando las áreas verdes por habitante;
- 3) la adopción de un parque, que supone recuperar y mantener los espacios verdes en alianza con el sector privado; y
- 4) una cancha por barrio, que consiste en la construcción y rehabilitación de espacios deportivos y recreativos. De 2009 a 2011, la Alcaldía de Sucre habilitó 110 espacios públicos: veintinueve parques, setenta y tres canchas deportivas, cinco plazas, un polideportivo, un boulevard y una pasarela. De acuerdo a una evaluación preliminar, la mejora de estos espacios públicos habría logrado reducir los homicidios en el municipio (García, 2012).

Estrategia II. Promoción de cambios culturales.

Esta estrategia pretende reducir los homicidios mediante la promoción de valores contrarios a la violencia, con el objetivo de transformar la cultura.

Tipo 4: Promoción de valores contra la violencia letal (21 programas)

Este tipo de programas se centra en la difusión de mensajes y valores que puedan transformar a las personas para construir una sociedad menos violenta. Algunas iniciativas son específicas contra la violencia letal, como la Campaña de Sensibilización para la Valoración de la Vida en Venezuela o la campaña Guatemala 24-0, que promueve 24 horas sin homicidios. En general, este tipo de experiencias no cuenta con evaluaciones que permitan medir su impacto efectivo sobre la violencia letal.

41 [Ver apartado 14.1.](#)

Campaña de Sensibilización para la Valoración de la Vida, Venezuela

La Campaña de Sensibilización para la Valoración de la Vida fue impulsada en 2012 por el Laboratorio de Ciencias Sociales (LACSO) y el Observatorio Venezolano de Violencia (OVV), e implementada con colaboración de la Central Única de Autos Libres y por Puesto. Esta iniciativa implicó la creación y colocación de más de mil afiches en puertas y vidrios traseros de ómnibus y taxis. El propósito fue sensibilizar sobre el valor de la vida por medio de una serie de mensajes que captaran la atención del potencial homicida para, idealmente, propiciar en él un cambio de conducta. El mensaje de los afiches buscaba enaltecer el valor de la vida, apelando a generar empatía a través de la figura de la madre como una víctima indirecta de la violencia homicida. Junto con el trabajo de piezas gráficas, en septiembre de 2012 también fue publicado un video que, con el mensaje “una bala trae más balas”, advertía sobre los círculos de venganza generados por la violencia. La población objetivo de la campaña fueron las personas de sexo masculino entre 15 y 25 años de edad, pertenecientes a sectores de nivel socioeconómico bajo. Para acceder a este grupo específico se priorizaron las líneas de transporte público con recorrido por los barrios más violentos y vulnerables de Caracas y otras ciudades de Venezuela.

Tipo 5: Protección a grupos de riesgo de sufrir homicidios (26 programas)

Estos programas tienen como objetivo la protección a personas, colectivos o categorías sociales que se encuentran en riesgo de sufrir homicidios. Las medidas de protección pueden ser variadas e incluyen custodia policial, retirada de las personas amenazadas de su entorno y diversas formas de apoyo económico, psicológico e institucional a las víctimas.

Entre los grupos beneficiarios de estas iniciativas se pueden mencionar:

- a) *Individuos concretos que han sufrido amenazas de muerte* y que pasan a integrar programas de protección a testigos o víctimas. Al respecto, cabe mencionar el Programa de Proteção a Crianças e Adolescentes Ameaçados de Morte (PPCAM) (Programa de Protección a Niños y Adolescentes Amenazados de Muerte) en Brasil.
- b) Miembros de *categorías profesionales o políticas de alto riesgo*, como es el caso del Programa de Prevención y Protección a Líderes Sindicales y Trabajadores Sindicalizados del Estado Colombiano.
- c) *Amplias categorías sociales que sufren riesgos específicos*. El ejemplo más claro es el de la mujer, que sufre riesgo de maltrato y, en última instancia, de muerte por parte de su pareja o expareja.

Estrategia III. Protección a grupos de riesgo.

Esta estrategia está dirigida a personas y grupos sometidos a un alto riesgo de ser víctima de homicidios, a los que se ofrece alguna clase de medida protectora. En algunos casos son grupos muy específicos, mientras que en otros se trata de amplias categorías sociales, como las mujeres.

En este sentido, varios países de la región han desarrollado leyes y programas para la prevención del femicidio o feminicidio. La introducción de la Ley “Maria da Penha” en Brasil habría generado efectos estadísticamente significativos sobre la disminución de los homicidios de mujeres asociados a la cuestión de género (Cerqueira *et al.*, 2015),⁴² aunque son pocos los casos en los que este tipo de medidas han sido evaluadas en nuestra región.

Estrategia IV. Mejoras en el funcionamiento del sistema de justicia criminal. Esta estrategia busca incrementar la eficacia del sistema de justicia en el enfrentamiento de los homicidios. Las medidas propuestas pueden estar relacionadas al patrullaje preventivo, a la investigación para incrementar la tasa de esclarecimiento de los crímenes de homicidio o a otros aspectos, como la reducción de la propia letalidad de las intervenciones policiales, en sí misma un tipo específico de homicidio.

Centros Emergencia Mujer (CEM), Perú

En el marco del Plan Nacional Contra la Violencia Hacia la Mujer (2009-2015), los CEM brindan atención a sobrevivientes de la violencia doméstica o a familiares de las mujeres que no lograron sobrevivir. Estos centros identifican casos de mujeres afectadas por hechos de violencia que, aunque no generaron evidencias para ser considerados como lesiones graves o tentativas de homicidio, reúnen elementos de riesgo que podrían conducir a un homicidio. Para la prevención del feminicidio se aplica una escala de riesgo que comprende tres ejes: agresor, víctima y relación entre ambos. El resultado de esa evaluación se expresa en categorías que clasifican al caso como leve, moderado o severo. Las acciones preventivas de los CEM están pautadas por los resultados de esa evaluación, pudiendo suponer, entre otras cosas, el traslado de las mujeres a refugios, el inicio de investigaciones policiales o procesos judiciales, además del apoyo jurídico, la atención médica y psicológica a las víctimas.

Tipo 6: Intervenciones policiales en áreas de riesgo (3 programas)

En algunos países de la región el poder público ha desarrollado intervenciones policiales innovadoras en locales de alta incidencia de homicidios, que tienen como objetivo explícito la reducción de la violencia letal o que, incluso cuando tienen objetivos más amplios, han obtenido impactos significativos en este sentido. Ejemplos del primer tipo de intervenciones son el programa Fica Vivo! en Minas Gerais o el Pacto Pela Vida en Pernambuco, mientras que un ejemplo del segundo tipo serían las Unidades de Policía Pacificadora (UPP) en Río de Janeiro, todos ellos en Brasil. La evidencia disponible apunta a que tanto el Fica Vivo! (Peixoto, Andrade y Azevedo, 2008b) como las UPP lograron reducir los homicidios (Cano, Borges y Ribeiro, 2012).

Fica Vivo!, Brasil

El Fica Vivo! es un programa diseñado por la Universidad Federal de Minas Gerais e inspirado por la experiencia de Ceasefire en Boston (Estados Unidos). En 2002 se implementó una experiencia piloto en la región de Morro das Pedras, y en 2003 la iniciativa fue institucionalizada por el gobierno estadual de Belo Horizonte y expandida a otras comunidades. El

42 Ver apartado 14.2.

objetivo de Fica Vivo! es reducir los homicidios dolosos en comunidades de alta incidencia combinando dos abordajes: la prevención social y la intervención policial o estratégica. Esta última se basa en dos elementos: el Grupo de Intervención Estratégica (GIE) y el Grupo Especializado en Policiamiento de Áreas de Riesgo (GEPAR). El GIE es un espacio que congrega a diversas instituciones del sistema de justicia para optimizar el análisis de informaciones criminales. Su actuación es clave para la “represión cualificada”, es decir, la definición de objetivos y estrategias basadas en la inteligencia policial. El GEPAR es una unidad policial de patrullaje preventivo, con vocación comunitaria, que se establece de manera permanente en los territorios donde funciona el Fica Vivo!

Las acciones de protección social, por otra parte, buscan ofrecer una red de atención local para prevenir la violencia, mejorar la convivencia y el ejercicio de ciudadanía. Para ello, se ofrecen distintos servicios a la población entre 12 y 24 años que reside en las comunidades intervenidas, como talleres de teatro, música o fútbol. Las personas que conducen los talleres son nativas de las mismas comunidades intervenidas por el programa. Más allá de su función educativa, ellas tienen el propósito de dialogar con los jóvenes, interesarlos en actividades socialmente productivas y contribuir a desactivar conflictos o a dirimirlos de forma no violenta.

Tipo 7: Otras intervenciones policiales de patrullaje (1 programa)

Se enmarcan en este tipo todos los programas policiales de prevención que no están focalizados en áreas de alta incidencia de homicidios y que pretenden reducir los homicidios o lo han conseguido en la práctica. Un ejemplo en este sentido es el Plan Nacional de Vigilancia por Cuadrantes de la Policía Nacional de Colombia, cuya evaluación reveló un impacto de disminución de los homicidios en las áreas en que fue implantado (Fundación Ideas para la Paz, 2012).

Plan Nacional de Vigilancia por Cuadrantes, Colombia

El Plan Nacional de Vigilancia por Cuadrantes es una iniciativa de la Policía Nacional de Colombia iniciada en 2010 para optimizar el servicio policial, mediante la asignación de responsabilidades concretas a los y las policías que prestan su servicio en un territorio pequeño, denominado cuadrante. El cuadrante es un sector geográfico fijo, que, a partir de sus características sociales, demográficas y geográficas, recibe distintos tipos de servicio policial, entre los cuales se cuentan la prevención, la disuasión, el control de delitos y contravenciones, y la educación ciudadana en seguridad y convivencia. Mediante esta intervención se espera obtener efectividad en la lucha contra la delincuencia, ofrecer un servicio de policía integral dentro del cuadrante y aumentar la cercanía de la policía con la ciudadanía.

- 43 Los incentivos económicos a los y las policías suelen otorgarse por estaciones o comisarías, para que reduzcan los homicidios en su zona. En estos casos, la evaluación debería distinguir diversos tipos de efecto posible: a) efecto sistémico o global de la introducción del premio sobre el conjunto del territorio; b) efecto sobre las unidades que intentan ganar el premio, por poseer indicadores que les permiten creer que pueden alcanzarlo; c) efecto posterior sobre las unidades que ya consiguieron el premio. Este último impacto puede ser tanto positivo, en razón de la motivación, como negativo, pues, dependiendo del indicador utilizado para ofrecer el estímulo económico, a veces el hecho de haber tenido éxito una vez puede tornar más difícil conseguirlo de nuevo (por ejemplo, porque cuando la tasa de homicidios ya es baja, sucesivas reducciones son cada vez más difíciles).
- 44 En varios estados de Brasil a la policía se le plantean metas de gestión por unidad (comisaría, batallón) que en caso de cumplirse generan premiaciones para los integrantes de esa unidad. En ese sentido, es común que los homicidios sean, entre otros indicadores, uno de los delitos cuya reducción se contempla. Sin embargo, el Premio Paraíba Unida por la Paz se distingue de estos casos, porque está focalizado en la reducción de los homicidios y no en el desempeño policial de un modo más general.
- 45 Más allá de este caso, el [Capítulo 14](#) presenta la evaluación de una reciente experiencia piloto de investigación criminal que fue desarrollada por la Fiscalía General de la Nación de Co-

Tipo 8: Incentivo económico a la eficacia policial (1 programa)

Estos programas buscan crear un incentivo para la reducción de homicidios mediante la concesión de bonificaciones económicas a los y las oficiales de policía por el logro de objetivos.⁴³ En principio, estos premios pueden ser individuales, grupales o institucionales. Por lo tanto, es un intento de priorizar la prevención de homicidios en el trabajo policial. Las directrices específicas para la acción no están necesariamente definidas, por lo que esta iniciativa podría llevar a la mejora de diferentes actividades, como la inteligencia policial, el patrullaje preventivo o la investigación. Un ejemplo al respecto es el Premio Paraíba Unida por la Paz (Prêmio Paraíba Unida pela Paz) en el Estado de Paraíba, Brasil.⁴⁴

Premio Paraíba Unida por la Paz, Brasil

Sancionado por la Ley 10327 de 2014 y liderado por la Secretaría de Seguridad y Defensa Social del Gobierno de Paraíba, este programa es un premio por resultados otorgado a policías militares, policías civiles y bomberos que logran reducir los crímenes violentos letales intencionales en sus áreas de actuación (AISP): homicidio doloso, latrocinio, lesión corporal dolosa seguida de muerte y otros crímenes dolosos que resulten en muerte. Al final de cada año se analizan los datos criminales previos y se definen las metas para el año subsiguiente, las cuales son revalidadas semestralmente por un Comité Gestor de la Seguridad y la Defensa Social, que es integrado por representantes de distintas instituciones del sector de seguridad pública.

Tipo 9: Mejora en la investigación de los homicidios (1 programa)

Son iniciativas que pretenden mejorar la tasa de esclarecimiento y de condena del crimen de homicidio, de forma que una proporción más elevada de los victimarios pueda ser objeto de castigo penal. Al respecto, se encontró apenas una iniciativa en Honduras, donde la Asociación para una Sociedad Más Justa, una organización de la sociedad civil, desarrolló un proyecto junto al poder público para intentar mejorar las investigaciones sobre homicidios.⁴⁵

Paz y Justicia, Honduras

Paz y Justicia es un programa de la Asociación para una Sociedad más Justa (ASJ) que desde 2005 colabora con operadores del sistema de justicia criminal en la investigación de los homicidios. La intervención se implementa en algunas colonias (barrios) de Tegucigalpa y San Pedro Sula, con tres componentes principales: el investigativo, el legal y el psicológico. La investigación criminal es realizada por investigadores

privados contratados por la institución, la mayoría expolicías, e implica el levantamiento de todas las evidencias necesarias para el esclarecimiento del homicidio. Al operar concentradamente solo en algunas colonias, los investigadores logran establecer vínculos de confianza y hacerse de informantes, construyendo así una red local de inteligencia cuya contribución es clave para el esclarecimiento de los crímenes.

La asistencia legal, a cargo de abogados, consiste en la representación jurídica de las personas beneficiarias y en el seguimiento de los casos a través de las distintas instancias del proceso penal. La asistencia psicológica es ofrecida por profesionales que realizan terapias individuales o colectivas para que los familiares y otras víctimas indirectas puedan superar el trauma de la violencia y recomponer sus vidas. Las y los beneficiarios son acompañados a lo largo de todo el proceso penal, y pueden recibir apoyos sociales circunstanciales, como cestas de alimentos, medicamentos o ayudas para la mudanza de domicilio en caso de que exista un riesgo inminente para su seguridad.

Tipo 10: Reducción de la letalidad policial (4 programas)

Algunos países de la región adolecen históricamente de exceso de uso de fuerza policial que, cuando se trata de fuerza letal, puede provocar un número elevado de homicidios. Venezuela y El Salvador son ejemplos de esta situación.⁴⁶ Sin embargo, esto no siempre es percibido como un problema por los gobiernos e incluso por sectores importantes de la sociedad, que llegan a demandar políticas “de mano dura” contra la criminalidad que alientan o toleran los abusos.

Ocasionalmente, algunas policías desarrollan programas para intentar reducir la letalidad policial. Un ejemplo lo constituye el Programa de Acompanhamento de Policiais Militares Envolvidos em Ocorrências de Alto Risco (PROAR) (Programa de Acompañamiento de Policías Militares Involucrados en Incidentes de Alto Riesgo), lanzado por el gobierno del Estado de São Paulo en 1995.

Índice de Aptitud para el Uso de la Fuerza Policial, Brasil

Esta iniciativa fue desarrollada en 2015 por la Policía Militar del Estado de Río de Janeiro (PMERJ) con el apoyo técnico del Laboratorio de Análisis de la Violencia (LAV-UERJ). El uso de la fuerza letal es elevado dentro de la PMERJ, institución que durante años aplicó un modelo de “guerra contra el crimen”.

El Índice de Aptitud tiene el objetivo de: 1) identificar policías que, por diversos motivos, no estén en condiciones adecuadas para ejercer una función policial que implique la posibilidad del uso de la fuerza, y 2) generar incentivos para la reducción en el uso de la fuerza.

lombia. Esta iniciativa, sin embargo, no fue incluida en el *Mapeo de programas de prevención de homicidios* (Cano y Rojido, 2016) por haber sido publicada después de dicho relevamiento.

⁴⁶ Más información disponible en: <http://www.monitorfuerzaletal.com>.

Estrategia V. Reinserción, mediación o negociación dirigida a victimarios. Esta estrategia aborda a los victimarios reales o potenciales como interlocutores y no simplemente como objeto de la represión del aparato del Estado. Por ese mismo motivo, estas son las iniciativas más controvertidas.

Las dimensiones contempladas en el Índice son: 1) aptitud física; 2) aptitud psicológica; 3) aptitud técnica; y 4) histórico de uso de la fuerza letal, medido a través del consumo mensual de municiones. De esta forma, los y las agentes podrían tener instancias de rehabilitación física, psicológica o técnica en función de las necesidades específicas de cada caso. Asimismo, los y las policías que presentaran un consumo mensual de municiones superior al esperado en razón de su función y del local en que trabajan, podrían tener limitado temporalmente el acceso al arma de fuego y ser reubicados, también temporalmente, en funciones que no prevean el uso de la fuerza.

Tipo 11: Reinserción o rehabilitación de autores de violencia armada (4 programas)

El objetivo es reinsertar en la sociedad a personas que se involucraron en violencia armada, particularmente las que pertenecen a grupos armados, promoviendo su incorporación a actividades educativas, laborales o recreativas que les ofrezcan una alternativa a la vida anterior. En esta línea, se puede mencionar el Proyecto Alcatraz en Venezuela, en el que una empresa privada ofrece una vida alternativa, a través del deporte, la formación y el trabajo, a jóvenes que integran bandas criminales de la zona.

Dado que algunos jóvenes podrían tener pendientes causas penales, estos programas son controvertidos y delicados desde el punto de vista político y jurídico. Por otro lado, considerando la baja capacidad de resocialización de la gran mayoría de los sistemas penitenciarios de la región, si este tipo de programas tiene éxito se podría alcanzar, potencialmente, un resultado más favorable para la sociedad que el obtenido por el procesamiento criminal tradicional.

Proyecto Alcatraz, Venezuela

El Proyecto Alcatraz es una iniciativa impulsada y financiada desde 2003 por una empresa privada: la Compañía Anónima de Ron Santa Teresa. Su principal objetivo es reducir los homicidios y para ello se reclutan integrantes de bandas delictivas o bandas completas en las localidades más violentas del municipio José Rafael Revenga, en el Estado de Aragua. Allí los jóvenes son llevados a convivir y compartir trabajo intensivo. El proyecto combina actividades de capacitación laboral, la práctica del rugby, formación en valores, asistencia psicológica, apoyo familiar, justicia restaurativa y trabajo comunitario. Al momento del egreso se buscan oportunidades laborales para los jóvenes y algunos incluso pasan a trabajar en la propia Compañía Anónima de Ron Santa Teresa.

Tipo 12: Mediación con grupos armados (13 programas)

En este tipo de intervenciones la interlocución no se realiza con individuos, sino con grupos armados en su conjunto, porque el objetivo no es desmovilizar algunas personas, sino obtener un cambio de conducta del grupo como un todo. Los protagonistas de este diálogo pueden ser miembros de la sociedad civil o el propio poder público, en cuyo caso los riesgos políticos del proceso son elevados, considerando la legitimación indirecta que el Estado puede prestar a grupos presuntamente criminales al entrar en comunicación con ellos. Por otro lado, este abordaje tiene la capacidad potencial de reducir los homicidios de forma muy intensa y rápida, especialmente en los contextos en que la violencia letal está vinculada a la actuación de grupos armados.

El ejemplo más común de este tipo es el programa Cure Violence, inspirado en la experiencia del Ceasefire en Chicago e implementado por organizaciones de la sociedad civil en un mínimo de nueve países de la región, generalmente con financiamiento de organizaciones internacionales como el BID o de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID, por su sigla en inglés). Esta intervención que ya había sido evaluada en contextos diferentes al latinoamericano, fue recientemente analizada en Trinidad y Tobago, con resultados que apuntan un impacto significativo en la reducción de los homicidios y otros delitos asociados a la violencia armada (Maguire, Oakley y Corsaro, 2018).

Por otro lado, en Venezuela un grupo de madres de miembros de bandas cumple esta función de negociación para reducir la violencia entre ellas, mediante las Comisiones de Convivencia de Catuche.

Peace Management Initiative, Jamaica

Peace Management Initiative (PMI) (Iniciativa de Gestión de la Paz) es una organización de la sociedad civil que desde el año 2002 implementa una intervención que lleva el mismo nombre en comunidades de Kingston, St. Catherine y Clarendon, en Jamaica. La experiencia de PMI está centrada en la idea de hacer cesar, a través de la mediación, los ciclos de venganza en los conflictos de grupos de jóvenes que habitan áreas periféricas del país y controlan estos territorios. Cuando se produce la muerte de un miembro de pandillas, los “interruptores” del programa entran inmediatamente en contacto para intentar evitar que se produzca una retaliación.

Además de su labor de mediación que resulta en treguas y “acuerdos de paz”, PMI se ha dedicado progresivamente a desarrollar acciones de apoyo a las víctimas y enfrentamiento del trauma, así como iniciativas de reinserción social que les ofrezcan alternativas a los jóvenes para que renuncien a la violencia.

Tipo 13: Negociación con grupos armados (1 programa)

En este caso, no se trata de una mediación del conflicto entre grupos armados, sino de una negociación entre representantes de estos grupos y del Estado, con el objetivo de reducir la violencia. El único caso público conocido de negociación con grupos armados es la llamada Tregua de las Maras, de El Salvador en 2012, en que el gobierno mejoró las condiciones penitenciarias de líderes de las pandillas y permitió el contacto con sus respectivas organizaciones a cambio de un compromiso de los grupos para lograr una reducción de los homicidios entre ellos y contra los agentes públicos.

Los riesgos políticos de este tipo de proceso son elevados, entre ellos la legitimación que el Estado puede otorgar a grupos criminales y la posibilidad de que estos últimos se fortalezcan a lo largo del proceso, incluso “administrando” los homicidios como forma de negociar con el poder público. Por otro lado, el caso de El Salvador también ejemplifica el enorme potencial de estas estrategias para evitar la pérdida de vidas humanas a corto plazo, especialmente en contextos donde pocos grupos armados son responsables de una gran parte de la violencia.

La Tregua entre Maras, El Salvador

Entre marzo de 2012 y mayo de 2014 El Salvador vivió una experiencia inusitada en la que los líderes presos de las principales pandillas (la Mara Salvatrucha o La MS13 y el Barrio 18 o La 18) se comprometieron a reducir los homicidios a través de acuerdos en los que renunciaban a ataques mutuos y a ataques contra las fuerzas del orden, además de cesar el reclutamiento obligatorio de personas y respetar los centros educativos. A partir de ahí, los dirigentes de esos grupos fueron transferidos a centros penales menos rigurosos y recibieron diversos beneficios penitenciarios, evidenciando el papel del Estado en el proceso, por más que el gobierno insistiera en que no negoció con grupos criminales. Posteriormente y más allá de la tregua, se declararon “municipios libres de violencia”, lo que abrió la puerta para diversos acuerdos locales que incluyeron programas sociales y económicos para los pandilleros y sus familias. A raíz de esos acuerdos, los homicidios en el país bajaron rápidamente a un tercio del nivel anterior y se mantuvieron de esa forma durante muchos meses, evidenciando la capacidad de las pandillas y de sus liderazgos para contener la violencia. Una vez rota la tregua, los homicidios aumentaron de forma abrupta y pronunciada.

Tipo 14: Planes de prevención de la violencia que incorporan un componente de homicidio (22 programas)

Estas intervenciones, comunes en la región, incluyen distintas acciones convergentes para reducir la violencia o promover la seguridad. Para que sean incluidas en el presente estudio precisan de un componente específico de reducción de homicidios o una meta en este sentido. Un ejemplo muy conocido es Todos Somos Juárez, en México, programa que implementó una amplia serie de medidas intersectoriales con el objetivo de superar la inseguridad y, entre otras cosas, reducir la elevadísima tasa de homicidios que la ciudad enfrentaba en 2009 y 2010. Algunas de estas intervenciones habrían tenido impacto sobre los homicidios, como la Iniciativa de Seguridad para América Central (CARSI) (Berk-Seligson *et al.*, 2014).

Citizen Security Programme, Trinidad y Tobago

Este programa, ejecutado por el Ministerio de Seguridad Nacional a partir del año 2007, busca contribuir a la reducción del crimen y la violencia en 19 comunidades de alta incidencia. Sus objetivos específicos son: 1) reducir los niveles de homicidios, robos y lesiones, 2) aumentar la percepción de seguridad, 3) reducir lesiones relacionadas con armas de fuego, maltrato infantil, violencia doméstica y violencia juvenil, y 4) aumentar la eficacia colectiva para prevenir la violencia en las comunidades asociadas.

La metodología de intervención se basa en tres pilares que combinan la acción comunitaria, la asistencia al servicio de policía y el fortalecimiento institucional del Ministerio de Seguridad Nacional. Con todo, la acción comunitaria es quizás el componente central del programa, que busca movilizar actores locales y diversas organizaciones no gubernamentales presentes en el país.

Estrategia VI. Estrategias integradas de reducción de la violencia letal. Estas iniciativas incorporan varias acciones de naturaleza diferente con el objetivo de prevenir la violencia letal. Al incluir intervenciones diversas estos programas suelen ser muy difíciles de evaluar, por el desafío de calibrar el impacto de cada una de las acciones que los componen y la práctica imposibilidad de generar grupos de control. En algunos casos, los proyectos contemplan la prevención general de la violencia, pero con un componente de homicidios. En otros, la estrategia en sí fue pensada específicamente para disminuir la violencia letal.

Tipo 15: Estrategias integradas para la reducción de los homicidios (3 programas)

A diferencia del tipo anterior, en el que la reducción de homicidios es apenas un componente, en este caso se contemplan estrategias que reúnen acciones diversas diseñadas específicamente con la intención de reducir los homicidios. Un ejemplo al respecto es la Estrategia Nacional contra el Homicidio, formulada por la Policía Nacional de Colombia en 2010. A pesar del breve tiempo en que estuvo en vigor y de la amplia variedad de medidas que incluyó, todo lo cual dificulta evaluar en qué medida tuvo un impacto, no deja de constituir una reflexión interesante sobre cómo una institución policial puede repensar su forma de actuar con el fin específico de contener los homicidios. En suma, el grado de focalización en la reducción de homicidios es máximo en este tipo de abordajes, aunque permanece la dificultad de evaluar el impacto de medidas simultáneas de diferente naturaleza.

Estrategia Nacional contra el Homicidio, Colombia

La Estrategia Nacional contra el Homicidio surgió en el año 2010 como parte de la Estrategia para la Consolidación de la Seguridad Ciudadana de la Policía Nacional. Más que una intervención concreta, la Estrategia Nacional contra el Homicidio representó una “caja de herramientas” destinada a mejorar el desempeño de la policía en el control de los homicidios.

A nivel táctico y operacional esta iniciativa contó con diversos componentes, a saber: intervención policial, interagencial, normativo, cultura ciudadana, participación ciudadana, gerencia de recursos, medios de comunicación, Sistema Penal Oral Acusatorio (SPOA) y evaluación. Estos componentes fueron combinados y aplicados de modo variable en el territorio, dado que en Colombia los alcaldes tienen autonomía para establecer prioridades en el trabajo policial, definir sus planes locales de seguridad y asignar los recursos económicos para desarrollarlos.

A continuación, se ilustra la frecuencia de los programas de prevención de homicidios en América Latina y el Caribe, según estrategia y tipo de intervención.

Apenas 18 (16,5%) programas de prevención de homicidios implementados en la región fueron objeto de al menos una evaluación de impacto. Algunas de estas intervenciones han sido evaluadas en más de una oportunidad, aunque esta es la excepción y no la regla. Lo más frecuente es que los programas cuenten con documentos internos de análisis e informes para el monitoreo de indicadores, pero no con evaluaciones rigurosas que permitan saber si los esfuerzos emprendidos brindaron o no los resultados esperados.

La instauración de una rutina de evaluación en la región es fundamental para desarrollar intervenciones eficaces y eficientes. De hecho, la gran mayoría de la literatura científica sobre prevención de la violencia se refiere a los Estados Unidos y a países de Europa, sociedades con bajas tasas de homicidio. En este sentido, es urgente desarrollar más investigación en lugares donde la violencia se halla más concentrada (Krisch *et al.*, 2015). De la misma forma, la adaptación de experiencias internacionales a la realidad de América Latina y el Caribe también debe pasar por la realización de evaluaciones de estas intervenciones en el contexto regional.

Esquema 3. Programas de prevención de homicidios en América Latina y el Caribe según estrategia y tipo de intervención



Fuente: [Homicidios Latam](#) (consultado en marzo del 2020).

Programas de prevención del homicidio

Frente a la gravedad y singularidad de los homicidios en América Latina y el Caribe cabría esperar una proliferación de políticas públicas específicamente destinadas a enfrentar el problema. Sin embargo, no es esto lo que ha sucedido hasta ahora.

Prevenir es reconocer el riesgo de que un evento acontezca y realizar alguna acción para eliminarlo o reducirlo. Por lo tanto, prevenir los homicidios implica anticiparse y evitar estos hechos.

Las iniciativas de prevención pueden ser categorizadas en función del tipo de población a las que se destinan. La prevención primaria o universal se destina a todos los individuos de una sociedad, la prevención secundaria o selectiva está centrada en grupos con alto riesgo de convertirse en víctimas o perpetradores de la violencia y la prevención terciaria o señalada se concentra en individuos que ya

RESUMEN

son víctimas o autores de la violencia para evitar la revictimización o la reincidencia.

Otro criterio utilizado para clasificar las iniciativas de prevención es su naturaleza. La prevención social busca intervenir sobre causas sociales para disminuir la probabilidad de que las personas se envuelvan en comportamientos violentos, la prevención situacional se compone de medidas orientadas a la reducción de oportunidades y la prevención policial consiste en estrategias desarrolladas por las instituciones policiales para reducir la violencia criminal.

Un estudio reciente se propuso identificar y analizar los programas de prevención de homicidios existentes en América Latina y el Caribe (Cano y Rojido, 2016). A pesar del número reducido (109), estas iniciativas destacan por su variedad. Los programas fueron clasificados en 15 tipos agrupados en 6 estrategias o áreas temáticas: a) control de factores de protección o de riesgo; b) promoción de cambios culturales; c) protección a grupos de riesgo; d) mejoras en el funcionamiento del sistema de justicia criminal; e) programas de reinserción, mediación o negociación dirigidos a victimarias; y f) estrategias integradas de reducción de la violencia.

Menos de un quinto de los programas de prevención de homicidios implementados en la región fueron objeto de al menos una evaluación de impacto. La instauración de una rutina de evaluación en la región es fundamental para desarrollar intervenciones eficaces y eficientes.

CAPÍTULO 3. POR QUÉ ES NECESARIO EVALUAR

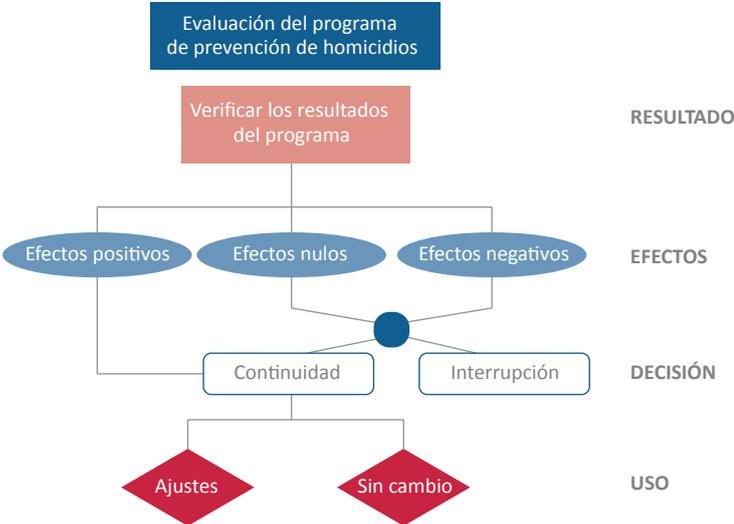
3.1 Utilidades de la evaluación

En general, los gestores y gestoras tienen una impresión positiva del impacto de los programas, incluso antes de que se implementen. A menudo, creen que basta ejecutar las medidas previstas para que los efectos se den naturalmente. Esta impresión, sin embargo, se suele basar en percepciones subjetivas afectadas por el interés profesional y personal, por lo que no garantiza que los programas sean exitosos. En este sentido, es importante resaltar que la evaluación es la única forma de conocer los efectos reales de un programa (ver apartado 4.1). No es posible saber concretamente si los programas han tenido resultado sin una evaluación metodológicamente válida e independiente, que intente aislar el impacto del programa de todos los otros elementos que pueden afectar los homicidios.

La evaluación representa un potente instrumento de gestión en la medida en que puede subsidiar las decisiones sobre la continuidad o interrupción de los programas y los ajustes a ser adoptados (Esquema 4). También puede analizar si hay efectos inesperados, diferentes de los previstos por el proyecto. Estos efectos imprevistos pueden ser positivos o indiferentes, pero también pueden ser perversos, tal que no compensen los éxitos obtenidos en otras dimensiones.

Si la evaluación arroja resultados positivos, es más fácil justificar la continuidad del programa, conseguir nuevos recursos o persuadir a otras instituciones para que se involucren en proyectos similares.

Esquema 4. Evaluación como instrumento de gestión



Fuente: Elaboración propia.

En suma, la evaluación debe identificar los efectos del programa para subsidiar sus decisiones.

Entre las funciones, complementarias e interdependientes que la evaluación desempeña, se pueden mencionar las siguientes:

- Comprobar si el programa se ha implementado tal como fue diseñado. En muchas ocasiones, los programas acaban siendo ejecutados de forma muy diferente a como fueron pensados, por dificultades del contexto o por diferencia de criterios entre los gestores locales.
- Medir en qué medida el programa ha alcanzado los resultados pretendidos. Por ejemplo, si la intervención limitó los horarios de venta de alcohol para reducir los homicidios nocturnos en locales cercanos a los bares, hay que comprobar si, de hecho, han disminuido los homicidios y si esta disminución ha sido resultado del programa.
- Analizar los resultados no previstos de la intervención y sus efectos colaterales, por ejemplo, la migración de los homicidios a otras áreas cercanas a las que fueron objeto de intervención.
- Calcular el costo de la iniciativa y comparar los beneficios con los costos, puesto que el programa podría haber reducido los homicidios, pero a un costo muy elevado, superior a otras alternativas existentes.
- Entender los obstáculos encontrados en la aplicación del programa, para mejorar su funcionamiento futuro.
- Explorar la percepción que los diferentes actores poseen sobre el programa y cómo influyó en la implementación y los resultados.
- De una forma general, ayudar a rendir cuentas sobre la intervención.

Para evaluar un programa no es suficiente decir que un servicio se proporcionó correctamente y que los usuarios están satisfechos. Si el objetivo era reducir los homicidios, hay que verificar si esto ha sucedido y, mucho más difícil, si esta reducción ha sido provocada por el programa.

Como regla, el resultado de la evaluación es relevante no solo para gestores y gestoras, sino para diversos actores sociales (en inglés, el término usado es *stakeholder*) que tienen un interés en los resultados y en las decisiones relativas al programa. Por tanto, evaluadores y evaluadoras deben pensar a priori en quiénes son esos actores, qué intereses u objetivos tienen, cómo les serán comunicados los resultados

de la evaluación y cómo maximizar la posibilidad de que los usen en sus respectivas tomas de decisión.

3.2 Qué riesgos existen si no se evalúa

Si no se realizan evaluaciones de calidad, las decisiones serán, básicamente, tomadas a ciegas.

Entre los riesgos que se corren, caben mencionar los siguientes:

- a) Continuar gastando recursos en programas que no reducen los homicidios, solo porque gestores y gestoras o quienes toman las decisiones están convencidos de lo contrario. Es común, por ejemplo, que una reducción de homicidios sea automáticamente atribuida al programa en vigor, pero si la atribución no es válida, seguirán empeñándose recursos y la reducción probablemente no se repetirá.
- b) Interrumpir un programa que salva vidas, porque carece de apoyo político. Este es el riesgo opuesto al anterior y probablemente aún más grave. Es posible, inclusive, que un programa haya conseguido disminuir los homicidios, pero que esto no sea aparente porque otros factores contextuales, políticos o económicos, hayan disparado la violencia. En estos casos, en ausencia de evaluación, el programa que salva vidas será interrumpido, porque se desconoce su impacto.
- c) Ocurrió una reducción de la violencia letal en los territorios intervenidos, pero esta no se debió al programa implantado, sino a otros que ya estaban en vigor anteriormente. Sin evaluación, el mérito será otorgado al programa ineficaz, que será perpetuado en detrimento del resto.
- d) El impacto del programa puede haber sido positivo en un principio, pero su capacidad de reducir los homicidios puede irse perdiendo a lo largo del tiempo, por diversos motivos. Solo una evaluación planeada para detectar la evolución de los efectos en el tiempo puede evidenciar este problema.
- e) El programa estaba bien orientado y podría haber producido resultados, pero algo cambió en el contexto externo o surgieron resistencias imprevistas que impidieron su eficacia. Una evaluación adecuada debería identificar estos factores que frustraron el éxito, para que la iniciativa tenga una nueva oportunidad y no sea descartada prematuramente.
- f) El programa puede haberse llevado a cabo de una forma muy diferente a lo previsto, de forma que, si no hay una evaluación que lo detecte, cualquier juicio sobre el efecto del proyecto será atribuido a otra intervención que no fue la aplicada, de

manera que las conclusiones extraídas serán equivocadas. Por ejemplo, no sabremos si la ausencia de resultados se debe a que el programa no tuvo realmente impacto o a que fue aplicado de forma equivocada (problema de integridad de la intervención). También es común que un programa sea ejecutado de una forma, pero vaya variando en la medida que avanza. Las consecuencias de no ser conscientes de ello son las mismas antes explicadas.

- g) El costo de la intervención subió mucho más de lo programado, de manera que, a pesar de ser eficaz, la iniciativa dejó de ser eficiente y no debería ser extendida, de esa forma, al conjunto del territorio.
- h) La aparición de resultados contrarios a los propósitos de la intervención o la incidencia de efectos colaterales se conocen como efectos perversos. Los efectos perversos pueden ocurrir con el grupo en el que se ejecuta el programa o en otros grupos que no están expuestos a la intervención. Una evaluación metodológicamente adecuada puede determinar si algunos tipos de resultados se deben a los efectos perversos de la intervención o si fueron causados por algún otro fenómeno.

Los efectos de los programas de prevención de violencia doméstica en el asesinato de parejas íntimas en 48 ciudades de los Estados Unidos*

La provisión de servicios públicos para mujeres que quieren abandonar relaciones abusivas puede, a pesar de las buenas intenciones, aumentar a veces el riesgo de femicidio íntimo cuando la protección no es brindada de forma adecuada. Un estudio analizó la relación entre el homicidio de una pareja íntima y las iniciativas para prevenir la violencia doméstica en 48 ciudades grandes de los Estados Unidos entre 1976 y 1996. La variable dependiente fue el recuento de homicidios de parejas íntimas divididas por el sexo, la raza (afroamericano, blanca, total) y relación marital con el agresor. Como variables independientes se incluyeron: recursos legales para prevenir la violencia doméstica, políticas locales, líneas telefónicas de apoyo legal a víctimas, beneficios estatales para familias con hijos dependientes, tasa de matrimonio y divorcio y tasa de homicidio para hombres adultos (25 años o más).

Los resultados del estudio mostraron que, controlando las diferentes influencias, varios tipos de iniciativas de prevención están vinculadas a niveles más bajos de homicidio por parte de una pareja íntima. Sin embargo, los resultados mostraron algunos efectos perversos. Así, cuando los fiscales adoptaron políticas que demostraron su disposición a procesar

criminalmente a los infractores de órdenes de protección, hubo un aumento en el homicidio de mujeres blancas, casadas y solteras, y de hombres afroamericanos solteros. Este resultado sugiere que estar dispuesto a demandar al agresor sin proporcionar la protección adecuada a la víctima puede generar impulsos vengativos que aumenten la victimización.

*Dugan, Nagin & Rosenfeld, 2003: 169-198.

Utilidades de la evaluación

La evaluación de los programas de prevención de homicidios es una herramienta de gestión importante para fundamentar las decisiones sobre: la continuidad del programa, los ajustes a adoptar y la interrupción del programa.

Es necesario analizar si hay efectos inesperados, diferentes a los pronosticados por el proyecto.

Si no se llevan a cabo evaluaciones de calidad, se corre el riesgo de prolongar programas inocuos o incluso contraproducentes y de interrumpir iniciativas que podrían salvar vidas.

RESUMEN

CAPÍTULO 4. ¿QUÉ ES UNA EVALUACIÓN DE PROGRAMAS?

4.1 Concepto: qué es una evaluación de programas

En primer lugar, vale la pena definir qué es un *programa social*. Un programa es una intervención planeada para modificar alguna dimensión de la realidad social a partir de un determinado momento del tiempo. Normalmente, los programas tienen un inicio y también un fin, aunque excepcionalmente hay programas que nacen con una vocación de permanencia. Un programa se diferencia de una política, un conjunto mayor de intervenciones que puede incluir diversos programas, no solo por su mayor especificidad, sino, justamente, porque los programas son acotados y tienen un momento de inicio, desarrollo y final, mientras que las políticas públicas pueden ser sostenidas en el tiempo. Y ese corte temporal es lo que permite comparar la situación anterior a la introducción del programa con la situación posterior, algo que será esencial para la evaluación. En ese sentido, la política de seguridad suele enfrentar el fenómeno de los homicidios como un crimen de extrema gravedad, merecedor incluso de una acción prioritaria, pero eso no la convierte en un programa. Un programa de reducción o prevención de homicidios es una intervención planeada específicamente para disminuir la incidencia de homicidios en una sociedad.

La *evaluación* (del latín *valere*) de un programa es un examen detallado y sistemático que permite atribuir un valor, positivo o negativo, a dicho programa. En otras palabras, la evaluación hace posible saber en qué medida un programa social es un éxito o un fracaso, para que puedan tomarse decisiones sobre su continuidad o su interrupción, y sobre alteraciones que puedan ser necesarias. La evaluación es más que nada una investigación, que debe revelar, entre otras cosas, el modo en que el programa ha sido implementado, cuáles han sido sus efectos y, si es posible, las causas de su éxito o fracaso. En nuestro caso concreto, la evaluación debe decirnos si el programa en cuestión vale o no la pena como estrategia para reducir los homicidios. A diferencia de la evaluación, que atribuye mérito al programa, la auditoría tiene como objetivo verificar la adhesión de un programa a un conjunto de procesos previamente definidos o de gastos preestablecidos.

Como cualquier investigación científica, la evaluación debe ser aplicada de forma neutra e impersonal, con base en la evidencia obtenida y sin prejuzgar los resultados. Las evaluaciones de programas no evalúan personas ni grupos políticos o sociales, sino intervenciones. En ese sentido, no pueden ser nunca pensadas como un ejercicio contra

o a favor de determinadas personas. De la misma forma, la investigación evaluativa debe ser conducida de forma ética y respetuosa con todos los sectores involucrados ([ver Capítulo 7](#)).

Como ya fue anticipado, aplicar programas sin evaluarlos es como conducir a ciegas. Por eso, los formuladores de los programas y las agencias financiadoras deben contemplar evaluaciones metodológicamente rigurosas junto con el diseño de las intervenciones. Por su parte, la sociedad civil tiene que exigir que los programas sean evaluados y que el resultado de estas evaluaciones sea difundido ampliamente, si se quiere reducir la violencia letal.

Sin embargo, a pesar de las profundas razones que justifican su existencia, las evaluaciones suscitan recelos comprensibles en los y las protagonistas de los programas y en las personas que tienen un interés sobre ellos. La posibilidad de que la evaluación arroje una conclusión negativa sobre el programa provoca la natural ansiedad de gestoras y gestores, personal técnico y responsables políticos. Todo ello generará resistencias, cuya superación constituye uno de los desafíos del proceso evaluativo.

Es casi una regla el hecho de que la evaluación será tanto mejor cuanto antes sea planeada. Lo ideal, por razones que serán mejor explicadas en el [Capítulo 6](#), es que sea diseñada en conjunto con la propia intervención. Esto permitirá recoger desde el principio las informaciones necesarias y crear los sistemas de información adecuados. Cuando la evaluación se piensa una vez que el programa ya comenzó o, peor aún, cuando el programa ya acabó, muchas informaciones relevantes pueden haberse perdido y las opciones metodológicas se reducen drásticamente.

De la misma forma que sucede con los programas, las evaluaciones están sujetas a diversas limitaciones políticas, económicas y sociales, de modo que deben ser pensadas de una forma realista y ajustada a su contexto y a los recursos disponibles.

4.2 Tipos de evaluación: de implementación, de impacto y de costo-beneficio

Existen tres tipos principales de evaluación: de implementación, de impacto y de costo-beneficio. Cada uno de ellas responde a una pregunta distinta, pero las tres son acumulativas, en el sentido de que cada una se suele construir sobre el resultado de la anterior. En primer lugar, se suele realizar una evaluación de implementación. Una vez que se tiene el resultado de esta, se aborda una evaluación de impacto. No es imposible realizar una evaluación de impacto sin haber realizado antes la de implementación, pero, en este caso, puede que no se sepa exactamente

en qué consistió el programa cuyo impacto se está evaluando. Por último, una vez que se ha determinado el impacto, se puede pensar en una evaluación de costo-beneficio. No tendría sentido iniciar una evaluación de este tipo sin saber antes si hubo algún beneficio y sin poder dimensionarlo.

4.2.1 Evaluación de implementación

Este tipo de evaluación responde a la pregunta de si el programa fue o no implementado según lo previsto. Esta es una cuestión central, puesto que antes de plantearse la eficacia de una intervención, se debe tener seguridad de qué es exactamente esa intervención que será evaluada. Es muy común que un programa sea aplicado de una forma diferente de lo planeado, ya sea por deficiencias en la gestión o por dificultades sobrevenidas. En programas nacionales o de amplia extensión, no es raro que las decisiones que los aplicadores toman a nivel local introduzcan un nivel de heterogeneidad significativo, transformando lo que era un proyecto único en un abanico de acciones diferentes.

Por todo ello, se suele iniciar el proceso evaluativo con una evaluación de implementación que permita conocer exactamente cómo se desarrolló el programa en la realidad, comparando el plan original con lo que efectivamente se puso en práctica. Para eso, es necesario verificar los registros existentes, recabar informaciones de beneficiarios y beneficiarias, técnicos y técnicas, y procurar, si existen, fuentes alternativas a las del propio programa. A través de esta información, es preciso saber si los medios empeñados en la intervención corresponden a los que fueron pensados inicialmente en cantidad y calidad, y si la temporalidad de esos medios coincide con la prevista. De la misma forma, hay que comprobar si la selección de beneficiarios y beneficiarias o de las áreas contempladas siguió los criterios prescritos o si hubo algún desvío.

Pensemos en un programa que intenta frenar los homicidios restringiendo el porte de armas en algunos municipios en algunas fechas concretas de mayor riesgo. Para evaluar su implementación, no basta saber la fecha en que la nueva normativa entró en vigor, sino que es necesario conocer, entre otras cosas, qué tipo de publicidad se le dio a la norma, cuántos puntos de fiscalización se establecieron en la ciudad, cuántas armas fueron incautadas en estas fechas, cómo se comparan con las incautaciones en otras fechas, etc. Si, por otro lado, un proyecto para reducir homicidios está basado en nuevas estrategias de patrullaje policial en áreas de alto riesgo, es preciso conocer cuántos policiales patrullaron, durante cuánto tiempo, exactamente qué áreas fueron objeto de este patrullaje y cómo fueron elegidas. Todo ello debe ser comparado

con el patrullaje que se realizaba anteriormente a la implantación del programa. Un tercer ejemplo son los programas de prevención social para jóvenes en riesgo de sufrir o cometer violencia. En este caso, hay que prestar especial atención no solo a cómo fueron ofrecidos los servicios a los jóvenes, sino a cómo esos jóvenes fueron seleccionados.

Obsérvese que nada de esto garantiza ni tampoco impide que se haya producido un impacto. Es decir, el programa puede haberse implementado exactamente como fue ideado y, aun así, no provocar resultados en su dimensión final, los homicidios. Lo contrario también es posible. La aplicación del programa puede distanciarse sobremedida del plan original y, sin embargo, puede haber tenido éxito en sus propósitos. De forma sucinta, la evaluación de implementación se ocupa de cómo los medios han sido aplicados, mientras que la de impacto lidiará con los fines, esto es, con los resultados.

Además de garantizar lo que se conoce como integridad del programa, esto es, si fue ejecutado tal como fue diseñado, la evaluación de implementación puede y debe proporcionar información relevante sobre las barreras y los facilitadores de la implementación.

4.2.2 Evaluación de impacto

La evaluación de impacto determina cuál fue el efecto que el programa tuvo sobre la realidad. Para ello, es preciso compararla con una situación hipotética o contrafactual, es decir, estimar qué hubiera ocurrido si el programa no hubiera existido.

Nunca está de más insistir en que el hecho de registrar los resultados pretendidos, una reducción de los homicidios, no garantiza que esto se deba a la intervención, pues podría ser fruto de muchos otros factores.

La “Donbernabilidad” en Medellín

En 2003, el gobierno colombiano negoció un acuerdo con los paramilitares para desarmarlos y desmovilizarlos a cambio de beneficios legales. El Bloque Cacique Nutibara (BCN), dirigido por el ex comandante paramilitar conocido como Don Berna, fue la primera unidad en desmovilizarse. Sin embargo, la desmovilización del BCN se consideró una estrategia para limpiar los antecedentes penales de algunos de los miembros más prominentes de la criminalidad en Medellín, que se unieron al grupo a principios de 2000, durante la adquisición de franquicias paramilitares. Asimismo, algunos miembros de pandillas bajo el control de la Oficina de Envigado (un plan dirigido por Don Berna para regular los Mercados Ilícitos) fueron formalmente “desmovilizados” en ceremonias simbólicas y públicas (Sandoval, 2017).

El BCN estaba formado por una extensa red de grupos criminales, que nunca se dismanteló durante la mediación y esta organización paramilitar terminó jugando un doble papel como actor «desmovilizado» y hegemónico en las estructuras ilegales de la ciudad. En otras palabras, la Oficina de Envigado fue tan eficiente que Don Berna logró controlar las tasas de asesinatos en Medellín (Sandoval, 2017).

Después de este acuerdo, la tasa de homicidios en Medellín se redujo de 184 por 100.000 en 2002 a 98.2 por 100.000 habitantes en 2003. Y después de la desmovilización total del grupo la tasa de homicidios disminuyó a solo 35 muertes por cada 100.000 habitantes por tres años consecutivos (2005-2007). Este período se conoció como “Donbernabilidad”, en vista de la consolidación de la dominación de Don Berna sobre la ciudad. Sin embargo, en 2008 Don Berna fue extraditado a los Estados Unidos y eso supuso el comienzo de un estallido de violencia entre facciones criminales que intentaron llenar el vacío de poder.

El aumento del presupuesto asignado a la justicia y la seguridad y las operaciones policiales contra las facciones criminales se entendió que, seguramente, tuvieron algún impacto, pero muchas personas están convencidas de que la reducción de los homicidios en Medellín no se debió al gobierno, sino, posiblemente, a la hegemonía de la estructura criminal liderada por Don Berna.

La evaluación de impacto es la más importante de todas, en la medida en que determina la eficacia del programa, sin la cual la propia intervención carece de sentido. Una estrategia para reducir la violencia letal puede ser bien ejecutada y requerir de pocos recursos, pero no sirve de nada si no consigue disminuir los homicidios. La evaluación de impacto es, por definición, una investigación causal, pues estima en qué medida el programa fue el causante de los efectos observados. Como la causa no puede ser observada sino inferida, este tipo de investigaciones requiere de una cuidadosa estrategia metodológica que será explicada más adelante ([ver Capítulo 6](#)).

La evaluación de impacto debe estudiar los efectos, en primer lugar, sobre las dimensiones que se pretendió alterar. En otras palabras, es preciso confrontar el efecto del programa con los objetivos que este se marcó al inicio. En nuestro caso, el impacto sobre los homicidios será siempre el indicador central, aunque pueden incluirse otras dimensiones intermediarias, según el caso (por ejemplo, el esclarecimiento penal, los conflictos violentos o el porte de armas).

En segundo lugar, este tipo de evaluaciones debe procurar efectos imprevistos o no deseados sobre otras dimensiones. Algunos de ellos pueden ser inocuos, mientras otros pueden ser francamente negativos. Por ejemplo, un programa de patrullaje policial puede acabar sometiendo

a los miembros de las minorías a un control selectivo y discriminatorio (*racial profiling*) que viole los derechos humanos y erosione la legitimidad policial. En ese sentido, es importante recordar que la evaluación es relevante para un conjunto de actores sociales (*stakeholders*) y que los impactos, especialmente los no intencionales, pueden ser diferentes para cada uno de ellos.

Es importante considerar el momento adecuado para una evaluación de impacto, de acuerdo con la temporalidad de los efectos pretendidos. Es común, especialmente en los proyectos de prevención social, que la evaluación se desarrolle algunos meses después de iniciado el programa, cuando la teoría del cambio ([ver apartado 5.3](#)) indicaría que serían necesarios años para que el impacto se manifieste. En el otro extremo, algunos impactos (como los de las prohibiciones ocasionales de porte de armas) son de corta duración, de modo que, si la evaluación no se hace inmediatamente después de la intervención, puede acabar no detectándolos.

Las evaluaciones que son realizadas durante la aplicación de un programa, para orientar un mejor funcionamiento, se denominan *formativas*. Y las que son aplicadas al final de un ciclo para hacer balance de un proyecto y decidir sobre su continuidad se llaman *sumativas*.

4.2.3 Evaluación de costo-beneficio

Una evaluación de costo-beneficio compara el valor o el beneficio de un programa con el costo que originó. El objetivo es determinar si el programa fue eficiente o si, por el contrario, hubiera sido posible obtener los resultados con un costo menor a través de otras intervenciones. Así, a diferencia de las evaluaciones de implementación y de impacto, la de costo-beneficio es siempre relativa, pues un programa será eficiente o ineficiente en comparación con las alternativas disponibles. De hecho, un programa que sea eficiente hoy puede dejar de serlo más adelante, cuando surjan otros de igual efecto y menor costo. La evaluación de impacto es un prerrequisito de la de costo-beneficio, pues solo tiene sentido plantearse el costo una vez que estamos seguros de que el programa consigue, al menos, algún resultado.

Los costos no precisan ser solamente económicos, sino que pueden ser sociales o de otro tipo. Los impactos no intencionales antes mencionados deben formar parte de este costo.

Se habla de evaluaciones de costo-beneficio propiamente dichas cuando los dos lados de la comparación (costos y beneficios) están monetizados, esto es, pueden ser medidos en dinero. En el caso de los homicidios, es posible pero complejo y bastante controvertido colocar

un valor a la vida humana (el beneficio). Por eso, en esta área es más común trabajar con evaluaciones de costo-utilidad, en las que la unidad del beneficio no está medida en dinero. Así, por ejemplo, se podrían calcular cuántos dólares costó evitar un homicidio y se compararían los programas según cuánto costase, en cada uno de ellos, cada homicidio que fue prevenido.

Tabla 1. Tipos de evaluación

Tipo de evaluación	Preguntas a ser respondidas	Utilidad
Evaluación de implementación	¿Se implementó el programa según lo previsto? ¿Qué tipos de problemas de implementación han surgido y cómo se pueden abordar?	Demuestra si los programas cumplen fielmente el modelo del programa
Evaluación de impacto	¿Cuál es el efecto del programa en la realidad? ¿Cómo logró el programa los resultados obtenidos?	Proporciona evidencia rigurosa sobre el efecto de un programa
Evaluación de costo-beneficio	¿Los beneficios del programa superan sus costos? ¿Es el programa más beneficioso que otros programas que abordan el mismo problema?	Compara el rendimiento de la inversión para varios programas que resuelven el mismo problema

Fuente: Elaboración propia.

4.2.4 Otros tipos de evaluación

Además de los tres tipos centrales ya descritos, hay otras clases de evaluación que pueden ser encontradas en la literatura.

Una *evaluación de diseño*, por ejemplo, examina la coherencia entre los objetivos, la metodología, el contexto, la teoría de cambio y los recursos para determinar, a priori, la viabilidad del proyecto y la probabilidad de que consiga sus objetivos. Se trata, en contraposición a los tres tipos anteriormente descritos, de una evaluación no empírica, sino teórica, *ex ante*, que debe basarse en la teoría, en la literatura existente y en la experiencia previa de evaluadores y evaluadoras. Por eso, debe realizarse en cuanto el proyecto haya sido diseñado y solo puede ser desarrollada por especialistas en el área.

Una *evaluación de necesidades* intenta medir las necesidades del público objetivo al que el proyecto pretende beneficiar y, a partir de

ahí, considera estas necesidades como objetivos, midiendo luego en qué medida han sido satisfechas por la intervención. En este sentido, es un tipo de evaluación de impacto en el que los objetivos no son definidos por el programa, sino por la población. En el caso específico de los programas de prevención de homicidios, la evaluación de necesidades no tiene demasiado espacio, porque el objetivo general del programa ya está predefinido: la reducción de los homicidios. Sin embargo, incluso cuando hay claridad en relación al objetivo central, puede haber diferencias en relación con los objetivos intermediarios o con el tipo de homicidio que se pretende reducir.

4.3 Monitoreo y evaluación

El monitoreo y la evaluación son conceptos próximos que a veces se usan casi de forma indistinta. En realidad, ambos términos se diferencian, sobre todo en el énfasis que conceden a unos u otros elementos.

El *monitoreo* puede ser definido como la recolección y el análisis sistemático y regular de informaciones sobre el andamio de un proyecto. El monitoreo pretende detectar problemas durante la aplicación del proyecto para poder tomar medidas correctivas. En general, se centra más en procesos que en resultados. Un sistema de alerta temprana, concepto común en epidemiología, encarna muy bien el espíritu del monitoreo.

Por su parte, la *evaluación* suele ser una iniciativa episódica, realizada al final de un ciclo, para determinar cómo fue la implementación, el impacto o el costo de un proyecto. En ese sentido, se focaliza más en resultados y menos en procesos. La evaluación está más preocupada con lo que sucedió hasta ahora, mientras que el monitoreo se concentra en el presente o el pasado inmediato.

El monitoreo no está distante de lo que es una evaluación formativa. De hecho, monitoreo y evaluación no son alternativos, sino complementarios. Un buen sistema de monitoreo es ya un gran paso para una evaluación, en la medida en que recoge informaciones que serán esenciales para esta última. Tanto el monitoreo como la evaluación utilizan indicadores y cobran pleno sentido en relación con objetivos y metas previamente definidos ([ver apartado 5.2](#)). No es imposible monitorear o evaluar sin metas, pero es mucho más complejo y los resultados serán menos sólidos.

El monitoreo suele ser realizado por el propio personal que ejecuta el proyecto, mientras que es común que las evaluaciones sean realizadas por personas externas, como forma de garantizar su neutralidad e independencia ([ver Capítulo 8](#)). El monitoreo suele estar centrado en la

implementación y no en el impacto, pero no es imposible que comience a registrar los efectos de la intervención. De forma general, la diferencia principal entre monitoreo y evaluación es la regularidad del primero y el carácter ocasional de la última.

Evaluación de programas

La evaluación de un programa es un examen detallado y sistemático que permite atribuir un valor, positivo o negativo, a dicho programa. La evaluación hace posible saber en qué medida un programa social es un éxito o un fracaso, para que puedan tomarse decisiones sobre su continuidad o su interrupción, y sobre alteraciones que puedan ser necesarias.

Tipos de evaluación:

Evaluación de implementación: responde a la pregunta de si el programa fue o no implementado según lo previsto.

Evaluación de impacto: determina cuáles fueron los efectos que el programa tuvo sobre la realidad.

Evaluación de costo-beneficio: compara el valor o el beneficio de un programa con el costo que originó y estima si sería posible obtener los mismos resultados con un costo menor por medio de otro tipo de intervenciones.

El monitoreo es un proceso regular de recolección y análisis de informaciones sobre el andamiaje de un proyecto para poder tomar medidas correctivas.

RESUMEN

CAPÍTULO 5. EVALUABILIDAD

Entre los diversos tipos de evaluación que se acaban de describir ([ver Capítulo 4](#)), el concepto de evaluabilidad está básicamente referido a las evaluaciones de impacto. Así, evaluabilidad, como el propio nombre sugiere, es el grado en que un programa es susceptible de ser evaluado de forma rigurosa para determinar su impacto (Wholey, 1994).⁴⁷

La reflexión sobre este concepto deriva justamente de la constatación de que algunos programas, por sus características o por el hecho de ya estar en curso, son difícilmente evaluables o imposibles de evaluar. Por un lado, es una señal de alerta para evaluadoras y evaluadores, de forma que estudien con cuidado la situación antes de decidir asumir la evaluación de un programa. Por otro, es un aviso para las entidades financiadoras y para los formuladores de las intervenciones, para que sus programas sean diseñados de forma que puedan ser evaluados adecuadamente.

Una idea implícita en el propio término es que hay situaciones en las que no se darían los requisitos mínimos para poder evaluar adecuadamente y, por tanto, sería mejor no evaluar hasta que esos requisitos fueran cumplidos.

5.1 ¿Cuáles son los requisitos mínimos para evaluar?

Para poder realizar una evaluación de impacto rigurosa es necesario que el programa cuente, en principio, con los siguientes elementos:

- a) Objetivos y, de preferencia, indicadores y metas claramente definidos.
- b) Un sistema de informaciones que permita recolectar datos precisos en distintos momentos: antes, durante y después de la intervención.
- c) Informaciones complementarias que permitan comparar otros grupos o individuos en situación similar a aquellos que experimentaron la intervención.

5.2 Objetivos, indicadores y metas del programa

Una evaluación mide el impacto de una intervención en función de sus *objetivos*, que deben estar establecidos en los documentos programáticos. Así, no son evaluadoras y evaluadores quienes deben escoger los objetivos, sino que estos deben estar definidos previamente por los formuladores del programa. Con todo, muchos programas no

⁴⁷ En el concepto originalmente defendido por Wholey, la plausibilidad de obtener un impacto y la posibilidad de utilizar los resultados de la evaluación serían también componentes de la evaluabilidad, pero estos elementos no serán incluidos en el concepto utilizado aquí. En ese sentido, la definición de Wholey incorpora lo que sería una “evaluación de diseño” ([ver apartado 4.2](#)).

tienen objetivos claros en sus documentos programáticos o simplemente carecen de este tipo de documento. En estos casos, el evaluador o la evaluadora debe indagar cuáles fueron los objetivos pretendidos por el programa, preguntando directamente a los formuladores o infiriéndolos a partir de los materiales e informaciones disponibles.

Existen iniciativas amplias dentro de las cuales la disminución de los homicidios es apenas uno de los objetivos específicos ([ver Capítulo 2](#)), como es el caso de los programas de seguridad ciudadana. Estos programas suponen mayores dificultades para la evaluación, en la medida en que plantean múltiples componentes y resultados a considerar. También hay programas con otros objetivos (como la mejora del patrullaje policial o la iluminación pública) que, sin embargo, podrían tener como consecuencia indirecta una reducción de los homicidios.

De forma general, las evaluaciones deben medir el grado de éxito en la consecución de los objetivos generales y específicos ([ver Capítulo 4](#)). Adicionalmente, es importante que se exploren los impactos no intencionales del programa, sean estos positivos o negativos, en dimensiones no anticipadas por sus formuladores.

En suma, la evaluación de impacto debe partir de los objetivos de una intervención. La evaluación será tanto más fácil metodológicamente y sus resultados serán tanto más sólidos en la medida en que los objetivos sean claros y estén formulados de manera precisa. Objetivos vagos o demasiado amplios dificultan una evaluación rigurosa. Cuando no hay un documento que especifique los objetivos de forma inequívoca, no es extraño que grupos diferentes dentro del programa posean visiones discordantes sobre cuáles son los objetivos centrales o sobre la prioridad de cada uno de ellos, lo que dificulta la tarea del equipo evaluador. En ese sentido, un programa en relación al cual los propios objetivos están en disputa es difícilmente evaluable, por lo que la existencia de objetivos es el primer requisito de la evaluabilidad.

En algunas ocasiones los programas alteran sus objetivos a partir de decisiones estratégicas claramente determinadas, mientras que en otros casos los modifican de forma gradual o tácita a lo largo del tiempo. Estos cambios afectan a la integridad programática y su identificación es fundamental para el proceso de evaluación, considerando, además, que los resultados esperados pueden variar en el tiempo. Situaciones de este tipo obligan a decidir si la evaluación será realizada en función de los objetivos iniciales o si será necesario establecer más de un conjunto de objetivos.

Del objetivo general y de los objetivos específicos deben derivarse *indicadores* que permitan estimar si se han logrado los resultados esperados. Los indicadores son variables operativas que sirven para medir, de forma indirecta, un fenómeno que no admite medición directa. La operacionalización es el mecanismo que permite pasar de conceptos abstractos (como la violencia) a aspectos observables (como la tasa de homicidios para cada 100.000 habitantes). La definición de indicadores es fundamental para la evaluación, dado que estos funcionan como medida de los objetivos del programa.⁴⁸ Por ese motivo, es deseable que los indicadores sean definidos por los formuladores del programa en la fase de diseño, junto con los objetivos.

Las *metas* traducen los objetivos en la práctica, ya que especifican qué valores de los indicadores deberían ser alcanzados para considerar el objetivo cumplido. Así, por ejemplo, una meta global sería la reducción del número de homicidios en 5% al año, durante 3 años, en la comunidad intervenida por el programa. Las metas deben ser verificables y realistas. Metas excesivamente ambiciosas llevarán el proyecto a ser percibido como un fracaso, incluso en situaciones en que su impacto haya sido considerable. Las metas deben ser determinadas en función del conocimiento teórico y de la experiencia de programas semejantes en el pasado. A veces, una de las conclusiones de una evaluación es justamente que las metas originales eran demasiado optimistas.

Una posibilidad es establecer gradualidad en las metas, en lugar de una meta única, y así tener umbrales de éxito parcial del proyecto. Siguiendo con el ejemplo anterior, una reducción anual del número de homicidios de más del 5% durante el período considerado podría considerarse éxito total del programa; entre 3% y 5%, éxito parcial; y menos del 3% como incapacidad de cumplir la meta.

La existencia de metas facilita enormemente la evaluación de impacto, porque permite establecer con claridad un parámetro de éxito o fracaso, generalmente cuantitativo, en relación con los objetivos originales.

5.3 Diagnóstico y teoría de cambio

El *diagnóstico* no es, en sentido estricto, un requisito de evaluabilidad, porque un programa que no partió de un diagnóstico puede, asimismo, ser evaluado de forma rigurosa. Sin embargo, la existencia de un diagnóstico previo no solo es un requisito para programas bien formulados, sino que constituye un paso adelante para la evaluación. Un diagnóstico bien elaborado ayudará a dejar claro cuáles son los objetivos de la intervención. De hecho, cuando no existe un documento programático, pero sí se cuenta con un diagnóstico previo, a veces es posible inferir

48 Un mismo objetivo medido mediante diferentes indicadores podría arrojar resultados disímiles. Por ejemplo, la reducción de la violencia podría ser mayor o menor al estimar este fenómeno a través de registros policiales, encuestas de victimización o artículos en la prensa.

los objetivos a partir de este último. Además, permite avanzar en la formulación de indicadores y metas.

La *teoría de cambio* o *lógica del modelo* (Weiss, 1995) es una explicación teórica detallada de los mecanismos por medio de los cuales las actividades desarrolladas deberían producir los cambios esperados en las variables intermedias y finales. En nuestro caso, las variables finales están referidas a los homicidios. Las variables intermedias son aquellas que deben ser modificadas como resultado de nuestras acciones para que ellas, por su parte, provoquen una disminución en los homicidios. Imaginemos, por ejemplo, que nuestra intervención pretende mejorar la capacitación de los jóvenes de forma que encuentren empleos dignos y no se dejen reclutar por grupos criminosos, lo cual debería reducir la violencia letal. En este caso, la primera variable intermedia sería el aumento de la formación o capacitación de los jóvenes. La segunda, una mejora de su empleabilidad y, en consecuencia, de la proporción de jóvenes que consiguen empleo formal. La tercera dimensión sería una reducción del reclutamiento de los grupos criminales de la zona. Si nuestra teoría de cambio está correcta, estos fenómenos deberían darse de forma sucesiva y ordenada en el tiempo, en un intervalo determinado, y al final de este proceso veríamos una caída de los homicidios.

En otras palabras, la teoría de cambio (TdC) es una teoría acerca de cómo los insumos y las actividades derivarán en productos y estos, en resultados que producirán el impacto deseado. Los supuestos acerca de las conexiones de cada eslabón de la cadena causal deben ser explicitados. A veces es posible encontrar intervenciones exitosas incluso si no se sabe nada de los mecanismos que producen el impacto (la llamada *teoría de la caja negra*). Sin embargo, lo ideal es contar con una TdC que abra esa “caja negra” y evidencie los mecanismos del programa.

La TdC proporciona una *hoja de ruta para el cambio*, basada en:

- 1) un análisis realista del contexto;
- 2) una valoración de las capacidades de los interventores;
- 3) una explicitación crítica de los supuestos del modelo, esto es, de las condiciones que tienen que cumplirse para que la intervención funcione.

Uno de los desafíos de la TdC es no dejar supuestos implícitos ni conexiones causales basadas en intuiciones o en el sentido común. Los diferentes elementos de la propuesta deben estar articulados y sus relaciones sustentadas en otros programas similares que hayan sido evaluados, en la literatura especializada y en la plausibilidad lógica y científica de los postulados.

Una TdC adecuada es esencial, tanto para el diseño como para la evaluación de los programas.

Entre otras cosas, permite:

- Estimar a priori la plausibilidad del programa.
- Diseñar intervenciones consistentes y con una lógica sólida.
- Ayudar a estimar la temporalidad de los efectos.
- Contribuir para el establecimiento de metas.
- Formular el orden de los efectos y de las variables intermediarias, con sus respectivos indicadores.
- Garantizar que todas las dimensiones relevantes serán medidas.
- Elaborar una narrativa sobre el sentido de la intervención.
- Promover una visión compartida y única sobre el programa.
- Anticipar efectos imprevistos o indeseados.
- Evitar acciones innecesarias o gastos no justificados.
- Analizar los resultados obtenidos e identificar las posibles causas del éxito o fracaso del programa.

Como ya se dijo a propósito del diagnóstico, una TdC tampoco es imprescindible para llevar a cabo una evaluación rigurosa. Sin embargo, un programa basado en una TdC sólida no solo tiene muchas más probabilidades de generar impacto, sino que facilita enormemente la tarea del equipo evaluador, pues ayuda a determinar objetivos, variables intermediarias e indicadores, además de apuntar posibles hipótesis alternativas sobre los efectos observados. Desgraciadamente, muchos programas de prevención de homicidios no cuentan con una TdC adecuada. De este modo, muchas veces se espera simplemente que una mejora general de la calidad de vida de la población o de la cohesión social, o un aumento de la cantidad de policías en servicio disminuyan los homicidios automáticamente, sin que se presenten evidencias en ese sentido.

Por ejemplo, entre las medidas contempladas por el gobierno brasileño en su tentativa de crear un Pacto Nacional de Reducción de Homicidios en 2015, que finalmente no se lanzó, estaba la mejora de la formación de los agentes penitenciarios. Aunque esta medida puede ser elogiada bajo diversos puntos de vista, no había ninguna indicación de cómo ella podría llegar a reducir los homicidios en el país en un plazo razonable.

La existencia de objetivos claros es un prerrequisito para poder formular una TdC. Es decir, cuando no se cuenta con objetivos definidos, es inútil intentar pensar en una lógica del modelo, pues no tiene sentido pensar en una cadena de efectos cuando ni siquiera se conoce el fin al que se dirige.

5.4 Temporalidad de los efectos y temporalidad de la evaluación

A menudo, se le presta escasa atención a la temporalidad de los programas y, aún menos, a la de la evaluación. Sin embargo, es imprescindible estimar, a partir de la TdC: a) el plazo en que la intervención puede provocar los efectos buscados; y b) la duración esperada de ese impacto. Ajustar esas temporalidades es algo esencial, por ejemplo, para no subestimar el tiempo necesario para que el programa tenga efecto. Desgraciadamente, muchos programas de prevención de homicidios, especialmente los que están basados en prevención primaria, como por ejemplo aquellos que se aplican sobre escuelas de enseñanza primaria ([ver apartado 2.2](#)), tendrían efecto a largo plazo y, sin embargo, muchas veces se espera que sean evaluados en un plazo breve de uno o dos años. De la misma forma, es de gran relevancia estimar cuál es el tiempo de permanencia de los efectos provocados por el programa, si son permanentes o temporales y, en este último caso, cuál es su duración prevista.

Estos tiempos (de espera hasta el impacto y de duración de este) son fundamentales, no solo para conseguir los efectos, sino también para la propia evaluación. Así, los tiempos de la evaluación están subordinados a los tiempos del proyecto (de espera y de duración), de forma que la evaluación no podrá ser inmediata si el efecto precisa de tiempo para acontecer, bajo riesgo de dejar de observarlo. De la misma forma, si el efecto es efímero, la medida deberá considerarlo para no dejar que pase desapercibido. Idealmente, debe verificarse la duración de un efecto y calcular su curva de pérdida en el tiempo. En la práctica, las evaluaciones tienen que ser concluidas en plazos relativamente cortos y no siempre pueden medir los efectos a medio y largo plazo. Sin embargo, evaluadoras y evaluadores pueden instalar sistemas de registro e información que permitan que otros, en el futuro, examinen la evolución del impacto en el tiempo.

Más allá de ajustar la temporalidad de la evaluación a la de la intervención, un problema común es que las evaluaciones a menudo son solicitadas cuando el programa está en marcha o incluso cuando ha terminado. En consecuencia, si el sistema de informaciones necesario no existe y debe ser instalado, muchas veces será imposible obtener los datos anteriores que puedan ser comparados con los registros posteriores

a la intervención. En este sentido, una de las mayores amenazas a la evaluabilidad de un proyecto es la imposibilidad de obtener informaciones retrospectivas, que serían esenciales para medir el impacto. Por eso, las evaluaciones de calidad son casi siempre diseñadas junto con el programa y no planificadas una vez que la intervención ha comenzado.

5.5 Otros elementos de la evaluabilidad

Además de los aspectos centrales para determinar la evaluabilidad que ya fueron mencionados (objetivos, metas, diagnóstico, TdC y temporalidad), hay otra serie de factores que, sin necesariamente impedir la posibilidad de una evaluación, ciertamente afectan de forma importante su calidad.

Entre ellos, se puede mencionar el pequeño número de homicidios en el momento inicial, el desplazamiento criminal, la regresión a la media, los cambios poblacionales, las alteraciones que afecten a todo el territorio y las intervenciones con múltiples estrategias simultáneas. Todos ellos serán tratados en el [Capítulo 11: Dificultades para evaluar programas de prevención de homicidios](#).

Existen protocolos para ayudar a recoger información que ayude a determinar la evaluabilidad de un proyecto, que en inglés se denominan *Evaluability Assessments*. Como ejemplo, cabe mencionar el protocolo propuesto por la UNODC.

Evaluabilidad

Para que un Programa sea evaluable debe contar con:

Objetivos definidos. De preferencia, también con indicadores, metas y una teoría de cambio que explique cómo las acciones deberían generar los efectos buscados y el tiempo en que cabe esperar esos efectos.

Sistemas de información que registren de forma regular los datos necesarios para la evaluación a lo largo de todo el proceso: antes, durante y después de que el programa haya sido implementado.

Informaciones complementarias que permitan comparar con otros grupos o individuos en situación similar, pero que no hayan sido objeto de la intervención.

RESUMEN

CAPÍTULO 6. DISEÑOS DE EVALUACIÓN

Una evaluación no puede ser algo improvisado ni decidido de repente. Necesita una planificación cuidadosa y un diseño adecuado que permitan, sobre todo, tener un alto grado de seguridad sobre si la intervención tuvo o no impacto. A esto se le llama en la literatura *validez interna* (Campbell & Stanley, 1963), que es la certeza de si los resultados observados fueron o no provocados por el programa que fue introducido.

En muchas ocasiones, se verifica que el cambio es positivo, pero no sabemos si fue producto de la intervención o de otros factores. Por ejemplo, los homicidios cayeron en el municipio en el que trabajamos, pero eso puede deberse, además de a nuestro programa, a una mejora del escenario socioeconómico o a una tregua entre los grupos criminales. En ese caso, carecemos de validez interna. Cuando, por el contrario, el estudio evaluativo nos permite estar seguros de en qué medida el resultado es explicado o no por el programa, entonces tenemos un alto grado de validez interna. Obsérvese que la validez interna no es sinónimo de impacto, pues con alta validez interna es posible concluir tanto la presencia como la ausencia de impacto.

La otra dimensión central es la *validez externa*, que puede ser definida como la capacidad de generalizar el resultado de nuestro estudio evaluativo a otros lugares, momentos, personas y formas de medir las variables independientes (causas) y dependientes (efectos) (Cook & Campbell, 1979). Si nuestro estudio es impecable desde el punto de vista de la validez interna, o sea, estamos absolutamente seguros de cuál fue la relación causal observada, pero estos resultados no pueden ser generalizados, la evaluación no será de gran utilidad. Si nuestra intervención consiguió reducir los homicidios, pero en condiciones muy específicas que dependían de contextos concretos, de forma que no tendrá éxito si la intentamos reproducir en otros barrios, otras ciudades o en otras instituciones, no habremos avanzado gran cosa. Esto suele suceder, por ejemplo, cuando los programas de reducción de homicidios dependen de personas concretas con gran influencia o capacidad (un juez que acepta aplicar principios de disuasión focalizada, un o una oficial de policía que consigue superar la desconfianza de la comunidad, un líder que logra reunir muchos recursos económicos o simbólicos, etc.). Algo parecido sucede cuando el resultado depende de un contexto muy específico que no será fácil de repetir, como un acuerdo entre grupos armados o un momento de expansión económica que permite ofrecer salarios atractivos a muchos jóvenes.

No podemos olvidar que las evaluaciones son realizadas no tanto para dar cuenta del pasado, sino para planear el futuro, continuando o expandiendo los programas exitosos y eficientes y modificando o interrumpiendo los que carecen de eficacia.

En general, aumentamos la validez interna equiparando el contexto que recibió el programa con otro contexto paralelo con el que se pretende compararlo, para verificar si existen diferencias que puedan ser atribuidas a la intervención. Ello dependerá fundamentalmente del diseño de la evaluación. Por otro lado, la validez externa es más difícil de precisar a priori y se obtiene a través de la representatividad de todos los elementos involucrados en la experiencia (personas, momentos, formas de aplicar el programa, etc.), con relación a los elementos a los que se quiere generalizar (otras personas, momentos, etc.).

Así, los esfuerzos para ajustar el diseño para aumentar la validez interna, o sea la calidad de la inferencia casual, pueden generar contextos cada vez más atípicos que dificultan la generalización. En otras palabras, suele haber una cierta tensión entre la validez interna y la externa, de manera que las tentativas de aumentar una de ellas pueden comprometer la otra. Por eso, es preciso un equilibrio entre un grado aceptable de validez interna, pues, si no, no habría nada que generalizar, y un grado mínimo de validez externa, esto es, de capacidad de generalización.

El desafío que toda evaluación de impacto enfrenta es responder a la pregunta de qué hubiera ocurrido si el programa no hubiera sido implementado, manteniendo todo lo demás constante. En nuestro caso, queremos saber cuál hubiera sido la incidencia de homicidios si el programa no hubiera existido. En la medida en que tengamos una respuesta para esta cuestión, podemos compararla con lo que efectivamente sucedió después de la introducción del programa, esto es, con los homicidios que efectivamente se registraron. La diferencia entre el escenario real (posintervención) y el escenario hipotético (sin intervención) nos revelaría, entonces, el impacto del programa.

En suma, estamos intentando estimar lo que hubiera ocurrido bajo un supuesto hipotético (la ausencia del programa que efectivamente existió), lo cual se conoce con el término de *contrafactual*. Dado que los contrafactuales no pueden ser observados, sino solo estimados, la causalidad solo puede ser inferida y esta inferencia nunca será infalible, sino aproximada y probabilística. Todas las investigaciones causales, no solo en el área de evaluación, enfrentan este problema.

La mayor dificultad de una evaluación es construir un escenario comparativo lo más similar posible al escenario intervenido antes de recibir el programa, es decir, un escenario hipotético de ausencia de intervención que sirva para compararlo con lo que efectivamente sucedió.

A este escenario comparativo se le denomina, en lenguaje experimental, *grupo de control*. Pero incluso cuando un grupo de control no es posible, la evaluación continúa obligada a generar este contexto hipotético de ausencia de programa de alguna forma medible.

En términos prácticos, todo lo anterior significa que no es posible probar el impacto de un programa de reducción de homicidios simplemente diciendo que la tasa de homicidios es reducida después de que la iniciativa entró en vigor. Obviamente, es posible que la tasa ya fuera baja antes y que el programa no haya cambiado la situación. Por tanto, en primer lugar, es preciso demostrar que hubo una *reducción* de los homicidios en un momento posterior al programa (que llamaremos *postest*), en comparación con un momento anterior (que denominaremos *pretest*). Para que esto sea posible es necesario tener mediciones anteriores a la intervención, lo que en la literatura se conoce a menudo como una *línea-base*. Una línea-base es una medida o un conjunto de medidas tomadas a intervalos regulares que permiten establecer la incidencia media de la variable dependiente (en nuestro caso, los homicidios) antes de que el programa sea implementado, para poder compararla con el desempeño posterior. El objetivo es poder separar la variación “natural” de la variable, tal como sucedía antes de la intervención, del impacto de esta última. En consecuencia, una medición única no nos dice prácticamente nada sobre impacto.

Sin embargo, tampoco basta mostrar que hubo una caída de los homicidios, pues este declive podría deberse a otros factores ajenos al programa: socioeconómicos, políticos, relativos al propio mundo de la criminalidad, etc. También podría suceder lo contrario, esto es, que los homicidios hayan aumentado, pero que lo hayan hecho en menor medida como consecuencia del impacto del programa. Así, contar con *pretest* y un *postest* son condiciones necesarias, pero no suficientes para mostrar causalidad. Para ello, es preciso que se construya el *contrafactual*, es decir, una estimativa confiable de la tasa de homicidios que hubiera existido en ausencia del programa, manteniendo todo lo demás constante.⁴⁹ Solo de esa forma se podrá calcular impacto.

Otra forma diferente, pero convergente de conceptualizar el problema, es afirmar que para demostrar la relación entre una supuesta causa (el programa) y un efecto (la reducción de homicidios) es necesario descartar todas las explicaciones alternativas y plausibles para esta asociación, tal como enseñaba el filósofo británico John Stuart Mill (1888).

49 Esta idea de mantener todo lo demás constante se denomina con la cláusula latina *ceteris paribus*.

6.1 Diseños experimentales

Los diseños experimentales son considerados la forma ideal de estimar causalidad. Su origen se sitúa en las investigaciones de agronomía, pero

posteriormente fueron adoptados sistemáticamente en áreas como la biomédica, donde hoy ejercen un dominio casi absoluto.

Hay tres requisitos fundamentales para que un estudio sea considerado un experimento:

- a) La causa debe ser introducida por el experimentador o la experimentadora, que controla así el momento en que ella se produce y puede obtener medidas antes y después. Esta condición suele ser fácil de cumplir en el caso de programas de reducción de homicidios, siempre que evaluadoras y evaluadores sean llamados antes de iniciar la intervención.
- b) El estudio se da en un ambiente controlado, para que se pueda evitar, en la medida de lo posible, la interferencia de otras variables, llamadas *variables intervinientes*. Estas podrían oscurecer o incluso eliminar el vínculo entre la causa y el efecto. En las ciencias naturales, este control experimental se consigue por medio de un laboratorio, que es un ambiente inerte y esterilizado donde solo actúan los elementos que los investigadores introducen en él. En el área social, esto es mucho más difícil, porque hay innumerables factores que inciden en los procesos que se pretenden modificar. A pesar de ello, es posible organizar el estudio para intentar minimizar la influencia de estos factores.
- c) Las personas o las unidades (escuelas, barrios, municipios, etc.) que van a participar en el estudio son distribuidas en dos grupos de forma aleatoria, o sea, por sorteo. Un grupo, llamado experimental, será sometido al programa que se va a evaluar. El otro grupo, denominado de control, experimentará las mismas situaciones que el grupo anterior excepto el programa, que estará ausente. La aleatorización es una forma de garantizar que los dos grupos son aproximadamente iguales en el momento inicial, de modo que cualquier diferencia posterior pueda ser atribuida al programa. En otras palabras, es una forma de operacionalizar la mencionada cláusula de *ceteris paribus* (todo lo demás, constante). Aunque puede parecer que hay otras formas de garantizar la igualdad inicial de ambos grupos, por ejemplo, escogiendo personas o unidades muy semejantes en los dos casos, el sorteo es la mejor opción, porque garantiza una equivalencia aproximada no solo en las dimensiones que sabidamente afectan al resultado, en nuestro caso a la violencia letal, sino en todas las dimensiones posibles, incluyendo aquellas que podrían tener un efecto inesperado. El requisito para que la aleatoriedad produzca equivalencia es que el tamaño de la muestra de ambos grupos sea suficientemente grande como para que el error aleatorio

sea reducido. El error aleatorio siempre existirá, pero puede ser minimizado aumentando el tamaño de las muestras y puede, además, ser estimado probabilísticamente con precisión. En el caso de los programas de reducción de homicidios, sortear a las personas que se beneficiarán de los programas no siempre es fácil, por limitaciones éticas, políticas o prácticas. De hecho, en el área de seguridad pública no existe una tradición de estudios experimentales, a pesar de que existen precedentes interesantes (ver cuadro sobre el experimento de Kansas City). En diversas situaciones no será posible aleatorizar. Por ejemplo, no se puede sortear en qué ciudades una nueva ley entrará en vigor y en cuáles no. Pero incluso cuando un programa aspira a ser universal, a veces es posible sortear las unidades o las áreas en que se implantará en primer lugar y usar el resto como grupo de control mientras no reciben el programa (ver Capítulo 12). De cualquier manera, si existe una posibilidad de hacerlo, la distribución aleatoria es la mejor opción metodológica y a menudo no supone un alto coste ni una gran complicación logística. Un ejemplo de evaluación experimental es el estudio que estima el impacto de la Iniciativa de Seguridad Regional Centroamericana (CARSI), un programa de prevención comunitaria de la violencia financiado por los Estados Unidos de América en El Salvador, Guatemala, Honduras y Panamá. Esta intervención escogió 127 barrios de alto riesgo y asignó aleatoriamente⁵⁰ 65 para recibir la intervención y 62 para componer el grupo de control (Berk-Seligson *et al.*, 2014).⁵¹

El experimento de patrullaje de Kansas City (1972-1973)

Este estudio es considerado el experimento pionero en el área de seguridad pública. El objetivo era medir el impacto del aumento de las patrullas policiales en las calles sobre el crimen y la percepción de seguridad. El distrito sur de la ciudad se dividió en 24 áreas, 15 de las cuales fueron elegidas. Las quince áreas se distribuyeron en cinco bloques, de modo que las tres áreas de cada bloque presentaban características urbanas y socioeconómicas relativamente semejantes. A partir de ahí, se sorteó, dentro de cada bloque, cuál de las tres áreas sería asignada a cada uno de los tres tipos siguientes de patrullaje: a) se eliminaron todas las patrullas preventivas y la policía solo respondía a llamadas de emergencia; b) el patrullaje preventivo se duplicó; y c) el patrullaje no cambió.

De esta forma, las 15 áreas se dividieron aleatoriamente en tres grupos: a) en las 5 áreas del primer grupo no había patrullaje preventivo; b) en las 5 áreas del segundo grupo se duplicó el patrullaje utilizando los recursos liberados en el grupo anterior; c) en las 5 áreas del tercer grupo, nada cambió.

50 En realidad, la selección no fue aleatoria en uno de los cuatro países: Honduras.

51 La variable dependiente en este caso no fueron los homicidios registrados, sino la percepción de los vecinos sobre si había habido o no asesinatos en los 12 meses anteriores.

Además, el diseño geográfico tomó el cuidado de que la policía pudiera, en todas las áreas, responder a las llamadas de emergencia con relativa rapidez incluso en ausencia de patrullaje preventivo. Esta era una forma de intentar evitar lo que, a ojos de la policía, podía llegar a ser un efecto perverso de la intervención (ver [Capítulo 3](#) y [Capítulo 7](#)).

El delito se midió por medio de encuestas de victimización y registros policiales. El miedo al delito y la satisfacción con el servicio policial se midieron con encuestas. Contra lo que se esperaba, no hubo diferencia entre las áreas sujetas a diferentes tipos de patrullaje en términos de delito registrado o denunciado en las encuestas, ni siquiera en los crímenes que se suponían más sensibles a la presencia policial (robos en residencias, robos de automóviles, vandalismo, etc.). Asimismo, la percepción de los ciudadanos sobre el servicio policial y sobre su propia seguridad o inseguridad no varió en función del nivel de patrullaje.

(Kelling *et al.*, 1974)

En ocasiones, hay excepciones a alguno de estos tres requisitos que definen un experimento, que no comprometen necesariamente el carácter experimental del estudio. Un ejemplo es el llamado *experimento natural*, situación en que el experimentador o la experimentadora no introducen la causa, pero esta se da de tal forma que la diferencia entre los que recibieron el programa y los que no lo hicieron es casi aleatoria, como si hubiera sido introducida por los evaluadores. Imaginemos que en un país los jóvenes fueron llamados a cumplir el servicio militar por sorteo durante varios años. Si quisiéramos saber el impacto del servicio militar en la probabilidad de ser víctima de homicidios a lo largo de la vida, solo tendríamos que comparar las tasas de homicidio de los que sirvieron al ejército con las de los que entraron en las listas, pero no fueron sorteados. La validez interna de esta comparación sería muy elevada, a pesar de que no hayan sido los evaluadores y evaluadoras quienes decidieron quién iba a hacer el servicio militar y quién no.

El diseño experimental más clásico es el de un grupo experimental y otro de control, conformados aleatoriamente, con pretest y postest. La notación más común es usar O_1 para el pretest y O_2 para el postest. La R indica que la asignación a los grupos fue aleatorizada (del inglés, *random*).

Grupo experimental (GE)	R	O_1	X	O_2
Grupo de control (GC)	R	$O_{1'}$		$O_{2'}$

El cálculo del impacto sería la diferencia entre el incremento observado en el grupo experimental y el registrado en el grupo de control:

$$\text{IMPACTO} = (O_2 - O_1) - (O_2' - O_1') \\ \Delta GE - \Delta GC$$

El ejemplo clásico de ese diseño experimental es la evaluación de las vacunas en el área biomédica, en la que los niños son sorteados y la mitad recibe la vacuna y la otra mitad un placebo, una sustancia de apariencia semejante, pero sin principios activos. Ni el médico que administra la vacuna ni su paciente saben si este último pertenece al grupo experimental o al de control, un procedimiento llamado *doble-ciego*, que es útil para garantizar que no haya interferencias en la asignación aleatoria. Al final del proceso, el experimentador o la experimentadora compara las tasas de morbilidad de un grupo con las del otro para determinar el impacto de la vacuna.⁵²

En las evaluaciones de programas de prevención de homicidios, una de las primeras cuestiones que deben ser decididas es cuál es la *unidad de análisis*. La unidad de análisis es aquella sobre la que se efectuarán las mediciones; puede ser el individuo, la familia, la institución, la comunidad o el barrio. Lo más común es que no sea el individuo, sino una cierta área territorial sobre la cual se calcula la tasa de homicidios.⁵³ En consecuencia, cada área es un caso, de manera que sería necesario tener un alto número de casos para conseguir significación estadística y poder identificar algún eventual impacto. Así, lo ideal sería contar con al menos 30 áreas que reciban el programa y otras 30 en el grupo de control. De hecho, cuando añadir más casos al grupo de control no tiene costos adicionales, como cuando los datos vienen de un registro oficial (por ejemplo, las estadísticas criminales), el grupo de control puede estar integrado por un número mayor de casos que el grupo experimental, lo cual mejorará la precisión de la estimativa del primero sin afectar a la de este último. De forma general, cuanto mayor sea el tamaño de las áreas que constituyen la unidad de análisis, menor será su número. Si trabajamos con barrios, por ejemplo, tenderemos a tener más casos que si comparamos ciudades.

Hay diversos diseños experimentales (Shadish *et al.*, 2002),⁵⁴ pero el más simple es el de dos grupos sin pretest.

GE	R	X	O
GC	R		O

52 En el área social, el concepto de placebo no es común, porque es raro contar con un programa alternativo que se parezca a aquel que se quiere evaluar en todo menos en el aspecto central a ser evaluado.

53 Con todo, en programas de prevención terciaria dirigidos a agresores o víctimas, por ejemplo, no es raro que la unidad de análisis sea el individuo.

54 Ver específicamente [Capítulo 8](#) de este libro.

Esta opción puede ser útil en los casos en que no exista un sistema de registro de informaciones previo a la intervención, pero esta haya sido aleatorizada. Idealmente, el número de casos debería ser elevado, puesto que se pierde el control adicional proporcionado por la observación inicial y la equivalencia entre los grupos depende exclusivamente de la aleatorización.

6.2 Diseños cuasiexperimentales

En muchas ocasiones no es posible realizar una evaluación experimental, porque no se puede asignar aleatoriamente la intervención o porque no es posible controlar suficientemente el proceso. En estos casos, se procura encontrar un parámetro de comparabilidad que sustituya al grupo de control aleatorizado y permita establecer causalidad; hablamos, entonces, de los *estudios cuasiexperimentales*. Es importante observar que estos estudios no son iguales a los observacionales, en los que simplemente se describe un grupo, sin que se busque un elemento de comparación.

Los diseños cuasiexperimentales suelen pertenecer a uno de los siguientes tres grupos:

a) Diseño con un **pseudo-grupo de control** (Shadish *et al.*, 2002),⁵⁵ muchas veces seleccionado después de la intervención. Se trata de buscar individuos/grupos/áreas que se parezcan a aquellos donde el programa fue aplicado. Idealmente, deben ser similares en todo menos en dicho programa. En la práctica, se busca sobre todo comparabilidad en la propia incidencia de homicidios, en las variables que tengan mayor influencia potencial sobre esta incidencia y en aquellas otras que puedan generar una interacción con el tratamiento experimental. Por ejemplo, si el programa fue aplicado en comunidades pobres con altos niveles de homicidio, sería preciso buscar otras comunidades con estas mismas características para componer el pseudo-grupo de control. Y si se supone que la composición étnica de la población puede afectar la forma en que el proyecto se desarrolla, las comunidades buscadas también deberían asemejarse a las intervenidas en esta dimensión. Como no hubo aleatorización, no es posible garantizar la equivalencia plena, pero la validez interna mejorará en la medida en que estos grupos se parezcan al grupo experimental. Este es un diseño muy común y su calidad depende de esa capacidad de obtener individuos/grupos/áreas similares.

En las últimas décadas se han desarrollado técnicas específicas para intentar mejorar la semejanza entre el grupo de control y el experimental cuando la aleatorización es imposible. Entre ellas se pueden mencionar la puntuación de propensión, el control sintético y las diferencias-en-diferencias.

55 Ver concretamente los [Capítulo 4](#) y [Capítulo 5](#) de este libro.

La *puntuación de propensión* (Guo & Fraser, 2010) comienza con la construcción de un modelo estadístico que estima la probabilidad de que un caso haya sido incluido en el grupo experimental (o en el de control) a partir de una serie de variables independientes relevantes a través, por ejemplo, de una regresión logística. Una vez obtenida esta probabilidad de selección se pueden escoger los integrantes del grupo de control, de modo que tengan la misma probabilidad a priori de pertenecer al grupo experimental, o sea, de forma que tengan un perfil muy parecido a los integrantes de este último. En nuestro caso, si fueron elegidas para el programa comunidades pobres y de alta incidencia de homicidios, se calcularía la probabilidad de cada una de las comunidades de haber sido escogidas y, a partir de ahí, se seleccionarían para el grupo de control aquellas que tuviesen la misma probabilidad de selección que las que fueron efectivamente elegidas. Esto permitiría incluir en ambos grupos comunidades con características muy parecidas en una serie de dimensiones que están empíricamente asociadas a esa probabilidad de selección. De esa forma, se espera que las comunidades del grupo de control sean semejantes a aquellas donde el programa fue implementado.

Otra posibilidad es ponderar los casos de ambos grupos por el inverso de la probabilidad de integrar el grupo experimental, homogeneizando así la representatividad de ambos. En ambos casos, la expectativa es que estos procedimientos equilibren la composición del grupo experimental y de control. Un ejemplo de la utilización de esta técnica fue la evaluación del impacto sobre los homicidios del proyecto Metrocable en Medellín ([ver apartado 14.1](#)). Otra investigación que usa la puntuación de propensión es la evaluación del programa Fica Vivo! en Minas Gerais, Brasil (Peixoto *et al.*, 2008b).

Por su parte, la estrategia del *grupo de control sintético* (Abadie *et al.*, 2014) sustituye un grupo de control único por una media de diversos grupos de control, cada uno de ellos con una ponderación que es proporcional a su similitud al grupo experimental (en un momento anterior a la intervención) en una serie de variables relevantes. Se trata de un abordaje muy similar al de la puntuación de propensión. La diferencia aquí es que, en vez de estimar la probabilidad de que todas las unidades sean seleccionadas dentro del grupo experimental, se estima directamente la similitud en el pretest entre el grupo experimental y todos los posibles integrantes del grupo de control y se les adjudica a estos últimos una ponderación en función de esa similitud. En otras palabras, las unidades más parecidas al grupo experimental tendrán más influencia en el grupo de control que las que son menos parecidas, pero todas pueden contribuir.

La técnica de las *diferencias-en-diferencias* (Foguel, 2016) puede ser aplicada a diversos tipos de diseños evaluativos y compara la diferencia

entre el grupo experimental, antes y después de la intervención, con la misma diferencia en el grupo de control. La premisa central es que, aunque haya diferencias significativas en los valores del pretest entre ambos grupos, la evolución a partir de ahí será igual en el grupo de control a lo que hubiera sido en el grupo experimental si no hubiera existido la intervención. La idea es que, de esta forma, se controlarían todas las diferencias entre ambos grupos que sean constantes en el tiempo. Por otro lado, como muestran los estudios históricos en el área de educación que utilizaban grupos experimentales y de control que estaban equiparados en el valor del pretest, subsisten aún amenazas a la validez interna como la regresión diferencial a la media de unos y otros, en función de su puntuación inicial. Un ejemplo de la aplicación de esta técnica es la evaluación del impacto en los homicidios de la Ley “Maria da Penha” contra la violencia doméstica en Brasil ([ver apartado 14.2](#)). Otros estudios evaluativos que mencionan la estrategia de diferencias-en-diferencias son la evaluación del Plan Nacional de Vigilancia por Cuadrantes en Colombia (Fundación Ideas para la Paz, 2012) y la evaluación de la creación de espacios públicos en Sucre, Venezuela (García, 2012).

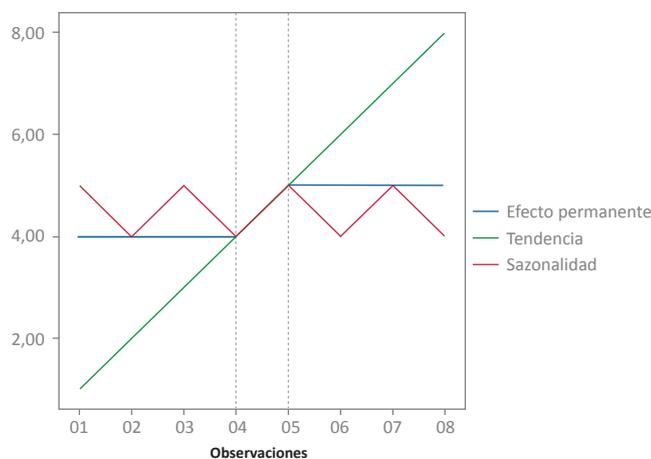
De forma general, las puntuaciones de propensión, el control sintético y las diferencias-en-diferencias han representado un avance en la búsqueda de grupos de control no equivalentes. Sin embargo, es importante recordar que estas estrategias no alcanzan la misma validez interna que la aleatorización, puesto que los modelos estadísticos no garantizan la equivalencia plena entre los grupos. En última instancia, se puede afirmar que los controles estadísticos a posteriori son importantes, pero no sustituyen la fortaleza de un buen diseño a priori.

b) Análisis de **series temporales**, que modelan la evolución de un fenómeno antes de la introducción del programa y comparan su desempeño posterior con la predicción que podía realizarse a partir de la serie anterior a la intervención (McDowall *et al.*, 1980). Una serie temporal no es nada más que la medición sistemática de un fenómeno a intervalos regulares, lo cual permite estimar sus tendencias o variaciones cíclicas. Esta es una estrategia metodológica comúnmente usada cuando no es posible contar con un grupo de control. Por ejemplo, una nueva ley nacional afecta al conjunto del territorio y no es posible restringir su aplicación apenas a determinados locales. La ausencia de grupo de control limita la validez interna, pero, a pesar de ello, la serie temporal permite intentar aislar el impacto del programa.

Observemos el gráfico siguiente, que presenta tres series temporales diferentes. Suponiendo que el programa es introducido entre O4 y O5,

vemos que, si tuviésemos apenas una medida antes (pretest) y otra después (postest), los resultados serían idénticos en los 3 casos. Sin embargo, un examen de la serie completa revela que la línea azul muestra un cambio justo posterior a la introducción del programa, que se mantiene posteriormente, lo cual constituye evidencia a favor de un impacto del mismo. Por su parte, la línea verde muestra una tendencia que existía antes de la introducción del programa y continúa después, lo que parece rechazar la interpretación de un posible impacto de la intervención. Por último, la línea gris registra un ciclo regular (estacionalidad), de subida y bajada, que tampoco parece estar afectado por el programa.

Gráfico 7. Ejemplos de posibles efectos reflejados en una serie temporal



Fuente: Elaboración propia. Nota: Las líneas verticales muestran los momentos de inicio y de fin del programa.

Cuanto mayor es el número de observaciones, mayor es la posibilidad de sofisticar las estimaciones del modelo y mejor la capacidad de predecir la serie a partir de sus registros pasados. Por otro lado, la validez interna aumenta significativamente cuando es posible obtener una serie temporal paralela correspondiente a un grupo de control. En este caso, se puede tener la ventaja de la aleatorización junto a la capacidad predictiva de la serie temporal. Imaginemos, por ejemplo, que tenemos un registro mensual de homicidios en un grupo de comunidades en las que vamos a introducir nuestro programa de prevención y, al tiempo, tenemos ese mismo registro para otras comunidades sin el programa. Si la elección de las comunidades se diera aleatoriamente, se daría la situación ideal y estaríamos ante un experimento. Pero incluso si la aleatorización no es posible, si conseguimos una serie temporal para un grupo de control parecido al experimental, habremos incrementado nuestra validez interna.

Un ejemplo de la aplicación de las series temporales es la evaluación del impacto en los homicidios del programa Unidades de Policía Pacificadora en Río de Janeiro (Cano *et al.*, 2012).

Cuando los efectos son temporales y de corta duración, es posible introducir y retirar el programa múltiples veces durante la serie y verificar el impacto que cada introducción tiene sobre esta.

O₁ X O₂ O₃ X O₄ O₅ X O₆ O₇ X O₈

Obviamente, cuando el impacto de la intervención es permanente, este diseño no tiene utilidad, pues los efectos de la primera introducción se confundirían con los de las siguientes. Un ejemplo del uso de la estrategia de las series temporales interrumpidas para intervenciones de corta duración es la evaluación del impacto de la prohibición de porte de armas en algunos fines de semana en Bogotá y Cali ([ver apartado 14.3](#)).

Mejor aún es si contamos con series temporales alternadas en las que el programa es introducido en ocasiones en un grupo y, en otras oportunidades, en el otro. De esa forma, los dos grupos se alternarían en la función experimental y de control, fortaleciendo significativamente la validez interna, pues sería muy difícil encontrar una explicación alternativa que justificase por qué la reducción sucede en cada grupo justamente después de la introducción del programa.

O ₁	X	O ₂	O ₃	X	O ₄	O ₅	X	O ₆	O ₇

O ₁	O ₂	X	O ₃	O ₄	X	O ₅	O ₆	X	

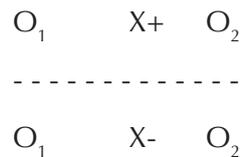
c) **Otros diseños cuasiexperimentales.** Existe toda una serie de diseños que, aunque son menos frecuentes que los anteriores, pueden ser empleados para mejorar la validez interna (Cook & Campbell, 1979).

Entre ellos se pueden mencionar:

- 1) Diseño con un único grupo, pero con múltiples variables dependientes, entre las cuales una debería ser, por razones teóricas, especialmente afectada por la intervención. Si el resultado es especialmente intenso en esta variable, mucho más que en las otras, ello refuerza la interpretación de que el programa tuvo impacto. Por ejemplo, si una nueva estrategia de policía, con mayor presencia policial, está pensada para reducir los robos en la calle, esperamos que los casos de violencia doméstica, que no

dependen en principio del patrullaje, se mantengan más o menos constantes, en contraste con los delitos contra la propiedad, que deberían caer. Si, por el contrario, caen todos los crímenes en la misma proporción, esto sugiere que la caída se debe a otros motivos y no al patrullaje policial.

- 2) Inversión de tratamientos, que consiste en aplicar el tratamiento contrario al original, para ver si se obtienen efectos también contrarios.

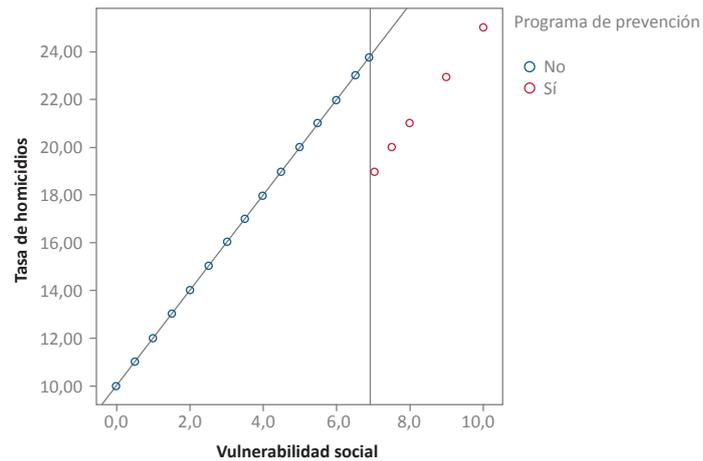


Para que este diseño sea posible es necesario, en primer lugar, que el concepto de *programa invertido* tenga sentido y no sea éticamente cuestionable, lo que no siempre ocurre en el área de prevención de violencia. Por ejemplo, lo opuesto a un programa de transferencia de renta para familias con hijos en situación de riesgo sería un impuesto adicional para estas familias, lo que sería éticamente inaceptable. Otro requisito es que exista un modelo teórico, una teoría de cambio que prediga que tratamientos inversos deberían provocar, de hecho, efectos inversos, lo que no siempre sucede. A pesar de que no son requisitos siempre fáciles de cumplir, a veces de hecho es posible utilizar este abordaje. El experimento clásico de Kansas City sobre el impacto del patrullaje preventivo, ya descrito, comparó el resultado de duplicar el patrullaje preventivo con el de eliminarlo completamente, añadiendo además un grupo de control en el que los niveles de patrullaje continuaron como estaban originalmente (Kelling *et al.*, 1974).

- 3) Discontinuidad en la regresión (Trochim, 1984). Este es un diseño adecuado a situaciones en que la atribución de cada unidad al grupo experimental o al de control se realiza de acuerdo a la puntuación en la propia variable dependiente en el momento inicial (pretest) o de acuerdo con alguna otra dimensión cuantitativa que esté fuertemente relacionada con ella. Los que superen (o, por el contrario, los que no alcancen) una determinada puntuación (punto de corte) integran el grupo experimental y el resto hacen parte del grupo de control. Una vez que la intervención se ha realizado, se hacen regresiones entre el pretest y el posttest de forma separada para los dos grupos, el experimental y el de control. Cada regresión tiene como variable dependiente la tasa final de homicidios y como variable independiente el valor del puntaje que condicionó la atribución del programa en cada área.

Imaginemos, por ejemplo, que hay una fuerte asociación entre la tasa de vulnerabilidad social de cada comunidad y su tasa de homicidios. Si el programa de prevención es otorgado solo a las comunidades más vulnerables y más violentas y consigue tener impacto positivo en la reducción de los homicidios, esperaríamos un resultado como el que muestra el Gráfico 8.

Gráfico 8. Ejemplo hipotético de discontinuidad en la regresión



Fuente: Elaboración propia.

De hecho, el salto o la discontinuidad entre las dos regresiones justamente en el punto de corte, aquel que determinó qué comunidades integrarían el grupo experimental, es el mejor indicador de impacto del programa.

En suma, la discontinuidad en la regresión es posible cuando se cumplen las siguientes condiciones: a) la concesión o no del programa depende de la puntuación en una dimensión cuantitativa, con un punto de corte claro, que debe ser respetado escrupulosamente; b) esta dimensión puede ser el pretest (o sea, el nivel de homicidios antes del programa) u otra variable con fuerte correlación con este; c) las relaciones entre ambas dimensiones, que serán usadas en la regresión, deben ser lineales.

6.3 Midiendo el rigor metodológico: la escala de Maryland

La *escala de Maryland* intenta medir la calidad metodológica de los diseños de evaluación a través de cinco niveles, de menor a mayor validez interna (Farrington *et al.*, 2002).

- **Nivel 1:** Asociación de una variable dependiente (en nuestro caso, homicidios) con una independiente (programa) en un único momento del tiempo, sin pretest ni grupos de control. Un ejemplo en este sentido puede ser la constatación de que las

estaciones policiales con mayor número de policías por habitante tienen menores tasas de homicidio. Los diseños de nivel 1 son extremadamente frágiles, dado que no descartan explicaciones alternativas a la asociación que se pretende probar. De hecho, la policía puede, por ejemplo, privilegiar la lucha contra los crímenes contra la propiedad en las áreas comerciales, que son áreas donde tradicionalmente la incidencia de homicidios es baja, independientemente de la presencia policial.

- **Nivel 2:** Comparación de la variable dependiente (homicidios) al menos en dos momentos del tiempo, antes y después de la intervención. Aunque se realizan pretest y postest, no se utiliza un grupo de control. Siguiendo con el ejemplo anterior, este caso podría responder a la observación de que los homicidios disminuyeron en algunas estaciones policiales después de que se incrementó la dotación policial en esas áreas. Los diseños de nivel 2 continúan siendo frágiles, ya que tampoco permiten descartar explicaciones alternativas, como podría ser un descenso de los homicidios en toda la ciudad por otros motivos que no tienen que ver con la presencia policial.
- **Nivel 3:** Comparación de la variación en la incidencia de homicidios entre lugares donde el programa está presente y otros en que está ausente. En este caso se puede hablar de un grupo experimental y un grupo de control, aunque estos últimos no son equivalentes porque no han sido aleatorizados. Un ejemplo en este sentido sería la comprobación de que los homicidios disminuyeron en los distritos en que se incrementó el número de policías, pero no en otros distritos en que la dotación se mantuvo constante. Aunque los diseños de nivel 3 son notoriamente mejores que los anteriores, la falta de equivalencia no permite descartar sesgos de selección en la conformación de los grupos, o sea, la existencia de perfiles diferentes entre el grupo de control y el experimental. De existir diferencias relevantes entre ambos con relación a variables que puedan influenciar la incidencia de homicidios, esto podría estar afectando a los resultados.
- **Nivel 4:** Comparación de la incidencia de homicidios en múltiples lugares con el programa y en múltiples lugares sin el programa, controlando estadísticamente la influencia de otras variables. En este caso se utilizan varios grupos experimentales y de control no equivalentes, realizando además un control estadístico de variables relevantes. Un ejemplo sería la observación de que los homicidios fueron reducidos en áreas donde se incrementó el número de policías, pero no en otras áreas donde no hubo tal aumento, controlando el efecto de la composición

sociodemográfica de la población, de la instalación de cámaras de vigilancia y del número de sentencias condenatorias contra homicidas residentes en las respectivas áreas. El diseño de nivel 4 otorga una validez mucho mayor que los anteriores, pero aún no permite garantizar que no existan efectos de selección o diferencias entre los grupos comparados.

- **Nivel 5:** Comparación entre la incidencia de homicidios en lugares donde el programa está presente y la incidencia en lugares donde está ausente, de forma que la atribución de dichos lugares a un grupo (experimental) o a otro (de control) sea realizada aleatoriamente. Siguiendo el ejemplo anterior, en este caso se llevaría a cabo un sorteo para decidir qué áreas de la ciudad recibirían un refuerzo policial y cuáles no, como fue realizado en el experimento de patrullaje de Kansas City anteriormente descrito. Este tipo de diseño experimental otorga la máxima validez a la atribución de causalidad a la intervención.

RESUMEN

Diseños de evaluación

Diseños sin validez interna

Solo medición posterior al programa (postest de un grupo): los homicidios son bajos después del programa. Tal vez eran bajos antes.

Medición anterior y posterior al programa, pero sin nada con qué comparar: bajaron los homicidios. Pueden haber bajado por muchas causas.

Medición anterior y posterior en grupos diferentes en diversas dimensiones, además del programa: bajaron los homicidios, a diferencia de otros lugares. Muchas explicaciones posibles para las diferencias entre los lugares.

Diseños cuasiexperimentales

Diseños con pseudo-grupo de control, en los que se busca un grupo parecido para comparar con el grupo experimental.

Técnicas para mejorar la comparabilidad: puntuación de propensión, grupo de control sintético, diferencias-en-diferencias y análisis de series temporales.

Otros diseños cuasiexperimentales: diseño con grupo único, pero con múltiples variables dependientes, de impacto previsto diferenciado; inversión de tratamientos; y discontinuidad en la regresión.

Diseños experimentales

Comparación entre un grupo experimental (con programa) y uno de control (sin programa) con asignación aleatoria de los casos a los grupos: máxima validez interna.

CAPÍTULO 7. DESAFÍOS ÉTICOS DE UNA EVALUACIÓN

7.1 Introducción

La ética se refiere a una práctica correcta o acordada dentro del contexto social y cultural en el que la evaluación es realizada, y busca garantizar el máximo beneficio del estudio minimizando el riesgo de daño a todas las personas involucradas (equipo de evaluación, participantes, etc.) y a lo largo de todo el ciclo de una evaluación (desde la planificación a la difusión de los resultados) (Economic and Social Research Council, 2015). Los juicios éticos suelen ser complejos, estar remitidos a circunstancias concretas y sujetos a diversas interpretaciones. Por tal motivo, no existen soluciones preestablecidas y universales para los dilemas éticos planteados por la investigación evaluativa.

Algunas organizaciones definen principios éticos en torno a lo que constituye un comportamiento “correcto” y establecen códigos de conducta para orientar el trabajo de evaluación (ver, por ejemplo, UNEG, 2008a, 2008b). Hay comunidades profesionales en las que es necesario someter la propuesta de evaluación a órganos oficiales de revisión, como los comités de ética, mientras que en otros casos no existen ámbitos formales, pero se recomienda la revisión de colegas para contar con un parecer externo e informado sobre la consistencia ética de los proyectos.

Los responsables de los programas, las evaluadoras y los evaluadores tienen una responsabilidad compartida de garantizar que una evaluación sea ética y que sus resultados no estén sesgados, sean fiables y creíbles. Lo contrario podría no solo invalidar la evaluación, sino poner en riesgo la integridad y el bienestar de las personas involucradas.

A pesar de que la evaluación plantea dilemas de difícil resolución, el propio hecho de evaluar puede considerarse desde una perspectiva ética. De hecho, no contar con evaluaciones posibilita que se malgasten recursos escasos (a menudo públicos) en aplicar intervenciones (ocasionalmente sobre poblaciones vulnerables) cuyos efectos se desconocen (a veces nulos o negativos). Ante la ausencia de evidencias sólidas, los programas de prevención son frecuentemente adoptados, continuados o interrumpidos en función de criterios subjetivos de los gobernantes o de acuerdo con su capacidad de lograr apoyo político o electoral, esto es, en función de su viabilidad para ser presentados al público como iniciativas exitosas. En países con economías débiles y niveles acuciantes de violencia letal, es imperioso que el destino de los recursos públicos esté guiado por evidencias rigurosas y que la calidad

del gasto sea la mayor posible. En este sentido, la propia evaluación puede ser entendida como una condición necesaria para la implementación ética de los programas.

7.2 Imparcialidad

La evaluación de programas no se desarrolla en el vacío, sino en el marco de actores sociales que representan intereses y mantienen entre sí relaciones de poder. Dichos actores suelen abrigar expectativas sobre los resultados de la evaluación y, en algunos casos, pueden intentar ejercer influencia para ajustarlos a su conveniencia. Este podría ser el caso, por ejemplo, de los gestores o las gestoras de un programa que necesitan que la iniciativa sea considerada eficaz para garantizar el financiamiento y la continuidad de los empleos. Contrariamente, otros actores podrían pretender que el programa se demuestre ineficaz, para provocar cambios en una intervención que creen políticamente inadecuada. Es importante reconocer de manera amplia la perspectiva de los diferentes actores sociales relevantes (*stakeholders*) y mantener un diálogo con ellos. En el caso de intervenciones para prevenir homicidios, por ejemplo, suele ser importante contemplar el punto de vista de los gestores y las gestoras del programa y de las autoridades públicas (como policías, fiscales y jueces), pero también el de familiares de las víctimas, vecinos de la comunidad y organizaciones sociales vinculadas al tema. En otras palabras, se deben analizar las perspectivas, los valores y los intereses de las distintas partes interesadas de manera imparcial e intentar que la evaluación responda, en la medida de lo posible, a estas perspectivas diferenciadas. Esta es una de las tareas de quien realiza la evaluación para evitar aproximaciones sesgadas e incompletas al programa evaluado.⁵⁶

La preocupación por la imparcialidad de una evaluación es de interés tanto para quienes evalúan, que tendrán expuesta su reputación profesional, como para quienes solicitan la evaluación, que necesitan confiar en la credibilidad de los resultados para orientar sus esfuerzos.

Por eso, es deseable que las evaluaciones sean conducidas de forma independiente y externa,⁵⁷ es decir, por un equipo que no tenga intereses económicos o políticos en los resultados de la evaluación. Un evaluador o una evaluadora independiente se encuentra en una posición imparcial, por lo que puede centrarse en el proceso de evaluación sin entrar en divergencias, enfrentamientos políticos o cualquier conflicto de interés que comprometa la evaluación. Dado que el rol de quien evalúa es conocer el impacto de una intervención, existe un conflicto de intereses cuando se podrían obtener beneficios personales o privados en función de los resultados de la evaluación, o cuando los beneficios pueden dirigirse al entorno o grupo al que pertenecen evaluador o

56 En la literatura sobre evaluación hay diferentes visiones sobre cuál debe ser el papel de los grupos de interés (*stakeholders*) en cada fase del proceso evaluativo: diseño, ejecución, interpretación, difusión y aplicación de los resultados. Este tema se discute con mayor profundidad en el capítulo 8.3.

57 Este tema es tratado en el [apartado 8.1](#).

evaluadora. La vigilancia ética de estos aspectos es fundamental para garantizar la objetividad, transparencia y credibilidad⁵⁸ del estudio. Los sesgos eventualmente introducidos no siempre son intencionales o evidentes, sino que pueden adoptar formas sutiles que afectan la toma de decisiones del equipo de evaluación (Eisner, 2009). Cuando no sea posible evitarlos, los conflictos de interés deben ser explicitados en los informes, así como en cualquier otro producto de la evaluación.

7.3 Consentimiento informado

Uno de los postulados éticos de los estudios con poblaciones humanas es que las personas participantes deben otorgar previamente su permiso, decidido de forma libre e informada, para ser parte del estudio. Comúnmente las evaluaciones de programas de prevención de homicidios son realizadas con datos que no requieren del contacto directo con personas para obtener información, de modo que la discusión sobre el consentimiento se vuelve irrelevante. Sin embargo, en caso de que se requieran datos primarios, será necesario el consentimiento informado. Al solicitar este consentimiento, quien evalúa necesita informar con claridad los objetivos del estudio, sus procedimientos, los beneficios y riesgos que implica, y también cómo se utilizarán las informaciones resultantes. A pesar de que el consentimiento haya sido conferido inicialmente, los y las participantes mantienen en todo momento el derecho a interrumpir su participación, sin que puedan existir imposiciones o sanciones en contrapartida.

Es frecuente que los comités de ética exijan términos individuales de consentimiento por escrito, los cuales deben ser firmados por las personas cuando aceptan participar del estudio.⁵⁹ Esos instrumentos, de uso extendido en el área biomédica, suelen ser cuestionados desde las ciencias sociales por su rigidez y formalidad (Fluehr-lobban, 1994). Además, cuando se entrevista a residentes en áreas dominadas por grupos criminales, portar documentos firmados por quienes fueron entrevistados puede poner a las personas en peligro, si esos términos de consentimiento caen en manos equivocadas. De hecho, es necesario extremar las precauciones para que la identidad las personas encuestadas sea preservada en relación con el propio contexto en que viven (Rojido y Cano, 2016).

7.4 Confidencialidad

La protección de la confidencialidad es una responsabilidad ética del equipo de evaluación, a fin de garantizar que no se revele la identidad de sus informantes. Sin embargo, existe la posibilidad de que las evaluadoras y los evaluadores sean presionados por diferentes grupos

58 La credibilidad está relacionada a la aceptación por parte de los grupos interesados del proceso y los resultados de evaluación.

59 El [formulario de consentimiento informado](#) de la Organización Mundial de la Salud puede servir como ejemplo de este tipo de instrumentos. Para una traducción al español: [aquí](#).

interesados para violar la promesa de confidencialidad que ofrecieron a quienes participan del estudio. La presión puede variar según el contexto político y social, para revelar directamente lo que una persona específica dijo o escribió, o para responder a solicitudes de datos que, analizados en conjunto, permitirían identificar indirectamente a ciertos individuos o grupos.

También existe la posibilidad de que los y las informantes teman ser identificados en la presentación de los resultados de la evaluación. Cuando el tamaño del universo es pequeño, por ejemplo, informar datos de acuerdo con las características demográficas y del programa puede permitir identificar individuos. Por ejemplo, puede que no haya demasiados hombres de 39 años con el oficio de panadero que, además, sean beneficiarios del programa. Por lo tanto, aunque el rigor técnico requiere que las fuentes estén correctamente indicadas, es importante que al publicar la evaluación se eviten detalles sobre el perfil de los informantes que permitirían identificar a estas personas.

Un tercer problema importante relacionado a la confidencialidad surge cuando el evaluador o la evaluadora se da cuenta de un comportamiento ilegal, peligroso o poco ético por parte del equipo de gestión o de las personas beneficiadas por el programa. Estos hallazgos generalmente ocurren durante la fase de recolección de datos y plantean un dilema entre los compromisos éticos de honrar la confidencialidad y proteger el interés público más amplio. Aunque no hay soluciones únicas para estas situaciones, es recomendable que quien hace la evaluación piense de antemano cuáles son los límites de la confidencialidad y los comunique a sus informantes al momento del consentimiento informado.

7.5 No causar daño

No dañar es uno de los principales postulados éticos de la evaluación. Los investigadores y las investigadoras deben intentar no perjudicar a las poblaciones estudiadas en la medida de lo posible y garantizar que sus derechos y bienestar sean preservados. Los cuidados generales deben extremarse cuando se trate de personas en situación de vulnerabilidad. Sin embargo, una evaluación no puede garantizar el principio general de no causar daños a la imagen de grupos sociales o instituciones, como sí tienen la obligación de hacer otros tipos de estudios sociales que pueden, por ejemplo, cambiar el nombre a los lugares estudiados en los informes para no generar estigmatización. Dado que la evaluación está pensada para orientar a los programas y las políticas públicas, un perjuicio a la imagen de grupos o instituciones puede llegar a ser inevitable.

7.5.1 Justicia en la distribución de riesgos y beneficios

Un aspecto a considerar es la justicia en la distribución de riesgos y beneficios durante la evaluación. Como se detalló en el capítulo anterior, la conformación aleatoria de los grupos experimental y de control es importante para poder determinar los efectos de una intervención. Sin embargo, puede surgir un problema ético cuando en el proceso de evaluación algunos individuos o grupos son seleccionados al azar para integrar el programa, privando al resto de beneficios potencialmente importantes.

Una opción para abordar este dilema es introducir gradualmente el programa a lo largo del tiempo, lo que, además, es necesario cuando no hay recursos suficientes para implementarlo de una sola vez en toda la comunidad, ciudad o país que se pretende cubrir. Cuando hay un grupo de personas que necesitan el programa en mayor medida que otros, existe la posibilidad de priorizar a este grupo y aleatorizar la selección dentro de otra subpoblación de menor necesidad. En cualquier caso, una estrategia de introducción progresiva permite que todas las personas tengan la misma probabilidad de ser seleccionadas para integrar los grupos experimental y de control, que nadie deba renunciar definitivamente a la intervención e incluso que quienes reciben el programa más tarde puedan beneficiarse con una versión mejorada en función de los datos recopilados en las fases anteriores (Gertler *et al.*, 2017). Dependiendo de las características del programa, también existe la opción de rotar a la población beneficiaria de modo que al final todos reciban la intervención. Además de satisfacer el criterio de justicia, la introducción del programa en diferentes momentos del tiempo puede dotar de mayor consistencia a la estrategia de evaluación, ya que permitiría conocer el impacto en distintos momentos del tiempo ([ver Capítulo 5](#)).

7.5.2 Protocolo de seguridad

En los casos en que la evaluación precise de datos primarios que deban ser recogidos en el terreno, es necesario considerar los riesgos de seguridad y elaborar un plan para gestionarlos, eliminarlos o mitigarlos. Los desafíos suelen ser mayores en contextos de violencia, donde el trabajo de campo puede comprometer a investigadoras e investigadores y, al mismo tiempo, acentuar la vulnerabilidad de las personas o grupos estudiados (Bush & Duggan, 2015). Para preservar la integridad de las personas y la consistencia del estudio se recomienda la elaboración de un protocolo que contenga procedimientos, precauciones y normas a seguir para maximizar la seguridad de todos los involucrados.

Este protocolo precisa ser elaborado en función de las especificidades de cada caso (tema abordado, técnicas de investigación, características locales, etc.) y ser objeto de una capacitación que incluya a todos los integrantes del equipo de investigación. Entre los temas a contemplar pueden mencionarse las estrategias para entrar y salir de campo, medidas a adoptar en casos de emergencia y procedimientos para gestionar las informaciones. A pesar de que este protocolo debe elaborarse al inicio de la evaluación, antes de la recogida de datos, es importante que sea revisado periódicamente a medida en que se avanza con el estudio. Una breve discusión sobre este tipo de herramientas puede encontrarse en Cano, Rojido y Aguilar (2018).

7.5.3 Revictimización

En situaciones en que los individuos investigados experimentaron trauma o dolor intenso derivado de la violencia sufrida, ya sea por sí mismos o por personas próximas a ellos, es imperativo que el investigador o la investigadora intente minimizar el riesgo de revictimización, es decir, de provocar que la víctima vuelva a sentir el dolor y el trauma mientras responde a las preguntas y relata lo sucedido. En ese sentido, las entrevistas a familiares de víctimas fatales de la violencia necesitan ser abordadas con sumo cuidado por investigadoras e investigadores capacitados y experimentados, considerando también la posibilidad de interrumpirlas en el caso de advertir signos de sufrimiento intenso. Asimismo, en la medida de lo posible, las investigadoras y los investigadores podrían buscar asistencia psicológica para las víctimas y sus familias. En verdad, la evaluación debe considerar, antes de comenzar, una estimación de costo-beneficio de la propia investigación, con el fin de decidir si los beneficios potenciales derivados de ella (no solo para el equipo de investigación, sino para la sociedad en general y para los propios sujetos del estudio) son mayores que los costos que puede generar en las personas participantes. Este análisis de costo-beneficio puede sugerir la búsqueda de otras fuentes de datos para la evaluación, aunque dichas alternativas no siempre son posibles.

7.5.4 Estigmatización

Otro aspecto a evitar es la estigmatización, lo que implica definir a las personas en función de un atributo que las tornaría parte de una categoría social inferior o menospreciada. En materia de prevención de la violencia, la estigmatización puede producirse al enfatizar sistemáticamente ciertas características de los agresores o las víctimas, como su raza, nivel socioeconómico o pertenencia a determinados barrios de la ciudad. Se produce así una asociación que podría conducir

a la falsa idea de que la mera presencia de estas características (ser negro, pobre o residente del barrio x) es un indicador de violencia.

La alta concentración social y geográfica del homicidio, examinada en el [apartado 1.2](#), puede propiciar este tipo de falacia. Por otro lado, el propio hecho de ser beneficiario o beneficiaria de un programa puede ser estigmatizador, cuando, por serlo, se tiene un acceso restringido a reconocimiento y oportunidades sociales. Esto sucede, por ejemplo, cuando ciertos certificados laborales son otorgados en prisión (de carpintería, panadería, etc.) para facilitar la reinserción social de las personas que luego son rechazadas por los empleadores en el mercado de trabajo. Es deber del equipo de investigación hacer lo posible por no fomentar estos procesos de estigmatización, ya que, además de inducir a distorsiones de la realidad, pueden comprometer el bienestar de las personas o grupos implicados. Sin embargo, como ya fue mencionado, la evaluación no puede garantizar que la imagen de ciertos grupos o instituciones no sea afectada negativamente por los resultados.

7.6 Publicación

Una vez que se han recopilado y analizado los resultados de la evaluación, es importante compartirlos con los y las responsables y las entidades financiadoras del programa, así como con los diversos actores sociales interesados (*stakeholders*), de la manera más abierta y directa posible ([ver Capítulo 9](#)). Cuando los resultados no son muy positivos, lo que dependerá en parte de las expectativas previas, la difusión de la evaluación a gran escala puede ser un tema delicado. No es raro, en esta situación, que gestores y agencias financiadoras prefieran mantener en secreto el informe de evaluación y traten de publicar solo un resumen políticamente aceptable de los resultados del estudio. Ya sean positivos o negativos, estos resultados deben discutirse con los actores estratégicos para que se puedan aprender las lecciones y hacer las adaptaciones necesarias. Aunque el uso de la evaluación no depende principalmente del equipo de evaluación, este debe esforzarse para que los resultados de su trabajo sean publicados. Es recomendable acordar entre las partes una estrategia de difusión al inicio de la evaluación, cuando todavía se desconocen los resultados.

Los resultados de la evaluación deben presentarse de la forma más clara posible. Dado que las evaluaciones suelen tener audiencias múltiples, es común elaborar diversos documentos de difusión con los estilos de comunicación más adecuados para cada caso. Es particularmente importante que los resultados estén disponibles para las personas que colaboraron con la investigación. No obstante, en contextos de violencia una devolución de resultados puede no ser deseable e incluso colocar

a las personas en situación de mayor vulnerabilidad. Por tal motivo, es aconsejable que el método de devolución sea pensado de antemano y acordado con las personas implicadas.

Por último, como se mencionó antes, las publicaciones deben preservar la confidencialidad y la seguridad de las personas, evitando el uso de etiquetas que refuercen los prejuicios hacia los grupos sociales que son objeto del estudio.

RESUMEN

Desafíos éticos de una evaluación

La ética se refiere a una práctica correcta o acordada dentro del contexto social y cultural en el que la evaluación es realizada.

Los y las responsables de los programas y los equipos de evaluación tienen una responsabilidad compartida de garantizar que una evaluación sea ética y que sus resultados no estén sesgados, sean fiables y creíbles. Lo contrario podría no solo invalidar la evaluación, sino poner en riesgo la integridad y el bienestar de las personas involucradas.

Para propiciar la imparcialidad, las evaluaciones deben ser conducidas de forma independiente y externa, es decir, por un equipo de evaluación que no tenga intereses económicos o políticos en los resultados.

Comúnmente las evaluaciones de programas de prevención de homicidios son realizadas con datos secundarios, pero cuando se requieren datos primarios, es necesario el consentimiento informado de los participantes.

La protección de la confidencialidad de las personas es una responsabilidad ética del equipo de evaluación, a fin de garantizar que no se revele la identidad de sus informantes.

No dañar es uno de los principales postulados éticos de las investigaciones con seres humanos. Los evaluadores y las evaluadoras deben intentar no perjudicar a las poblaciones estudiadas en la medida de lo posible, aunque a veces el daño reputacional a instituciones o grupos puede ser inevitable.

Planificar una justa distribución de riesgos y beneficios entre las personas participantes, contar con protocolos de seguridad, y evitar la revictimización y la estigmatización son algunos de los aspectos a considerar.

Una vez que se han recopilado y analizado los resultados de la evaluación es importante compartirlos con responsables y agencias financiadoras del programa, así como con los diversos actores sociales interesados (*stakeholders*), de la manera más abierta y directa posible.

CAPÍTULO 8. ¿QUIÉN HACE LA EVALUACIÓN?

8.1 Evaluación interna vs. externa

Una *evaluación interna* es la realizada por parte del equipo gestor del programa, mientras que la *evaluación externa* la llevan a cabo personas ajenas a gestores y gestoras. Por ejemplo, la evaluación externa la pueden realizar especialistas, institutos de investigación o empresas de consultoría contratadas al efecto. La gran diferencia es que la evaluación externa la ejecutan personas cuya única responsabilidad en relación con el proyecto es, justamente, la evaluación y no tienen ninguna participación en el diseño o gestión del programa ni tampoco responsabilidades políticas en relación a este. La idea central es que el futuro profesional o político del equipo de evaluación no debe depender del resultado de la evaluación, de forma que, al final de esta, evaluadoras y evaluadores continúan su vida profesional ajenos al programa. Sin embargo, es importante no ser ingenuos, porque a menudo hay otras formas de condicionamiento profesional a quien realiza la evaluación. Por ejemplo, en la medida en que los resultados de la evaluación sean positivos, puede ser más probable que sean contratados de nuevo por las agencias financiadoras o por los cargos públicos para evaluar otros programas. Por ello, hay que tomar cuidado siempre para preservar la independencia de evaluadoras y evaluadores, incluso cuando son externos.

Las *ventajas de una evaluación externa* son las siguientes:

- Como ya fue explicado, un equipo de evaluación externo no posee en principio un interés personal, institucional o político en las conclusiones de la evaluación, por lo que puede desarrollar su trabajo de manera más independiente y neutral.
- Al ser externos al programa, evaluadoras y evaluadores tienen también mayor probabilidad de ser percibidos como neutrales por parte de todos los sectores afectados por él (gestores y gestoras, población beneficiaria, instituciones financiadoras, sociedad civil, etc.) y, por tanto, de que la evaluación sea considerada como legítima.
- Las evaluadoras y los evaluadores externos pueden tener mayor conocimiento técnico sobre evaluación y más experiencia en evaluación que quien gestiona el programa o sus técnicos y técnicas. A veces, el equipo de evaluación posee también un conocimiento profundo sobre el área del programa, en nuestro caso, sobre prevención de homicidios, pero ello no siempre es así.

- Las evaluadoras y los evaluadores externos pueden ofrecer una perspectiva nueva sobre el programa, que no esté presa a la rutina o al consenso interno que se haya formado al interior del programa.

Las posibles *desventajas de una evaluación externa* serían las siguientes:

- Contratar a un equipo evaluador externo puede demorar más tiempo y aumentar los costos en relación con una evaluación interna, aunque en este cálculo comparativo es preciso incluir el coste del tiempo que el equipo técnico le dedicaría a la evaluación. En algunos casos, la dedicación a la evaluación podría inclusive comprometer la ejecución normal de las actividades del programa.
- Las evaluadoras y los evaluadores externos pueden no tener una comprensión plena de las cuestiones pertinentes para el programa o para población destinataria, o de la dinámica de su funcionamiento. Del mismo modo, pueden no estar suficientemente familiarizados con los datos o las fuentes que el programa utiliza. En este sentido, un equipo evaluador experimentado tiene como desafío escuchar atentamente a todos los actores para minimizar estos riesgos.
- La confidencialidad de algunas informaciones, que no podrían ser difundidas fuera del proyecto. Obviamente, una evaluación precisa de acceso a todas las informaciones relevantes. En estos casos, lo recomendado es elaborar un protocolo de confidencialidad que obligue legalmente al equipo de evaluación a mantener, posteriormente, la confidencialidad de las informaciones sensibles a las que hayan tenido acceso, o que obligue a eliminar algunas de estas informaciones de los bancos de datos al final del trabajo. El caso extremo de confidencialidad es, por ejemplo, la evaluación de un programa de protección a personas amenazadas de muerte, cuya identidad es secreta, incluso por imperativo legal, en muchos países. En estos casos, la agencia de protección puede proporcionar a evaluadoras y evaluadores una lista de personas beneficiarias con códigos en vez de nombres y facilitar la entrevista con estas personas sin que el equipo de evaluación sepa su identidad real o su lugar de residencia.
- Mayor dificultad para que el resultado de la evaluación sea usado por gestoras y gestores y por el equipo técnico del programa, una

vez que se origina en una fuente externa. Ante esto, es importante concienciar a los integrantes del programa sobre la importancia de la evaluación y de que esta sea utilizada de forma apropiada ([ver Capítulo 9](#)).

Las *ventajas y desventajas de una evaluación interna* son el reflejo especular de las ya citadas para la evaluación externa. Por un lado, es una evaluación con riesgos de parcialidad, en la medida en que de sus resultados depende la vida profesional de gestoras y gestores. Por más que una persona se considere imparcial y ponderada, es imposible esperar que sus intereses y su experiencia en la institución no afecten a los resultados de la evaluación. En general, una evaluación desarrollada por las gestoras y los gestores, las técnicas y los técnicos, tenderá a ser benevolente con los fallos y a ignorar otros puntos de vista. Nadie puede ser el mejor evaluador de sí mismo. Además, es raro que los integrantes de un programa tengan conocimientos especializados y experiencia en el área de evaluación. Por otro lado, la evaluación interna puede ser más rápida y menos costosa (aunque, como fue mencionado, hay que considerar también los costos del tiempo dedicado por el equipo para llevar a cabo la evaluación), garantiza en mayor medida que los puntos de vista del equipo técnico serán incorporados en el esfuerzo evaluativo y puede promover la aplicación de los resultados.

Colocando en una balanza los pros y los contras de los dos tipos de evaluación, el saldo es inequívoco: es necesario hacer evaluaciones externas, por lo menos cada cierto tiempo. Las evaluaciones internas son complementarias a las externas, pero no las sustituyen. De lo contrario, la imparcialidad y la credibilidad del proceso evaluativo permanecerán bajo sospecha. Pero sí es posible y recomendable que, entre las evaluaciones externas, realizadas periódicamente a cada ciclo, existan evaluaciones internas con metodologías parecidas, que puedan incluso alimentar a las primeras.

El hecho de que la evaluación sea externa no significa que los y las participantes en el programa no tengan un papel fundamental. Las gestoras y los gestores, las técnicas y los técnicos deben ser escuchados y sus posiciones sobre el diseño de la evaluación, la recogida de informaciones y el análisis deben ser cuidadosamente tomadas en cuenta por el equipo evaluador. Inclusive, pueden participar directamente en la recogida de informaciones, según el caso. Lo esencial para mantener la independencia de una evaluación externa es que la última palabra sobre el diseño de la evaluación, el análisis de las evidencias y las conclusiones correspondan a personas ajenas al programa.

Tabla 2. Virtudes y riesgos de las evaluaciones externas

	Virtudes	Riesgos
Evaluación externa	<p>Es probable que la evaluación externa sea más objetiva, ya que el equipo de evaluación tendrá independencia en relación al programa.</p> <p>Los evaluadores y las evaluadoras deben tener una variedad de habilidades de evaluación y experiencia.</p> <p>A veces la gente está más dispuesta a hablar con personas de afuera de la organización que con las de adentro.</p> <p>El uso de un equipo evaluador externo otorga mayor credibilidad a los hallazgos, particularmente a los positivos.</p>	<p>Alguien externo a la organización o proyecto puede no entender la cultura o incluso lo que el trabajo está tratando de lograr.</p> <p>Las personas directamente involucradas pueden sentirse amenazadas por personas externas y es menos probable que hablen abiertamente y cooperen en el proceso.</p> <p>La evaluación externa puede ser muy costosa.</p>

Fuente: Elaboración propia.

A veces se habla también de *evaluación mixta*, como aquella en cuyo equipo de evaluación se cuentan integrantes internos y otros externos, por ejemplo, con la mitad de personas elegidas por gestores y gestoras y la otra mitad escogidas por la agencia financiadora. En estas circunstancias, si los nominados por los ejecutores quisieran sesgar el resultado, los otros podrían denunciarlos. Por otra parte, la presencia de aquellas garantiza que la visión del equipo de gestión será también contemplada. En la práctica, ambos equipos estarían interesados en llegar a un acuerdo, pues una discrepancia permanente entre ambas partes no le interesaría a nadie.

De cualquier forma, si la evaluación es ejecutada por un equipo mixto, es preciso estipular con claridad quién tiene la responsabilidad última de diseñar la evaluación y elaborar el informe final.

8.2 Elementos a ser tomados en cuenta a la hora de contratar una evaluación externa

Un equipo de evaluación externa debería cumplir los siguientes requisitos:

- Competencia técnica y experiencia en el tipo concreto de evaluación.
- Credibilidad y fiabilidad.

- Independencia y ausencia de vínculos con el programa.
- Capacidad de relacionarse con los diferentes actores sociales con interés en el proyecto (*stakeholders*).
- El conocimiento técnico sobre el área temática. En este caso, sobre programas de prevención de violencia o, mejor aún, de prevención de homicidios es una clara ventaja, pero no siempre es posible encontrar un perfil tan específico. En la duda, es mejor contar con experiencia en evaluación de programas sin el conocimiento de prevención de violencia que lo contrario: experiencia con el tema de violencia, pero sin conocimientos sobre evaluación.

Más allá de los requisitos, será necesario considerar factores como el costo y el tiempo necesarios para realizar la evaluación, a la hora de escoger el equipo de evaluación.

Una vez que se ha elegido a una evaluadora o a un evaluador externo, ya sea un especialista, una institución o una consultora especializada, será preciso negociar un *plan de evaluación* específico, que debe contener, entre otros, los siguientes elementos:

- 1) Los objetivos de la evaluación: qué tipo de evaluación se realizará, a qué preguntas se pretende responder y qué variables serán evaluadas, incluyendo las dimensiones de posibles impactos no intencionales.
- 2) El diseño metodológico de la evaluación, que estipule la estrategia básica de la evaluación para responder a las preguntas anteriormente mencionadas.
- 3) Cuáles serán las informaciones que serán recogidas, entregadas al equipo evaluador y analizadas, qué fuentes serán utilizadas y quién repasará todas estas informaciones en el equipo de evaluación.
- 4) El cronograma de la evaluación, incluyendo fases, entregas de informes parciales y finales, y pagos. Es posible que este cronograma deba luego ser modificado en función de las condiciones reales, pero es importante partir de una propuesta inicial.
- 5) El costo y el presupuesto de la evaluación, las formas y los montos de cada pago.
- 6) Qué productos deberán ser entregados por el equipo evaluador: informe final, informes parciales, bancos de datos, materiales audiovisuales, etc. Además, debe quedar claro de quién será la propiedad final de estos productos.

- 7) Cómo serán los contactos entre el equipo ejecutor del programa y el equipo evaluador (por ejemplo, quienes serán los encargados de esta relación por ambas partes) y quiénes entregarán las informaciones relevantes.
- 8) Cómo serán publicados los resultados de la evaluación, quién o quiénes recibirán los informes o las otras formas de entrega de resultados, y en qué momento. Es muy importante que evaluadoras y evaluadores comprometan a la institución financiadora con la divulgación amplia del resultado y con la difusión a todos los actores sociales que tienen un interés legítimo en este ([ver apartado 7.4](#)). Es razonable que la entidad financiadora o quienes gestionan el programa exijan ser informados de los resultados en primer lugar y, en ocasiones, que demanden un tiempo de exclusividad (para tomar medidas, para preparar sus respuestas ante el público, etc.) antes de hacerlos públicos de forma amplia. Sin embargo, no es razonable y constituye un serio problema ético el hecho de que la agencia financiadora espere que el resultado de la evaluación permanezca confidencial indefinidamente o que condicionen su difusión al tipo de resultados que se obtengan.
- 9) Cuáles serán los criterios técnicos de aceptación de los productos y quiénes tomarán esta decisión. En las consultorías técnicas, en general, es común que los pagos sean condicionados a la aceptación de la calidad de los resultados y a la confirmación de que responden específicamente a lo que fue contratado. Sin embargo, todo esto, siendo razonable, es mucho más delicado en el caso de evaluaciones, pues podría existir la tentación de rechazar el informe final si no corresponde con las expectativas del financiador. Por eso, los parámetros de aceptación del resultado final deben proteger, por un lado, a la agencia financiadora contra un trabajo mal realizado o deficiente, pero también deben proteger, por otro, al equipo de evaluación contra eventuales rechazos de su trabajo en función de si el resultado final de la evaluación es considerado favorable o no para el proyecto.

Es esencial que todos esos puntos hayan sido acordados y aceptados por la entidad financiadora, el equipo de evaluación, gestoras y gestores del programa, para que no surjan discrepancias más adelante. En la medida de lo posible, todo este plan de evaluación debe reflejarse en los contratos firmados entre quien evalúa y quien financia el trabajo. No obstante, siempre pueden surgir obstáculos o novedades en el camino que hagan aconsejable alterar algunos de sus puntos, principalmente el cronograma por posibles atrasos, pero también las dimensiones a ser estudiadas e incluso, eventualmente, los objetivos de la evaluación. Estas alteraciones deben ser realizadas de común acuerdo entre quienes

acordaron el plan de evaluación, pero este debe contemplar desde el principio algún grado de flexibilidad que permita acomodar estos cambios sin romper los compromisos contractuales y algún mecanismo para decidir estos cambios.

8.3 Participación social en la evaluación

Como ya ha sido mencionado en diversas ocasiones, existen siempre distintos actores sociales (*stakeholders*) que se ven afectados por una evaluación, como gestoras y gestores, el equipo técnico, la población beneficiaria, las víctimas y sus familias, las agencias financiadoras, los gobiernos, la sociedad en general, etc. El impacto que el programa presenta para cada uno de ellos debería ser contemplados en la evaluación.

Sin embargo, el poder que cada actor posee sobre la definición de la evaluación es muy diferente. Las entidades financiadoras, por supuesto, tienen un poder muy amplio para determinar el curso de la evaluación, ya que pagan por ella y pueden demandar ciertos objetivos o métodos. Los cargos públicos que deciden sobre los programas, que pueden o no ser quienes financian la evaluación, también suelen poseer gran influencia. Además, es común que las opiniones del equipo técnico, gestoras y gestores sean tomadas en cuenta. En consecuencia, cuando se piensa en participación social, se está pensando en actores sociales cuya visión no es necesariamente contemplada en la evaluación. Entre ellos, en nuestro caso, podríamos citar las autoridades locales o comunitarias, las familias de las víctimas, los grupos sociales que componen el perfil de alto riesgo o la sociedad en general.

La pregunta clave es en qué medida es posible y deseable la participación de estos grupos en la evaluación. No se debe olvidar que la visión de todos estos grupos, como la de cualquier actor social afectado, debería ser necesariamente contemplada como parte de la evaluación, indagando, por ejemplo, su percepción del programa, los impactos no intencionales de este, las necesidades no cubiertas y cómo el resultado de la evaluación les puede afectar. La cuestión aquí, sin embargo, va más allá, pues se interroga sobre en qué medida pueden tener una participación más activa, no apenas como sujetos pasivos u objetos de la investigación, que ayude a determinar la propia evaluación. En última instancia, ¿pueden estos grupos ser, en parte al menos, evaluadores?

Este no es un tema sobre el que exista acuerdo en la literatura, sino que, al contrario, genera bastante polémica. Algunos profesionales del área de evaluación defienden un modelo en el que los grupos sociales son consultados sobre intereses y necesidades, pero no tienen ninguna participación en la evaluación, que es considerada un ejercicio técnico.

Otros, en cambio, proponen que esos grupos deben ser protagonistas de la evaluación desde el principio, influyendo y participando en todas las etapas. Su idea es intentar disminuir el desequilibrio de poder, por un lado, entre los que toman las decisiones y los que son más afectados por ellas, y, por otro, entre estos últimos y el propio equipo de evaluación. En este sentido, no hay duda de que la evaluación es también un proceso de aprendizaje del cual, idealmente, tanto los diversos actores sociales como evaluadoras y evaluadores saldrán beneficiados. Para estos últimos, inclusive, el objetivo central de la evaluación debe ser empoderar a los grupos sociales con menor poder (Fetterman, 1994). De cualquier forma e independientemente de la posición que tome cada evaluador o evaluadora, no se debe ser ingenuo y pensar que una evaluación por sí misma puede alterar el escenario de poder en una sociedad.

En suma, no hay recetas universales sobre qué grado de participación social es deseable en una evaluación y la decisión final dependerá en buena parte del contexto, de las circunstancias y del perfil del equipo de evaluación.

De todas maneras, a continuación, se presenta un análisis somero de las oportunidades y los límites de esta participación en cada una de las fases de la evaluación:

- 1) *Objetivos de la evaluación.* Como la designación de los objetivos es un juicio de valor y no de hecho, la participación de los diversos sectores sociales no solo es posible sino deseable, para que la evaluación responda a los intereses de todos ellos. Eso puede hacerse desde el inicio, mediante algún órgano de representación que decida sobre la evaluación, o lo pueden hacer los propios evaluadores y evaluadoras mediante consultas antes de iniciar su trabajo. Sin embargo, la idea de la investigación-acción, que existe en otras áreas de investigación social, no se corresponde con la tarea de la evaluación, pues no es posible intentar evaluar algo que, al mismo tiempo, se pretende transformar. La transformación puede suceder a partir de los resultados de la evaluación, pero no como objetivo de esta.
- 2) *Diseño de la evaluación.* Este es el aspecto más técnico, por lo que la decisión final debe corresponder al equipo de evaluación, que no solo está más cualificado para hacerla, sino que es, en teoría, independiente. Todos los sectores pueden ser escuchados, pero no es en absoluto recomendable que se cedan las decisiones técnicas a uno u otro grupo.
- 3) *Recogida de informaciones.* En esta fase, la participación de diversos sectores es perfectamente posible, siempre que ello no coloque en riesgo la neutralidad del proceso. Una posibilidad frecuente es entrenar a residentes locales en la comunidad

para que ellos trabajen como entrevistadores o investigadores a servicio de la evaluación. Esto significa que, además de recibir alguna renta y formación, estos residentes se implicarán en mayor medida en el proceso evaluativo, lo que aumenta las posibilidades de utilización de los resultados.

- 4) *Análisis de datos*. Este es un paso también eminentemente técnico y muy sensible desde el punto de vista político. Por tanto, una vez más debe ser el equipo de evaluación, a partir de su competencia técnica y su independencia, el que tome las decisiones en este campo. Ello no obsta la posibilidad de escuchar las opiniones y sugerencias de diversos sectores, sin renunciar nunca a la decisión final.
- 5) *Conclusiones*. Este es el momento probablemente más delicado, porque incluso si hay consenso sobre el análisis, la forma en que las conclusiones son formuladas puede generar diferencias y controversias. Por ello, de nuevo evaluadores y evaluadoras tienen que tener la última palabra. Pero es muy recomendable que circulen versiones preliminares de las conclusiones entre los diversos grupos sociales y que sus reacciones sean escuchadas con atención antes de redactar la versión final ([ver Capítulo 9](#)).

Independientemente de la fase en cuestión, es de gran importancia que el equipo de evaluación escuche y les otorgue un papel a los sectores que puedan ser críticos con el programa y no solo a aquellos que lo apoyan o se benefician de él.

Por último, cabe recordar que, en su contacto con los diversos actores sociales afectadas, evaluadores y evaluadoras deben tomar cuidado en usar un lenguaje y un abordaje que sea comprensible y adecuado a cada uno de ellos. No es obligación de los diversos sectores sociales adaptar su forma de comunicarse a los integrantes del equipo de evaluación, sino al contrario; son estos últimos quienes deben adaptar su estrategia de comunicación para adecuarla a cada uno de sus interlocutores.

¿Quién hace la evaluación?

Evaluación interna: es realizada por parte del equipo gestor del programa.

Evaluación externa: la llevan a cabo personas ajenas al programa.

Evaluadoras y evaluadores externos:

- No poseen en principio un interés personal, institucional o político en las conclusiones de la evaluación.

RESUMEN

- Tienen también mayor probabilidad de ser percibidos como neutrales.
- Suelen tener mayor conocimiento técnico en el campo de la evaluación.
- Pueden ofrecer una perspectiva nueva sobre el programa.

El papel de la participación social en el proceso evaluativo:

- Algunos evaluadores y evaluadoras defienden un modelo en el que los grupos sociales son consultados sobre intereses y necesidades, pero no tienen participación en la evaluación propiamente dicha, que es considerada un ejercicio técnico.
- Otros y otras profesionales proponen que esos grupos deben ser protagonistas de la evaluación desde el principio, influyendo y participando en todas las etapas.

CAPÍTULO 9. RESULTADOS Y USOS DE LA EVALUACIÓN

Una evaluación que no se utiliza carece de sentido. Sin embargo, al planificar una evaluación no siempre la cuestión del uso es cuidadosamente contemplada. En una encuesta realizada en 2006 a 1140 miembros de la American Evaluation Association, el 68% de las personas consultadas indicó la no utilización de los resultados como uno de los problemas más importantes de su campo (Patton, 2008). Datos como ese deben alertar a agencias financiadoras, gestores y gestoras, y evaluadores y evaluadoras sobre la necesidad de prestar mayor atención al uso de las evaluaciones.

En sentido estricto, el uso de una evaluación depende de las personas responsables del programa, ya que son quienes decidirán, en última instancia, en qué momento, grado y forma se van introducir cambios en su intervención. Con todo, cabe hacer dos observaciones en relación con este punto. La primera, es que el evaluador o la evaluadora puede propiciar en mayor o menor medida el uso de la evaluación, entre otras cosas, previéndolo en el proceso y comunicando adecuadamente los resultados de su trabajo. Tanto es así que hay quienes defienden el concepto de una evaluación focalizada en resultados (Patton, 1986). La segunda, es que la evaluación puede ser utilizada por los distintos grupos de interés (*stakeholders*) involucrados, de modo que en realidad no existe una usuaria o usuario único de la evaluación.

La evaluación también puede servir a varios propósitos. El primero y más evidente es ayudar a la toma de decisiones para mejorar la eficacia y eficiencia de los programas, y con ello, el bienestar de las personas. Una buena evaluación proporciona información rigurosa y relevante sobre los efectos de una intervención, señalando, en la medida de lo posible, los componentes que requieren de modificaciones.⁶⁰ Esto generalmente orienta la introducción de cambios, como expandir el programa, hacer pequeños ajustes o modificaciones sustanciales, como la propia interrupción de la iniciativa (ver apartado 3.1). Dichas modificaciones dependerán tanto de los resultados de la evaluación como del uso que hagan de estos las diferentes partes involucradas (gestores y gestoras del programa, agencias del poder público, sociedad civil, etc.).

Por otro lado, la evaluación también tiene un uso político, por ejemplo, cuando se utiliza para rendir cuentas a la ciudadanía de una iniciativa desarrollada por el gobierno (lo que en inglés se conoce como *accountability*). Asimismo, la evaluación podría utilizarse como un

60 Como ya se señaló en el [apartado 4.2](#), la evaluación es formativa cuando se realiza durante la intervención con el propósito de mejorarla, y sumativa cuando se hace después de la intervención, para determinar si esta debe continuar.

pretexto para diferir una decisión que no se quiere tomar o para justificar retrospectivamente decisiones que fueron tomadas por otros motivos. La evaluación podría servir a diversos fines, inclusive contradictorios entre sí, pero la obligación de todas las partes implicadas es que las conclusiones, propuestas y decisiones estén basadas en datos objetivos. Por encima de las distintas perspectivas e intereses en juego, las evaluadoras y los evaluadores deben intentar que sea eso lo que suceda.

El uso efectivo de una evaluación no ocurre por accidente o casualidad y, si bien excede la atribución del equipo evaluador, este debe propiciarlo de varias maneras: planificando la entrega de resultados en función de los procesos de toma de decisiones, involucrando a los grupos de interés (*stakeholders*) y efectuando recomendaciones específicas, basadas en evidencias, realistas y orientadas a la acción.

El compromiso de divulgar de forma adecuada los resultados de la evaluación a todas las partes interesadas es una cuestión ética ([ver Capítulo Z](#)), pero también un elemento que maximiza las probabilidades de su utilización. Lo ideal es que exista un *plan de divulgación* pensado desde el inicio de la evaluación, con objetivos claros. Allí debe establecerse qué, cuándo, cómo y con quién se debe compartir la información, teniendo además en cuenta cómo se podría recibir, interpretar y utilizar en cada caso. Evidentemente, será necesario asignar recursos para que dicho plan pueda implementarse.

Uno de los aspectos principales es definir usuarios y usuarias potenciales de la evaluación, que pueden incluir a las agencias financiadoras, la dirección del programa, la población beneficiaria, la localidad donde el programa fue aplicado, las autoridades públicas, la comunidad académica, organizaciones sociales interesadas en el tema y la sociedad en general. Solo una vez pensadas las audiencias, cabe preguntarse qué informaciones podrían encontrar útiles y cuál sería la mejor manera de comunicarlas.

La publicación de los resultados de la evaluación, sobre todo cuando son poco auspiciosos, puede ser un tema sensible para las instituciones financiadoras y para las personas responsables de los programas. En el área de reducción de los homicidios este problema es particularmente delicado, dada la importancia del tema. De hecho, no es raro que ante esta situación los y las responsables del programa prefieran mantener el informe en secreto o publicar apenas un resumen políticamente aceptable. El equipo de evaluación, por el contrario, deben velar para que su trabajo sea publicado de manera íntegra o, al menos, equilibrada. Para evitar conflictos es recomendable *que se acuerde la difusión al inicio de la evaluación*, cuando los resultados aún son desconocidos. En este sentido, es de suma importancia que las personas responsables del

programa asuman el compromiso de publicar los resultados, aunque ello sea hecho después de un tiempo necesario (por ejemplo, dos o tres meses) para introducir las medidas correctivas que consideren pertinentes. Con todo, este tiempo no podría extenderse demasiado, dado que la oportunidad determina la utilidad de la información para otros usuarios y usuarias. Además, la capacidad de quien hace la evaluación para demandar la divulgación de los resultados tiende a decrecer con el pasar del tiempo.

La presentación de los resultados de la evaluación puede adoptar múltiples formas: informes escritos detallados, resúmenes ejecutivos, artículos, audiovisuales, folletos, presentaciones con diapositivas, etc. Los canales de comunicación también pueden ser diversos: sitios web, listas de correo electrónico, periódicos, radio y televisión, revistas, eventos y seminarios, etc. Es recomendable que los canales o espacios utilizados para la divulgación sean simbólicamente neutros, es decir, que no estén asociados apenas a algunas perspectivas o grupos de interés. Suele ser efectivo combinar formas y canales de comunicación en función de los usuarios y usuarias, quienes también determinan la selección de contenidos y el tono de la comunicación (obviamente, sin alterar los resultados), que puede ser, por ejemplo, más o menos técnico.

Con todo, el informe escrito suele ser el producto principal de la evaluación. Este debe dar cuenta de todo lo requerido en los términos de referencia e incluir al menos las siguientes secciones: una descripción del programa estudiado, los objetivos de la evaluación, la metodología (fuentes, diseño, técnicas de análisis y limitaciones), los resultados, las conclusiones y las recomendaciones. Este informe es fundamental para la transparencia de la propia evaluación, en la medida en que permite que otras personas puedan valorar la calidad del proceso evaluativo. Sirve, a su vez, para elaborar otros materiales, como los resúmenes ejecutivos, que facilitan la difusión al ofrecer apenas una síntesis del trabajo.

La versión preliminar del informe debe circular entre los sectores interesados (*stakeholders*), para obtener su visión sobre él antes de elaborar la versión final. Los evaluadores y las evaluadoras deben escuchar las sugerencias recabadas como un insumo para enriquecer el informe y considerar la posibilidad de hacer correcciones de última hora. Esto cumple varios objetivos importantes: a) disminuye errores, b) garantiza que la visión de cada grupo será tomada en cuenta, c) maximiza la posibilidad de que los resultados sean utilizados, d) contribuye a anticipar las reacciones que despertará la publicación de los resultados y las eventuales polémicas vinculadas.

Alentar la participación social durante la evaluación puede favorecer la receptividad y el uso de los resultados, ya que de algún modo se habrán atendido las demandas de las partes implicadas.

Más allá de la calidad y pertinencia de las recomendaciones producidas por la evaluación, uno de los mayores obstáculos para su aplicación suele ser la falta de un sistema de incentivos que refuerce positivamente las conductas deseadas y negativamente las que se pretenden desalentar. Dichos incentivos pueden ser materiales o simbólicos, individuales o grupales, pero, en cualquier caso, son un mecanismo fundamental para el cambio organizacional. Aunque el diseño de un sistema de incentivos trasciende las atribuciones comunes de un evaluador o una evaluadora, es deseable que las recomendaciones sugieran incentivos que podrían emplearse para generar los cambios pretendidos. Por ejemplo, además de sugerir una mejora en el registro de los datos en el sistema de informaciones del programa, el equipo evaluador podría proponer tentativamente establecer controles periódicos de calidad en ese sistema, que servirían para premiar con una compensación económica a gestoras y gestores que realicen adecuadamente el registro de los datos. De manera complementaria, los individuos que no consigan la compensación podrían ser encaminados a jornadas de capacitación. De esta forma, el equipo evaluador no solo estaría orientando a los tomadores de decisiones sobre *qué se debería hacer* para mejorar el programa, sino que ofrecería, además, ideas preliminares acerca de *cómo se podría hacer*. Este tipo de sugerencias no supone mucho trabajo adicional para el equipo de evaluación y puede redundar en una mayor probabilidad de uso de los resultados.

RESUMEN

Resultados y usos de la evaluación

Agencias financiadoras, gestores y gestoras, y evaluadores y evaluadoras deben propiciar el uso de la evaluación. Entre otras cosas, esto implica planificar la entrega de los resultados en función de los procesos de toma de decisiones, involucrar a los grupos de interés (*stakeholders*) y efectuar recomendaciones específicas, basadas en evidencias, realistas y orientadas a la acción.

Lo ideal es que exista un plan de divulgación pensado desde el inicio de la evaluación, determinando qué, cuándo, cómo y con quién se debe compartir la información. Para evitar conflictos es recomendable que se acuerde la difusión al inicio de la evaluación, cuando los resultados aún son desconocidos.

La evaluación puede servir a diversos propósitos, y usuarios y usuarias, pero la obligación de todas las partes es que las conclusiones, propuestas y decisiones estén basadas en datos objetivos.

Suele ser efectivo combinar distintas formas y canales de comunicación en función del perfil de las posibles audiencias.

La versión preliminar del informe de evaluación debe circular entre los sectores interesados (*stakeholders*) para obtener su visión sobre él antes de elaborar la versión final. En general, alentar la participación social durante la evaluación puede favorecer la receptividad y el uso de los resultados.

Es preferible que las recomendaciones estén vinculadas a un sistema de incentivos que refuerce positivamente las conductas deseadas y negativamente las que se pretenden desalentar. De esta forma, el equipo evaluador no solo estaría orientando a los tomadores de decisiones sobre qué se debería hacer para mejorar el programa, sino que ofrecería además ideas preliminares acerca de cómo se podría hacer.

CAPÍTULO 10. COSTO Y FINANCIAMIENTO DE UNA EVALUACIÓN

Este capítulo trata brevemente de los aspectos económicos de una evaluación, tanto su costo como su financiamiento. Es difícil discutir estos temas en pocas páginas, ya que las variaciones posibles son innumerables. De todos los elementos que tienen que ver con diseñar y aplicar una buena evaluación, el costo es probablemente aquel en que es más difícil dar consejos concretos, en función de la gran variabilidad de circunstancias que los programas enfrentan (programas del sector público versus privados, iniciativas nacionales versus locales, intervenciones de gran presupuesto versus iniciativas modestas, etc.). El objetivo aquí es simplemente ofrecer algunas recomendaciones básicas y generales que puedan ser de utilidad, tanto a ejecutores de programas de prevención de homicidios como a agencias financiadoras y a la sociedad civil.

10.1 ¿Son caras las evaluaciones?

La primera cuestión que se plantea es cuántos recursos deberían ser dedicados a la evaluación. Es evidente que sin evaluaciones de calidad los programas serán dictados por impulsos y convicciones personales y los recursos continuarán siendo aplicados de manera ineficaz o ineficiente. En el otro extremo, es posible imaginar un escenario, por más que parezca atípico en el momento actual, en el que un exceso de recursos dedicados a la evaluación pudiera inhibir la inversión en programas necesarios. El punto de equilibrio entre ambos riesgos no está definido, aunque parece claro que estamos hoy más cerca del primero que del segundo. Por otro lado, puede argumentarse que la amenaza de un exceso de evaluaciones es solo la pérdida de eficiencia en la inversión, esto es, un costo mayor para obtener el mismo resultado. Por su parte, el riesgo asociado a un déficit evaluativo es la ausencia completa de eficacia, o sea, una inversión sin resultado ninguno. Ocasionalmente, es posible incluso que la intervención pueda generar algún impacto negativo.

Es común que la presión por realizar una evaluación proceda de las entidades financiadoras, de los órganos de fiscalización o del debate público. En ese sentido, la ausencia de una cultura de evaluación en el área de seguridad pública, a diferencia por ejemplo del área de salud, implica que la evaluación es percibida como una demanda externa y no como una necesidad. Como consecuencia, es posible que cualquier gasto pueda ser percibido como superfluo y que algunas personas encargadas

de la gestión piensen que constituye un recurso desperdiciado que podría haberse dedicado a potenciar la intervención.

De hecho, entre las personas que gestionan programas de prevención de violencia, y de homicidios en particular, es frecuente la creencia de que las evaluaciones son de alto costo y que apenas deben ser realizadas para iniciativas de gran envergadura. Sin embargo, los mayores obstáculos para una evaluación rigurosa son muchas veces políticos y técnicos, más que económicos. Por ejemplo, la aleatorización del grupo experimental y de control no siempre es fácil de conseguir y a veces es inviable, pero cuando las condiciones existen para ello, su costo suele ser reducido.

De forma general, puede afirmarse que:

- a) las evaluaciones de implementación, que se suelen confundir con las auditorías, pueden y deben ser realizadas por todos los programas de prevención de homicidios independientemente de su dimensión;
- b) las evaluaciones de impacto son más complejas técnicamente y necesitan de una serie de requisitos metodológicos (ver [Capítulo 5](#) y [Capítulo 6](#)), pero es perfectamente posible ejecutarlas también en proyectos de pequeño y mediano porte.

Una forma simple de resolver el dilema sobre el volumen de recursos que debe ser dedicado a la evaluación es *dedicar un porcentaje fijo del presupuesto del programa* a tal fin. Diversas instituciones públicas y privadas estipulan un porcentaje que va desde 5% hasta 20% o incluso más en algunos casos, dependiendo del nivel de complejidad (ver por ejemplo las [recomendaciones de agencias del gobierno de los Estados Unidos](#)). En la literatura de evaluación, valores entre 5% y 10% del presupuesto total son considerados una referencia para instituciones privadas y públicas, aunque este porcentaje no siempre se alcanza en la práctica. En una encuesta a fundaciones y organizaciones sin ánimo de lucro en 2010, el porcentaje medio encontrado fue de 3,4% (Hewlett Foundation, 2014).

En general, se suele dar una cierta economía de escala, de forma que cuanto mayor es el volumen y el presupuesto de un proyecto, menor tiende a ser el porcentaje de su presupuesto dedicado a la evaluación. En una muestra de proyectos de gran volumen financiados por el Banco Mundial, por ejemplo, la media del presupuesto de evaluación fue de 4,5% y los programas de mayor presupuesto tendían a comprometer un porcentaje menor (Gertler *et al.*, 2011).⁶¹ Esto sucede, entre otras cosas, porque parte de los costos de la evaluación son fijos (diseño, análisis de datos) y no dependen del tamaño del proyecto. Según esta lógica, cuanto mayor sea el volumen del programa, menor tendría que ser el

61 [Ver Capítulo 10.](#)

porcentaje dedicado a la evaluación, pero el riesgo de este abordaje es que los programas pequeños, que ya tienen limitaciones presupuestarias, tendrían que dedicar una proporción mayor. De hecho, quienes gestionan intervenciones locales suelen argumentar que las iniciativas de bajo presupuesto no cuentan con recursos para contratar evaluadoras y evaluadores externos ni para implementar los sistemas de información necesarios para evaluar impacto de forma rigurosa. En la práctica, las evaluaciones más rigurosas tienden a ser más costosas, en función de la información necesaria, pero el costo no aumenta linealmente con el rigor. Toda esta complejidad reforzaría la opción de dedicar un porcentaje mínimo fijo a la evaluación, como forma de protegerse contra riesgos en una u otra dirección, aunque esta es una cuestión que está lejos de suscitar un acuerdo general.

Para iniciativas de organizaciones pequeñas, con poco presupuesto y limitaciones técnicas, una posibilidad puede ser reunir varias instituciones o proyectos e intentar contratar un servicio técnico para todos ellos, creando economías de escala. Las universidades o centros de investigación también pueden prestarse a colaborar a cambio de tener acceso a los datos.

En realidad, el costo efectivo de una evaluación depende de diversos factores vinculados al propio programa, al diseño de la evaluación y al plan de diseminación de resultados.

Algunos de los elementos que influyen en el costo de la evaluación son los siguientes:

- el número de dimensiones que deben ser monitoreadas,
- la necesidad de recoger datos primarios,
- la existencia o no de sistemas de registro de información,
- la localización y dispersión de la población beneficiaria,
- el intervalo temporal dentro del cual los efectos deben ser estimados,
- el número de grupos de interés que deben ser incorporados al proceso e informados del resultado final,
- la forma en que los resultados serán difundidos.

De forma general, se puede decir que el componente más caro de una evaluación suele ser el proceso de recogida de informaciones.

En los casos en los que los recursos disponibles para la evaluación sean, por un motivo u otro, insuficientes, esto puede llegar a comprometer la posibilidad de evaluar el programa de una forma rigurosa. El dilema

que se plantea es si vale la pena realizar una evaluación de inferior calidad o si es mejor no hacer ninguna. La respuesta a esta pregunta es afirmativa para las evaluaciones de implementación, pues, como ya se ha señalado, siempre es necesario realizarlas. En el caso de las evaluaciones de impacto, no hay, desgraciadamente, una respuesta clara, pues, de hecho, hay situaciones en que ejecutar una evaluación de escasa fiabilidad que lleve a conclusiones erróneas puede tener peores repercusiones que no hacer nada. La recomendación es llevar a cabo evaluaciones de impacto siempre que las limitaciones metodológicas, operacionales o económicas no comprometan seriamente la validez interna, esto es, la solidez de la inferencia causal sobre la eficacia o no del programa.

10.2 ¿Quién debe financiar las evaluaciones?

Si los programas cuentan con una entidad financiadora distinta de la institución que los ejecuta, lo ideal es que esa entidad corra también con los gastos de la evaluación y que esta sea directamente contratada por quien financia y no seleccionada por quienes gestionan la iniciativa. Esto sirve a diversos propósitos. Por un lado, la agencia financiadora tendrá el mayor interés en conocer genuinamente el resultado de su inversión. Por otro, cuando pensamos en evaluaciones externas al programa, el hecho de ser retribuido directamente por quien financia, específicamente para este cometido, subraya la independencia del equipo de evaluación ante las personas responsables por la gestión del programa y la relevancia de la propia evaluación ([ver apartado 8.1](#)). En los casos en los que se optó por dedicar un porcentaje fijo del presupuesto total a la evaluación, la institución financiadora puede escoger, separadamente, al ejecutor de la intervención y a los evaluadores y evaluadoras dentro del mismo proceso. El riesgo es que el equipo de evaluación externo, remunerado por quien financia, refuerce la percepción por parte de quienes se encargan de la gestión de que la evaluación es una imposición externa y una amenaza al programa. De todas formas, el riesgo contrario, el de que quien evalúa sea escogido al gusto de las personas responsables de la gestión para conseguir un resultado favorable, es mucho más grave. No tiene sentido apostar por una evaluación externa y sus potenciales beneficios si después el equipo de evaluación es elegido para producir o favorecer determinados resultados.

Cuando la institución financiadora y la gestora son una sola, no tiene sentido pensar en una conexión directa entre el ente financiador y el equipo de evaluación que deje al ejecutor a un lado, pero aún es necesario, tal vez incluso más, que el proceso de selección de ese equipo se dé con independencia de quien gestiona el programa.

Costo y financiamiento de una evaluación

RESUMEN

Algunas personas responsables de la gestión piensan que las evaluaciones son algo reservado para programas de gran envergadura, pero siempre es posible realizar una evaluación de implementación y, frecuentemente, también es posible llevar a cabo una evaluación de impacto, incluso con un presupuesto reducido.

Algunas instituciones reservan una proporción del presupuesto del programa para la evaluación: 5%, 10% o incluso más, aunque no hay consenso sobre este tipo de criterios.

La mayor parte del presupuesto se suele dedicar a la recogida de datos, mientras que los costos técnicos suelen ser fijos. En general, cuanto mayor sea la calidad de la evaluación, mayor será su coste, pero este último no aumenta linealmente en función del rigor.

CAPÍTULO 11. DIFICULTADES PARA EVALUAR PROGRAMAS DE PREVENCIÓN DE HOMICIDIOS

En este capítulo se discuten algunos de los principales obstáculos para evaluar programas de prevención de homicidios. Algunos de estos problemas son propios y específicos de este tipo de evaluaciones, otros son comunes a evaluaciones de programas de seguridad pública o incluso a programas sociales en otras áreas, pero todos ellos tienen especial relevancia a la hora de evaluar programas de prevención de homicidios.

11.1 Ausencia o fragilidad de la lógica del modelo o teoría de cambio

Como ya fue explicado ([ver apartado 5.3](#)), la lógica del modelo o teoría del cambio es una explicación teórica de los mecanismos a través de los cuales las acciones emprendidas producen los efectos deseados. La lógica del modelo es uno de los requisitos de la evaluabilidad. Sin embargo, muchos programas de prevención de homicidios no cuentan con una lógica del modelo explícita y algunos ni siquiera poseen un documento programático desde el cual se pueda pensar la evaluación. Esta es una deficiencia común en los programas de seguridad pública, pero es particularmente grave en el caso de los programas de prevención de homicidios, que muchas veces presumen que una promoción general de la convivencia o una mejora del ambiente urbano automáticamente resultarán en una reducción de la violencia letal. En otras palabras, muchas iniciativas asumen implícitamente que la violencia constituye una dimensión única que va desde las tensiones mínimas de la vida cotidiana hasta los homicidios, pero no presentan ninguna evidencia en este sentido. Otras iniciativas de reducción de homicidios contemplan a beneficiarios y beneficiarias que poco tienen que ver con el perfil de los autores o las víctimas más comunes de este crimen. De hecho, el fenómeno del homicidio tiene varias etiologías que corresponden a procesos, perpetradores y víctimas con perfiles diferentes ([ver Capítulo 1](#)). Por eso, es importante que se defina cuáles tipos de homicidio están siendo contemplados por la intervención.

La ausencia de una teoría del cambio dificulta la elección de los indicadores intermediarios o de producto. Aunque los indicadores finales puedan ser medidos de cualquier forma, pues están referidos a los propios homicidios, en caso de un resultado negativo no será posible saber en qué momento o qué supuesto de la cadena causal originó ese fracaso. Por otro lado, en el caso de lograrse una reducción de los homicidios tampoco será posible saber con exactitud el mecanismo que la provocó.

11.2 Pequeño número de homicidios en el momento inicial

El homicidio es un fenómeno relativamente raro, incluso en áreas de alta incidencia. Una comunidad de 10.000 personas con una elevadísima tasa de 100 homicidios por cada 100.000 habitantes produciría un número aproximado de 10 homicidios por año. Un programa de alto impacto que redujera los homicidios en 50% al año resultaría en un número de 5 homicidios al año siguiente. Esto sería en principio un gran éxito, pero con números tan reducidos es prácticamente imposible obtener resultados estadísticamente significativos que descarten la explicación alternativa de que esta diferencia pueda haber sido producto de error aleatorio. En consecuencia, no podremos saber si el descenso es una simple fluctuación casual de un año a otro o si es genuinamente un impacto del programa. Este es un problema especialmente grave en el caso de programas locales dirigidos a comunidades de pequeño tamaño.

Además de la dificultad para obtener resultados significativos, estos números reducidos implican que las distribuciones estadísticas no serán análogas a la distribución normal, que es la más común en los test, sino fuertemente sesgadas hacia los números bajos. Si se mide, por ejemplo, el número de homicidios para cada manzana de casas, lo que nos daría en principio un alto nivel de precisión, la mayoría de ellas tendrá 0 homicidios, algunas tendrán 1 homicidio y poquísimas tendrán más de uno. Como resultado, esta distribución se parecerá mucho más a una binomial (con dos valores posibles: 0 y 1) que a una normal. En estos casos, lo indicado es dejar de utilizar distribuciones normales y recurrir a otras, como la distribución de Poisson, adecuada para conteo de fenómenos, o la propia binomial. Otra posibilidad consistiría en cambiar la unidad de análisis a niveles mayores de agregación, interviniendo en barrios, por ejemplo, en vez de manzanas, lo cual hará que las distribuciones se vayan aproximando a la normal.

De cualquier forma, la posible superación del problema de la falta de normalidad no nos soluciona la dificultad para obtener significación estadística debido a los números reducidos. Las únicas alternativas posibles aquí son, de nuevo, la ampliación geográfica de las áreas contempladas, para conseguir así más población y más homicidios, o la ampliación de los tiempos considerados (de uno a varios años, por ejemplo), con el mismo propósito. Esto es de poca ayuda para quien trabaja en programas locales en áreas pequeñas, que no pueden ser expandidas con facilidad. En estos casos, lo único que cabe recomendar, además de la posible ampliación de los tiempos, es la elección de lugares particularmente violentos para aplicar los programas que serán evaluados. Así, cuanto mayor sea el número inicial de homicidios, más fácil será después la comprobación del posible efecto de la intervención.

11.3 Desplazamiento criminal

Cualquier intervención en seguridad pública focalizada en territorios reducidos puede provocar desplazamiento geográfico del crimen a otras áreas en que haya menos represión contra ese tipo de delito (Weisburd *et al.* 2006). El ejemplo clásico es la saturación policial. Si se aumenta drásticamente el número de policías en una zona con alta incidencia de robos, es muy posible que los ladrones busquen otras calles con menor vigilancia para actuar. Pero este fenómeno no se limita a la presencia policial. Por ejemplo, si se desarrollan programas para la juventud que ofrezcan oportunidades de formación y empleo para los jóvenes, de modo que no se integren en las bandas de narcotraficantes, es posible que estas procuren otros lugares para reclutar nuevos miembros.

La consecuencia de este desplazamiento geográfico es que la incidencia criminal disminuye en el lugar objeto de la intervención y aumenta en otras zonas, hacia las que los criminales expulsados del local original se han dirigido. Este desplazamiento puede ser total o parcial. Si el programa no tiene impacto más allá del desplazamiento, la incidencia global del fenómeno no variará, sino que apenas se alterará su distribución espacial. En consecuencia, se habrán gastado recursos significativos solo para alterar los locales en que el crimen ocurre. Peor aún, si la posibilidad de desplazamiento no es tomada en cuenta en el diseño de la evaluación, la disminución de la criminalidad en el lugar escogido será erróneamente interpretada como impacto positivo global de la intervención. De esa forma, el programa podrá ser extendido a otras áreas, a pesar de su falta de efecto. Como la posibilidad de desplazamiento disminuye con la proliferación de los locales de intervención, pronto quedará claro que el impacto no es el esperado, contradiciendo el resultado aparente de la evaluación original.

Se trata, en suma, de una seria amenaza a la validez interna de la evaluación, esto es, una deficiencia que puede cuestionar la conclusión de qué fue lo que originó el efecto observado. El riesgo es mayor cuanto menor sea el área geográfica intervenida (por ejemplo, una esquina). Inversamente, cuanto mayor sea el lugar donde el programa se desarrolla (un municipio, una ciudad, etc.), mayores serán las dificultades de los delincuentes para buscar locales alternativos.

Además de la vertiente geográfica, existen otros tipos de desplazamiento (Guerette & Bowers, 2009), como el temporal (el crimen cambia de hora), el cambio del tipo de víctima o de la propia naturaleza del delito (crímenes más perseguidos dan lugar a otros menos vigilados). Con todo, en el caso de prevención de homicidios, el desplazamiento geográfico es el que más puede comprometer los resultados de la evaluación. Incluso podría argumentarse que, si hay desplazamiento a

otros crímenes, necesariamente menos graves que el homicidio, esto ya podría considerarse como un impacto positivo.

La posibilidad de protegerse contra la amenaza del desplazamiento geográfico en el diseño de evaluación depende, críticamente, de la capacidad para anticipar los locales hacia los que el desplazamiento se puede producir. La hipótesis más común es que el desplazamiento se dirija hacia áreas limítrofes o contiguas a la intervención, hacia donde podría ser más fácil para los delincuentes. Sin embargo, las vías de comunicación pueden facilitar el desplazamiento en algunas direcciones en detrimento de otras. Además, los mercados criminales pueden también favorecer a determinados locales, como por ejemplo territorios dominados por el mismo grupo criminal, a pesar de estar situados más lejos. Por eso, *el primer ejercicio es pensar cuáles son los locales hacia los cuales el desplazamiento criminal es más probable*. Para ello, la lógica del modelo y la teoría del cambio son esenciales, pues permiten desarrollar hipótesis sobre el desplazamiento.

Si, como suele ser habitual, el desplazamiento es más probable hacia locales contiguos, la estrategia recomendada es contar con *dos tipos de grupos de control*: contiguos y no contiguos. Si se da una disminución de los homicidios en el grupo experimental comparado con los grupos de control contiguos, pero no en relación con los no contiguos, esto es una fuerte indicación de desplazamiento, especialmente si los homicidios aumentaron en los grupos contiguos. Si el resultado es el mismo en los dos tipos de grupos de control, eso apunta a la ausencia de desplazamiento. Otra estrategia alternativa es *adoptar como grupo de control un área lo más amplia posible*, siempre que sea comparable con el grupo experimental, de forma que esto diluya cualquier posible desplazamiento. Por ejemplo, si el grupo experimental es un barrio de 5000 habitantes y el grupo de control son 20 barrios que totalizan 100.000 habitantes, cualquier desplazamiento del primero hacia los segundos tendrá un efecto mínimo sobre la incidencia de los homicidios en estos últimos. Obsérvese, sin embargo, que esta opción no resuelve completamente el problema, pues si bien evita en buena medida la sobrestimación del grupo de control, no previene la subestimación del grupo experimental. Así, los homicidios que antes ocurrían en el área intervenida podrían continuar siendo desplazados a lugares vecinos.

Además, hay que tomar cuidado con el posible *desplazamiento de los locales de registro de los homicidios*. En ese sentido, no es raro que las autoridades locales argumenten que no hay tantos homicidios en su jurisdicción, pero que les vienen a tirar en su barrio o en su ciudad los cadáveres de homicidios acontecidos en otros lugares. Esto ocurrió, por ejemplo, en la evaluación del proyecto Municipios Libres de Armas en El Salvador (Cano, 2007). En estas circunstancias, se recomienda verificar el local de residencia de las víctimas. Si de hecho se produjo un aumento

en la tendencia a arrojar cadáveres en otros locales, entonces se debería encontrar, en paralelo, un aumento del porcentaje de cadáveres que eran residentes en otros lugares.

11.4 Regresión a la media

En la literatura sobre evaluación, la regresión a la media es un problema clásico que surge cuando se eligen personas con una puntuación extremadamente alta o extremadamente baja en un primer momento (pretest) y se acompaña su desempeño en un segundo momento o posttest (Cook & Campbell, 1979). Como el altísimo (o bajísimo) resultado inicial se debe en parte a error aleatorio, digamos a la suerte, es más que probable que en un segundo momento los resultados sean menos extremos, evolucionando en la dirección de la media incluso en ausencia de cualquier impacto del programa. De ahí el nombre, pues esta regresión sería un fenómeno automático e independiente de cualquier intervención, debido apenas a la existencia de error en las mediciones.

Es común que los programas de prevención de la violencia letal sean concebidos e introducidos en momentos en que el problema se tornó excepcionalmente grave, como el año posterior a un pico de los homicidios. Así, la disminución al año siguiente podría deberse al impacto del programa, pero también a una disminución natural, esperable cuando esas circunstancias excepcionales de violencia hayan cesado. Un ejemplo es el de Ciudad Juárez, que alcanzó un pico elevadísimo de homicidios en 2010 y luego descendió, también abruptamente. Está claro que el programa Todos Somos Juárez tuvo algo o mucho que ver con esa evolución, pero no se sabe hasta qué punto parte de esa reducción hubiera sucedido naturalmente después de la emergencia inicial.

La regresión a la media, en la medida que ocurre, no puede ser evitada, pero el problema metodológico puede ser resuelto contando con un grupo de control que registre los mismos valores extremos iniciales. Si se cuenta con varios locales que integren el grupo experimental y otros tantos que compongan el grupo de control, la garantía es aún mayor. La principal cautela a ser tomada en los programas de reducción de homicidios es intentar no elegir como locales aquellos con una incidencia extrema, sobre todo si no es posible encontrar grupo de control en la misma situación. Si, por ejemplo, se eligen lugares con incidencia media-alta, en lugar de extrema, el problema se verá aminorado en buena medida. Sin embargo, obsérvese que este consejo va en la dirección contraria al que se ofrece en el [apartado 11.2](#) (pequeño número de casos), pues valores bajos dificultan la obtención de significación estadística. Por tanto, es preciso conciliar ambos riesgos.

11.5 Cambios poblacionales

La violencia extrema puede llegar a expulsar las poblaciones de sus áreas de residencia, ya sea mediante la eliminación física de las víctimas o, en mayor medida, por el desplazamiento de personas que, justamente, intentan evitar la victimización. Entre otras consecuencias, los denominadores de las tasas de homicidio están sobrestimados y las tasas, subestimadas. Eso sucede en los últimos años, por ejemplo, en Venezuela, después de que millones de venezolanos hayan dejado el país como consecuencia de la crisis económica y política.

Así, los datos censitarios, que también son la base de los marcos muestrales, ya no corresponden a la realidad, pues mucha gente ya no vive allí. Estas migraciones no suelen ser homogéneas, sino que afectan predominantemente a determinadas áreas y a ciertos perfiles poblacionales que sufren altos niveles de violencia, lo que puede introducir sesgos significativos. Cuando se realiza una investigación de campo evaluativa que necesita recoger datos en el terreno, a veces es necesario hacer un relevamiento para actualizar los marcos muestrales de forma que se ajusten a la realidad. Ocasionalmente, puede ser necesario buscar a las poblaciones desplazadas en sus nuevos locales de residencia, lo que implica un serio desafío metodológico.

En lo que se refiere a las evaluaciones, es importante considerar estas posibles pérdidas poblacionales y es esencial que, de producirse, se den de forma aproximadamente semejante en el grupo experimental y en el de control. De lo contrario, los homicidios podrían bajar en el grupo experimental, simplemente como consecuencia de la ausencia de población y esto podría confundirse con un impacto positivo del programa.

11.6 Alteraciones en las categorías o en los sistemas de registro de homicidios

Como ya fue explicado en el [Capítulo 1](#), la medición de los homicidios no es algo trivial por dos motivos. En primer lugar, existen varias definiciones posibles de homicidio defendidas por diversas instituciones. La inclusión o no de las personas muertas por la policía, de los homicidios en legítima defensa o de las víctimas de guerras y conflictos políticos son ejemplos de ese tipo de diferencias. En segundo lugar, los registros de homicidio pueden adolecer de problemas de validez y fiabilidad, en función de diversos problemas, como la falta de cobertura o la existencia de categorías ambiguas o residuales que pueden incluir homicidios (“muertes por causa externa de intención desconocida”, “muerte sospechosa”, etc.).

El riesgo para la evaluación es que se produzca algún cambio, ya sea en las categorías utilizadas para medir el homicidio, ya sea en los sistemas de registro, durante el período que está siendo evaluado. Si esto llegara a ocurrir, los cambios en la incidencia de homicidios observada pueden ser debidos simplemente a alteraciones en el procesamiento de la información y no al impacto de un programa. Por eso, es muy importante garantizar que no se produzcan cambios en los conceptos y en los sistemas de información que recogen los homicidios a lo largo del período que es sometido a evaluación. Caso contrario, sería necesario tomarlos en cuenta e intentar eliminar su influencia en el análisis final.

11.7 Cambios que afectan a todo el territorio e inviabilizan a un grupo de control

Cuando el programa de prevención de homicidios se ejecuta por medio de o viene acompañado por un cambio en la legislación o una alteración en el funcionamiento del sistema de justicia criminal que afecta a todo el país de una sola vez, no es posible generar un grupo de control, dado que la modificación sucede al mismo tiempo en todas partes. Un ejemplo de ello puede ser un programa de reducción de homicidios basado en una nueva ley de control de armas o en una reestructuración nacional de los departamentos de investigación de homicidios.

En estos casos, cuando no es posible la introducción paulatina del programa, la estrategia metodológica alternativa más común es aplicar un análisis de series temporales para intentar estimar el impacto ([ver apartado 12.2](#)).

11.8 Intervenciones con múltiples estrategias simultáneas

Algunos programas de reducción de homicidios incorporan múltiples medidas de diversa naturaleza aplicadas a territorios reducidos, usualmente de alta incidencia, buscando muchas veces un efecto demostración que transmita el mensaje a los ciudadanos de que es posible reducir la violencia letal incluso en áreas consideradas muy peligrosas. En este tipo de programas tanto la inversión como la focalización son elevadas para conseguir un efecto significativo (ver el cuadro siguiente sobre Todos Somos Juárez).

Estos programas presentan dos tipos de problemas, incluso cuando consiguen comprobar un impacto positivo, que están relacionados entre sí. Uno es de corte metodológico y consiste en la extrema dificultad, o imposibilidad, de saber cuáles entre las numerosas estrategias provocaron el impacto y cuáles fueron inocuas o incluso contraproducentes. El

segundo problema es que la inversión económica de tales proyectos es tan elevada que no es factible expandirlos a un gran número de lugares, mucho menos al conjunto del país. En consecuencia, como no se conocen cuáles son las medidas que tuvieron impacto y no se pueden replicar todas, la utilidad de la evaluación y la del propio programa son reducidas, más allá del efecto demostración.

Para poder evaluar estas iniciativas, lo ideal es implementar parcialmente sus componentes en lugares distintos, combinando diversas estrategias en cada uno de los locales. Eso permitiría estimar el impacto de cada una de las intervenciones o de combinaciones precisas de varias de ellas, extrayendo así lecciones importantes para futuras expansiones. Estas aplicaciones parciales de las distintas medidas en locales diferentes implican una mayor dificultad logística y de diseño, pero mejoran significativamente la evaluabilidad y la utilidad de los resultados. Idealmente, sería deseable aplicar todas las combinaciones posibles de estos componentes⁶² para poder detectar posibles interacciones entre ellos, esto es, efectos conjuntos que difieran de la suma de los efectos individuales de cada componente.

Todos Somos Juárez, en Ciudad Juárez, México

Juárez es una ciudad en el estado mexicano de Chihuahua, en la frontera con los Estados Unidos. Hasta 2007, el número promedio de homicidios era de 200 víctimas por año, pero a partir de 2008 hubo una explosión de violencia. En 2010, el Instituto Nacional de Estadística y Geografía (INEGI) registró 3.798 homicidios, una cifra superior a 200 homicidios por cada 100.000 habitantes. La violencia en la ciudad comenzó debido a la guerra entre los carteles. En 2008, autoridades municipales y estatales exigieron la presencia del Ejército, que coincidió con la estrategia de guerra contra las drogas del gobierno del presidente Calderón. Esta estrategia dejó al Ejército a cargo de la lucha contra el crimen organizado y se concentró en neutralizar a los líderes de estos grupos criminales. En enero de 2010 hubo una masacre de jóvenes que el presidente inicialmente atribuyó a un ajuste de cuentas entre grupos criminales. Cuando se comprobó que las víctimas eran simplemente estudiantes, el gobierno mexicano se disculpó y se comprometió a invertir en prevención de la violencia en Ciudad Juárez. La visita oficial del presidente a la ciudad para conocer la situación y escuchar las demandas puede considerarse como el punto de partida de la iniciativa Todos Somos Juárez (TSJ).

El objetivo central del proyecto TSJ, que articula tres niveles de gobierno y sociedad civil, era la “reconstrucción del tejido social”, que habría sido destruido por la violencia, mediante la reducción de la delincuencia y del comportamiento violento. Aunque la intervención no se dirigió exclusivamente a los homicidios, su reducción se consideró como el

62 Esto se conoce en el lenguaje metodológico como un diseño de *cuadrado latino*.

objetivo central. La intervención puede resumirse en tres ejes centrales: a) inversión social; b) cambios en los órganos del sistema de justicia penal, como la policía y el Ministerio Público; c) mecanismos de participación social y creación de una arquitectura institucional que apoyase el programa. El programa se articuló alrededor de 160 compromisos específicos, agrupados en seis áreas: seguridad; economía; empleo; salud; educación, cultura y deportes; y desarrollo social. TSJ contó con un financiamiento de aproximadamente USD 400 millones ejecutado entre febrero de 2010 y noviembre de 2012.

Los homicidios disminuyeron drásticamente desde la segunda mitad de 2010, de modo que, en la segunda mitad de 2012, ya habían regresado a un nivel similar al de 2007.

Esta intervención no se sometió a ninguna evaluación de impacto. De hecho, sería extremadamente difícil evaluar el efecto de cada una de las medidas tomadas bajo el paraguas de TSJ, en función de su diversidad y de su aplicación simultánea en el tiempo y en el espacio. Por lo tanto, es natural que un proceso tan complejo y con tantas dimensiones esté abierto a múltiples interpretaciones, comenzando con el posible impacto de las iniciativas de prevención y de las acciones policiales. Sin embargo, muchos atribuyen a los propios carteles criminales o a la mejora de la situación económica la responsabilidad por la reducción de la violencia. Por otro lado, considerando la naturaleza “epidémica” del estallido de violencia en Juárez desde 2007, la regresión a la media podría ser también una hipótesis alternativa para ayudar a explicar el posterior declive ([ver apartado 11.4](#)).

11.9 Distinta temporalidad de los efectos

Los programas de reducción de homicidios pueden presentar efectos a corto, medio y largo plazo, dependiendo de las estrategias utilizadas y de las respectivas teorías del cambio. Por ejemplo, las negociaciones entre grupos criminales pueden reducir los homicidios de forma inmediata, los cambios policiales para mejorar la tasa de esclarecimiento podrían contener las muertes violentas en plazos cortos, las intervenciones de prevención secundaria o terciaria podrían hacer lo propio a medio plazo, mientras que los programas de prevención primaria para niños podrían evitar los homicidios a largo plazo. De hecho, algunas investigaciones constatan correlaciones entre los indicadores socioeconómicos de los municipios en un determinado momento y la tasa de homicidios que se registra veinte años después (Ribeiro & Cano, 2016).

Por otro lado, hay efectos que son relativamente duraderos, como los de la prevención social, y otros que se manifiestan solo mientras el programa está en vigor, como las prohibiciones de porte de armas en días de alto riesgo ([ver apartado 14.3](#)).

Tanto la demora hasta producir el efecto como su permanencia dependen críticamente de la naturaleza de las intervenciones. Como ya se vio en el [apartado 5.4](#), la evaluación debe registrar los efectos en sintonía con la temporalidad de los efectos esperados. Sin embargo, es muy común que programas de prevención social de los cuales solo cabría esperar efectos a largo plazo sean evaluados meses después de su aplicación. Además, no es raro que las decisiones sobre los programas deban ser tomadas en poco tiempo, lo que dificulta que las evaluaciones registren efectos demorados y midan la curva de permanencia temporal de los impactos. De hecho, muchas evaluaciones cuentan con una única medición posterior a la intervención, lo que torna imposible el monitoreo de la evolución de los efectos. A pesar de que los tiempos de la evaluación puedan ser limitados por imperativos administrativos o políticos, es importante dejar instalados los sistemas de registro de información para que, meses o años después, los efectos ulteriores puedan ser estimados.

RESUMEN

Dificultades para evaluar programas de prevención de homicidios

- Ausencia o fragilidad de la lógica del modelo o teoría de cambio.
- Pequeño número de homicidios en el momento inicial.
- Desplazamiento criminal.
- Regresión a la media.
- Cambios poblacionales.
- Alteraciones en las categorías o en los sistemas de registro de homicidios.
- Cambios que afectan a todo el territorio e inviabilizan un grupo de control.
- Intervenciones con múltiples estrategias simultáneas.
- Distinta temporalidad de los efectos.

CAPÍTULO 12. ELEMENTOS QUE CONDICIONAN EL DISEÑO DE LA EVALUACIÓN

Las personas responsables de la gestión y las entidades financiadoras de intervenciones para la reducción de homicidios se suelen enfrentar a la cuestión de cómo evaluar un programa específico. La respuesta no es simple, pues la decisión sobre qué diseño de evaluación debe ser escogido en cada caso depende de muchos factores y del contexto concreto.

Es posible, sin embargo, formular una serie de consideraciones generales que, sin ofrecer recetas concretas para los casos individuales, ayuden en el proceso de reflexión sobre las opciones metodológicas disponibles. En esa línea, este capítulo sopesa los elementos que mayor influencia pueden ejercer al momento de definir qué posibilidades metodológicas existen para la evaluación de un determinado programa.

El primer punto a ser considerado es la *extensión territorial* del programa, básicamente si este se aplica al conjunto del territorio o apenas a algunas áreas. La limitación puede deberse a la restricción de los fondos disponibles o a la propia naturaleza del programa. Esto es de gran importancia para determinar si es posible crear un grupo de control o no. Si la iniciativa actúa sobre áreas específicas, como ciertos barrios o comisarías de policía, la primera sugerencia sería intentar generar un grupo de control aleatorio sorteando los barrios o las comisarías que serán objeto de la intervención y dejando el resto como parte del grupo de control. Si el programa no se aplica por territorios, sino sobre individuos concretos, la opción de aleatorizar quién será beneficiado o beneficiada suele ser más viable. Aquí la pregunta es si todos los individuos de interés serán beneficiados o si el programa atenderá apenas a una parte, de forma que los restantes puedan componer el grupo de control.

Por otro lado, si la intervención se aplica a todo el territorio sin excepción, como sería el caso de un cambio legislativo sobre el femicidio, entonces no existe posibilidad de crear un grupo de control, ni siquiera un pseudogrupo de control o de aplicar una puntuación de propensión. En estos casos de cambios bruscos y universales, el abordaje metodológico más adecuado es el de las *series temporales interrumpidas*, que pueden determinar si existe una variación en el nivel de homicidios una vez que el programa entró en vigor, comparando el resultado con la proyección elaborada a partir de los registros anteriores de la serie (ver un ejemplo de esa estrategia en el [apartado 14.2](#), sobre la evaluación de la ley contra la violencia doméstica en Brasil).

Otra cuestión relacionada con el punto anterior de la territorialidad es la *temporalidad de la implantación* del programa. Si el programa, aun siendo universal en su desarrollo final, puede ser implementado poco a poco, como por ejemplo una nueva doctrina nacional de patrullaje policial, sería posible comenzar la aplicación en algunas áreas, que serían el grupo experimental, y atrasarlo en otras, que constituirían el grupo de control. Esto no supone necesariamente una alteración de los planes de introducción del programa, pues normalmente el proceso de implantación necesita recursos y otros elementos (formación, infraestructura, etc.) que no pueden ser administrados a la vez en todo el territorio. Un ejemplo de ello es el Plan Cuadrante de la Policía Nacional colombiana, que fue evaluado comparando las estaciones policiales donde se había dado capacitación específica a los agentes sobre el mismo con aquellas otras donde esa capacitación aún no había sucedido (Fundación Ideas para la Paz, 2012).⁶³ De esa forma, aunque el cambio de estrategia se daba a nivel nacional, se tomaba la capacitación como variable operativa para definir el inicio del programa en cada local.

Un tercer elemento conectado con la territorialidad y la temporalidad de la implantación es la *temporalidad de los efectos* del programa. Si el programa tiene un efecto inmediato pero pasajero, puede ser aplicado varias veces a lo largo del tiempo, lo cual ofrece muchas opciones para la evaluación, la más destacada de las cuales es, de nuevo, la estrategia de las series temporales interrumpidas. Ese es el caso de la evaluación de la prohibición temporal del porte de armas en espacios públicos en Bogotá y Cali en algunos fines de semana de alto riesgo ([ver apartado 14.3](#)). Como el impacto de la medida se limitaba al fin de semana en que era aplicada, y como la norma era adoptada en algunos fines de semana y no en otros, eso permitió comparar ambos momentos y refinar la capacidad de estimar el impacto.

Por otro lado, si el efecto del programa es permanente y su impacto solo puede ser medido una única vez, la validez interna del estudio es consecuentemente menor de lo que sería si las aplicaciones fueran múltiples. De la misma forma, si el efecto no es inmediato sino dilatado, la validez interna también se reduce incluso en presencia de múltiples aplicaciones, pues los impactos de diversas aplicaciones podrían superponerse y confundirse entre sí. De cualquier forma, no siempre se conoce con exactitud cuál es, a priori, la temporalidad de los efectos de una intervención.

En realidad, los tres elementos que hemos considerado hasta ahora (territorialidad, temporalidad de la implantación y temporalidad de los efectos) sirven, sobre todo, para decidir si es posible o no obtener un grupo de control. Vamos ahora a analizar las opciones en ambas situaciones.

63 [Ver apartado 2.3](#).

12.1 Diseños evaluativos con grupos de control

Como ya fue explicado, la opción ideal es contar con un grupo de control equivalente, esto es, escogido de forma aleatoria. La posibilidad de aleatorizar o no depende de diversos factores. En muchas ocasiones, escoger las áreas, unidades o personas beneficiadas por sorteo es imposible (como en un cambio legislativo), antiético (como en un programa para proteger personas amenazadas)⁶⁴ o difícil de ejecutar. Otro de los elementos que influencia la posibilidad de aleatorizar es la elección de la *unidad de análisis* (ver apartado 6.1): individuos, familias, áreas territoriales (comunidades, barrios, estaciones de policía), etc. Lo más común es que las unidades sean divisiones territoriales y que la variable dependiente sea, por tanto, la tasa de homicidios para cada una de ellas. De forma general, tiende a ser más fácil aleatorizar individuos o áreas de pequeña extensión que territorios amplios. Cuando el número de unidades (zonas de la ciudad, por ejemplo) es grande, ello fortalece el poder estadístico de los test y la validez interna. Cuando el número de unidades es reducido, disminuye la validez interna y aumenta el riesgo de que otros eventos específicos y concomitantes en una de esas unidades afecten al resultado final.

Cuando no es posible aleatorizar, es frecuente optar por la búsqueda de grupos de control no equivalente, de forma que se asemejen lo más posible a los integrantes del grupo experimental en la tasa de homicidios y en las dimensiones principales que afectan al nivel de homicidios. Estrategias modernas que persiguen este mismo fin son la puntuación de propensión y diferencias en diferencias (ver apartado 6.2).

Uno de los factores que condiciona la elección de grupos de control es la *posibilidad de desplazamiento criminal* como consecuencia de la propia intervención (ver apartado 11.3). Para evitar este riesgo, es fundamental que los grupos de control sean elegidos en áreas lejanas o donde el desplazamiento criminal sea imposible o muy improbable. Idealmente, podría incluirse también un segundo grupo de control, en áreas donde el desplazamiento sea más probable, justamente para intentar medir este desplazamiento.

Otra dimensión que afecta a la formulación de los grupos experimental y de control es la *naturaleza de la intervención*, específicamente si se trata de una acción única o de un conjunto de intervenciones de distinto tipo (ver apartado 11.7). En efecto, cuando hay un gran número de acciones diversas que se ejecutan sobre el mismo territorio, incluso si se consigue evaluar el impacto del programa en su conjunto a través de un grupo de control, difícilmente se sabrá cuál o cuáles de estos componentes es el responsable por el efecto. En estos casos, la estrategia recomendada es descomponer el programa en sus diversos componentes

64 En estas situaciones, a veces es posible proteger a todas las personas en alto riesgo y aleatorizar la protección a personas de riesgo medio, bajo el supuesto de que no habría recursos para incluir a todo el mundo.

básicos y aplicar combinaciones diferentes de dichos componentes en varios territorios para poder establecer cuáles son efectivos.

12.2 Diseños evaluativos sin grupo de control

Como ya se mencionó, cuando no es posible contar con grupos de control, la estrategia metodológica más seguida es el análisis de series temporales. Lo más común es que el programa sea introducido una única vez, y su impacto se estime comparando la proyección de la serie a partir de los registros anteriores a la intervención con lo que realmente sucedió a partir de ese momento. Cuando el programa presenta un impacto temporal e inmediato, como ya se vio, es posible estimar un modelo de series temporales interrumpidas en cada introducción del programa, lo cual fortalece notablemente la validez interna de la evaluación.

Un hecho que hay que tomar en cuenta, pues puede comprometer la calidad de la evaluación a través de series temporales, es si el programa fue ejecutado en un *momento de crisis*. Es normal que las intervenciones sean demandadas y aplicadas inmediatamente después de una crisis, cuando los gobiernos y otras organizaciones se proponen hacer algo para reducir una subida brusca o un nivel anormalmente elevado de homicidios. Sin embargo, en estas circunstancias el riesgo de regresión a la media es elevado ([ver apartado 11.4](#)). En otras palabras, es posible que el descenso posterior de la tasa de homicidios responda, en alguna medida, a que el momento excepcional de crisis haya pasado y no necesariamente al impacto del programa. Para reducir este peligro, es de gran importancia contar con una serie temporal prolongada que revele si existe un ciclo en la tasa de homicidios y, en caso afirmativo, en qué medida se podría esperar una disminución posterior, para no confundirla con el impacto del programa.

Otra estrategia posible en ausencia de un grupo de control puede ser formulada a partir de una *medida de intensidad* de la aplicación del programa, cuando esta existe. Eso sucede, por ejemplo, cuando la implantación del programa se da poco a poco, progresivamente, y se puede medir de alguna forma el grado en que se va introduciendo. Cuando existe esta medida de intensidad del programa, se comparan los resultados obtenidos con el nivel de intensidad, de modo que si la reducción de homicidios es mayor donde el programa fue aplicado con mayor intensidad, eso fortalece la conclusión de que hubo un impacto positivo de la intervención. Imaginemos una norma que prohíba la venta de alcohol a partir de determinado horario, en una ciudad en la que se haya comprobado que los homicidios son frecuentes a la salida de los bares durante la madrugada. Aunque la norma sea aplicada a todo el territorio, podríamos medir el número de operaciones de fiscalización

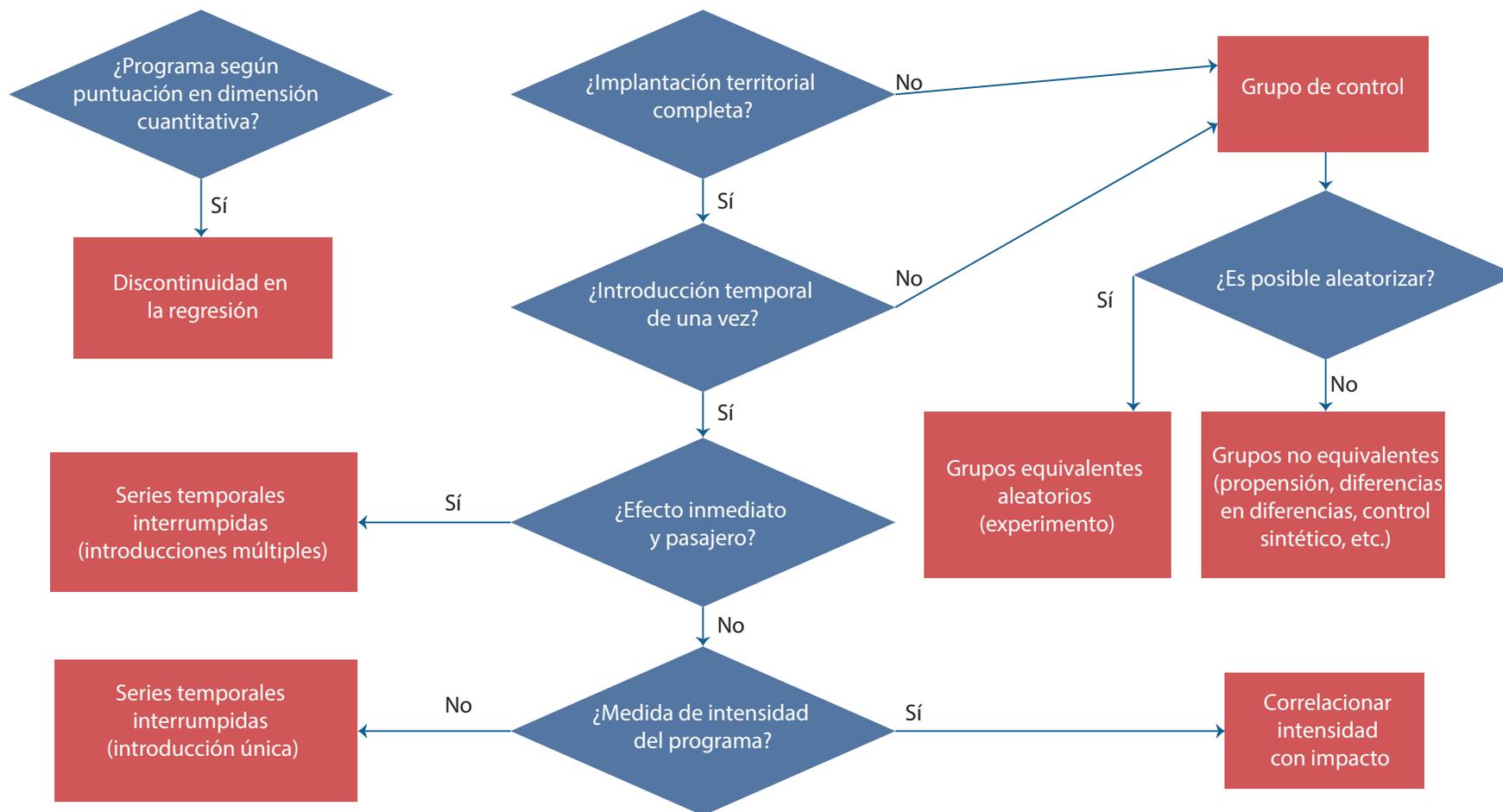
de bares en cada región y usar esta medida como un indicador de intensidad del programa. Para que eso tenga sentido, sería preciso que hubiera suficiente variación de los números de operaciones entre unas zonas y otras, y que se comprobara también una relación entre el número de operaciones y el grado de cumplimiento efectivo de la nueva norma.

En el otro extremo, tenemos los cambios legislativos nacionales, que, por definición, se aplican al conjunto del territorio nacional de una sola vez, a partir de la fecha de entrada en vigor de la nueva ley. Además, supuestamente no debería haber variación en la intensidad de la aplicación, puesto que la nueva ley es la misma para todos. En estos casos, como ya se explicó, la alternativa posible son las series temporales interrumpidas.

Un caso excepcional ocurre cuando el programa ha sido concedido a personas o áreas en función de la puntuación en una dimensión cuantitativa que correlaciona fuertemente con la tasa de homicidios. Imaginemos, por ejemplo, un programa de prevención nacional que selecciona las áreas de aplicación en función de una puntuación en un índice de riesgo (calculado a partir de variables socioeconómicas y de los propios indicadores de violencia). Barrios con un valor superior a una determinada puntuación recibirían el programa de prevención, mientras que barrios con puntuación inferior no serían considerados. En estas circunstancias, el *diseño de discontinuidad en la regresión* presenta una oportunidad única de medir impacto con alta validez interna, analizando la existencia o no de una discontinuidad en las regresiones anteriores y posteriores a este punto de corte que decidió la aplicación del programa ([ver apartado 6.2](#)).

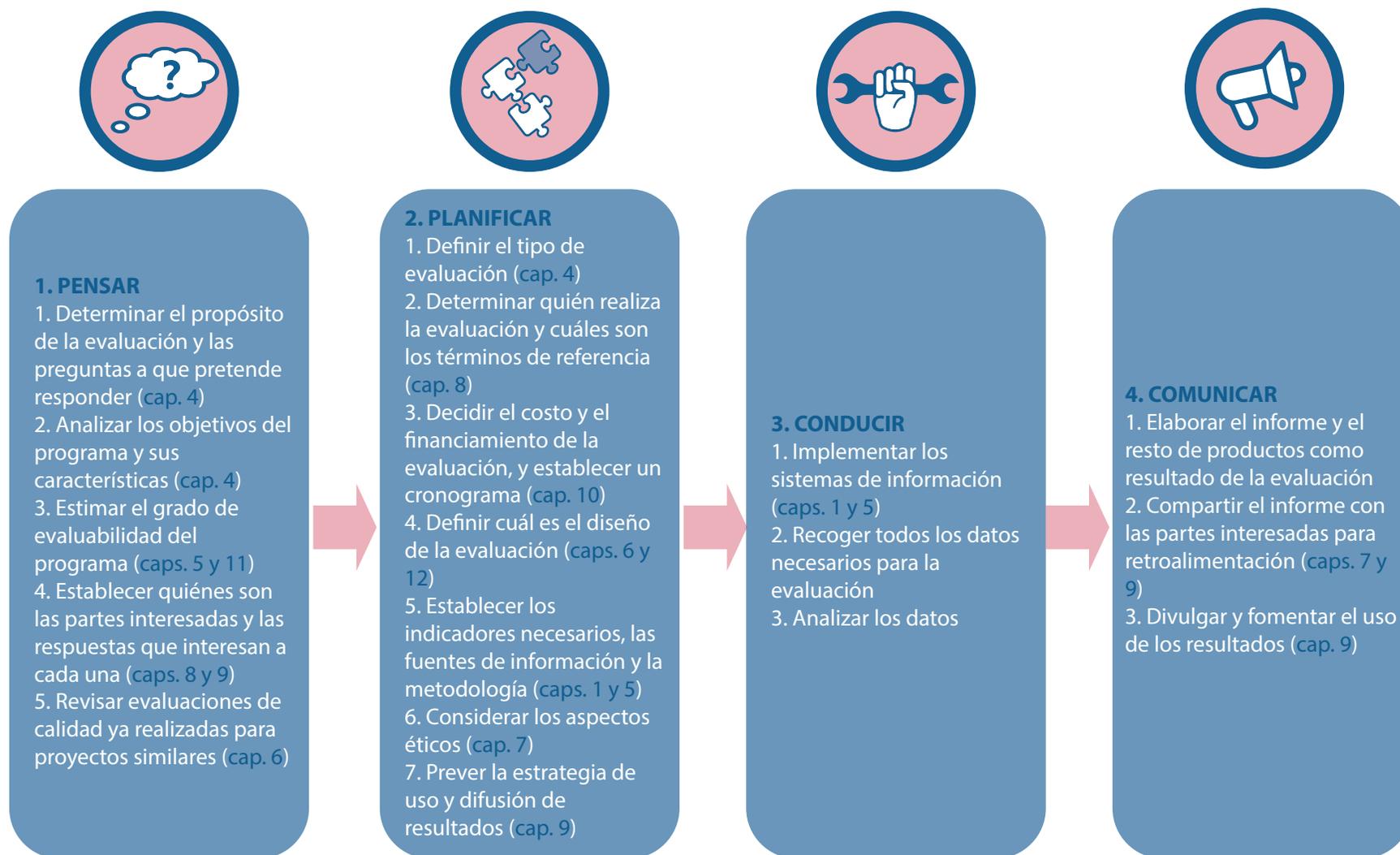
Finalmente, a modo de resumen de este capítulo, presentamos un flujograma con algunas de las decisiones que condicionan la elección del diseño evaluativo.

Esquema 5. Elementos que condicionan el diseño de la evaluación



Fuente: Elaboración propia.

CAPÍTULO 13. PASO A PASO DE UNA EVALUACIÓN DE PREVENCIÓN DE HOMICIDIOS



Fuente: Elaboración propia.

CAPÍTULO 14. EJEMPLOS DE EVALUACIONES DE PROGRAMAS DE REDUCCIÓN DE HOMICIDIOS

Después de observar las pautas descritas en esta guía sobre estrategias metodológicas, políticas e institucionales para ser consideradas en una evaluación, esta sección presentará una breve selección de ejemplos de evaluaciones de programas de reducción de homicidios. Los ejemplos fueron elegidos considerando dos criterios principales. El primero es el interés metodológico y la posibilidad de presentar diferentes diseños de evaluación discutidos en la guía. El segundo es que se tratara de evaluaciones de programas de reducción de homicidios implementados en América Latina y el Caribe. Estos ejemplos pueden ser útiles para ilustrar y ayudar a comprender aspectos específicos del marco teórico y conceptual expuesto hasta ahora.

14.1 Impacto en los homicidios de la creación del Metrocable en Medellín (2004). Evaluación realizada por Cerdá, Morenoff, Hansen, Hicks, Duque, Restrepo y Diez-Roux⁶⁵

14.1.1 Descripción del programa

En 2004, las autoridades municipales de Medellín, Colombia, construyeron un sistema público de teleférico, conocido como Metrocable, para conectar barrios de bajos ingresos aislados del centro urbano de la ciudad. El gobierno municipal realizó mejoras en los vecindarios que pasaron a ser atendidos por el teleférico, incluyendo: iluminación adicional de espacios públicos; nuevos puentes peatonales y calles; parques biblioteca; edificios para escuelas, centros recreativos y centros de promoción de microempresas; más patrullas policiales; y una estación de policía familiar al lado de cada estación del teleférico.

Según la evaluación, los lugares donde se llevó a cabo la intervención fueron seleccionados por un panel de autoridades municipales en función de la viabilidad topográfica y geográfica del emprendimiento. Sin embargo, no está muy claro cuáles fueron exactamente los criterios de esta selección.

El objetivo del programa era afectar positivamente el entorno físico de algunos territorios y, sobre todo, mejorar la vida de los residentes. Entre otros resultados, la idea era lograr una reducción de la violencia y los homicidios al mejorar las condiciones en los vecindarios, fortalecer la eficacia colectiva, aumentar la confianza de los ciudadanos en el sistema de justicia penal y en la policía, además de poner a disposición servicios

Estrategia de evaluación: Diseño cuasiexperimental con grupo de control no equivalente.

65 Cerdá, M.; Morenoff, J. D.; Hansen, B. B.; Tessari Hicks, K. J.; Duque, L. F.; Restrepo, A. & Diez-Roux, A. V. (2012): "Reducing violence by transforming neighborhoods: a natural experiment in Medellín, Colombia". *American Journal of Epidemiology*, 175(10), 1045-1053.

de barrio, como parques y actividades culturales. En este sentido, se trata de un programa que posee componentes de prevención social y de prevención situacional ([ver apartado 2.2](#)).

14.1.2 Metodología de la evaluación

Para verificar si los cambios tuvieron un efecto sobre las condiciones sociales y la violencia en el vecindario, la estrategia metodológica fue un diseño cuasiexperimental que estimase los efectos del programa en los 25 vecindarios de intervención y en 23 vecindarios de control comparables, antes (2003) y después (2008) de la finalización del programa.

Para el período anterior a la intervención, los datos se recopilaron en 2003 mediante una encuesta de hogares sobre barrios y violencia conocida como Prevención de la Violencia en el Valle de Aburra (PREVIVA). Esta encuesta estaba compuesta de una muestra representativa de la población de Medellín entre 12 y 60 años de edad, incluidos los barrios que participaron en la intervención. En 2008, se realizó una encuesta de seguimiento a parte de las personas que habían sido entrevistadas en PREVIVA: a) personas que en 2003 residían en uno de los 25 barrios donde se instaló el primer sistema de teleféricos (localizados en los distritos 1 y 2); y b) personas que en 2003 vivían en de los 23 barrios ubicados en vecindarios comparables en la ciudad (localizados en los distritos 4 y 8) que no fueron atendidos por el teleférico. Así, se realizaron entrevistas de seguimiento con 466 (225 encuestados de los barrios de intervención y 241 encuestados de los barrios de control) de los 599 participantes de PREVIVA (78%).

Las mediciones a nivel de barrio fueron analizadas en un modelo multinivel.⁶⁶ Las medidas de violencia incluyeron: 1) respuestas a las preguntas de la encuesta sobre eventos violentos que ocurrieron en el vecindario en los últimos 6 meses; y 2) registro de homicidios, por barrio, de la Fiscalía. Otras dimensiones consideradas relevantes y que deberían cambiar con la intervención fueron: a) la eficacia colectiva, esto es, la capacidad de los vecinos de intervenir para restaurar el orden; b) la confianza en la justicia; y c) la confianza en la policía.

Inicialmente, el equilibrio previo a la intervención entre los grupos de intervención y control se testó con respecto a las variables dependientes y a las covariables. Para equiparar ambos grupos, se usó la puntuación de propensión.⁶⁷ Es decir, se estimó la probabilidad de cada barrio de ser escogido para recibir la intervención en función de las variables disponibles, y se usó esta probabilidad (propensión) para escoger barrios del grupo de control semejantes a los incluidos en el grupo experimental.

66 Los *modelos multinivel* son modelos estadísticos apropiados para proyectos de investigación en los que los datos de los participantes se organizan en más de un nivel (es decir, datos anidados). Las unidades de análisis suelen ser individuos (el nivel inferior) anidados en unidades agregadas como hogares, barrios, municipios o ciudades (el nivel superior).

67 Para saber más sobre la puntuación de propensión [ver el apartado 6.2](#).

14.1.3 Resultados de la evaluación

Los niveles de homicidio disminuyeron en ambos grupos, pero la disminución fue más pronunciada en el grupo de intervención, donde la tasa de homicidios cayó 84%, en comparación con 60% en el grupo control.

La confianza en el sistema de justicia penal y la confianza en la policía también aumentaron significativamente en el grupo de intervención. La proporción de encuestados que dijeron que llamarían a la policía si veían actividad sospechosa aumentó en un 77% con el tiempo en el grupo de intervención, en comparación con un aumento del 14% en el grupo de control. Esto refuerza la hipótesis de que los cambios en los niveles de violencia pueden deberse, entre otras cosas, a una actitud más favorable en relación con el sistema de justicia criminal.

La existencia de parques y áreas de ocio no parece haber cambiado con el tiempo en ninguno de los grupos, pero los residentes de las áreas de intervención manifestaron un aumento en las oportunidades locales para participar en actividades culturales y educativas.

El análisis que comparó los grupos experimental y de control, controlando 23 covariables, mostró que las diferencias entre ellos con respecto a la variación de homicidios fueron estadísticamente significativas. En comparación con el grupo de control, el grupo intervenido experimentó una caída del 66% en el logaritmo de la tasa de homicidios y una reducción del 75% en las chances de violencia reportada por los residentes.

Sin embargo, en las variables intermediarias (confianza en la justicia, en la policía, eficacia colectiva) las diferencias entre los grupos experimental y de control no fueron significativas en el modelo multivariado.

14.1.4 Limitaciones de la evaluación

La primera dificultad de la evaluación radica en que los criterios que fueron utilizados en su día para seleccionar los barrios que recibirían el Metrocable no están suficientemente claros. A pesar del intento de igualar el grupo de control con el experimental utilizando la puntuación de propensión, la tasa promedio de homicidios antes de la intervención en el grupo experimental (188 por cada 100.000 habitantes) era mucho mayor que en el grupo control (104 por cada 100.000 habitantes). Esto en sí mismo cuestiona la equivalencia entre ambos grupos. En el período posterior a la intervención (en 2008), la tasa de homicidios fue aún menor en el grupo experimental que en el grupo control, lo que apunta a que la evaluación podría estar subestimando el impacto de la

intervención. De cualquier forma, el impacto exacto de la intervención solo podría establecerse si hubiera habido una equiparación entre los grupos experimental y de control, algo que no es factible en función de la gran diferencia inicial en la variable dependiente del estudio.

Muchas de las variables intermediarias, como la eficacia colectiva y la confianza en la policía y la justicia, no presentaron resultados estadísticamente significativos, a pesar de que los autores en ocasiones intentaron suavizar esta conclusión aludiendo a test estadísticos menos rigurosos (probabilidad de error tipo I⁶⁸ más elevada o test de una sola cola).⁶⁹ De esa forma, aunque esto en sí mismo no niega la reducción de los homicidios, sí puede cuestionarse la teoría del cambio y el proceso teórico que, según los autores, habría provocado este resultado.

14.2 Impacto en los homicidios de la Ley ‘Maria da Penha’ para inhibir la violencia contra las mujeres en Brasil (2006). Evaluación realizada por Cerqueira, Matos, Martins y Junior⁷⁰

Estrategia de evaluación: Modelo de diferencias en diferencias.

14.2.1 Descripción del programa

En 2006, Brasil aprobó una ley federal para proteger a las mujeres de la violencia doméstica y familiar, llamada Ley “Maria da Penha” (LMP). Esta normativa no solo imponía una pena mayor al agresor, sino que ofrecía también un conjunto de mecanismos con el objetivo de prevenir la violencia y proteger a las mujeres maltratadas. Además, la LMP incluía instrumentos para preservar la propiedad de la víctima y sus derechos, medidas para mejorar la asistencia jurisdiccional, así como la creación de un programa de rehabilitación de agresores.

Se esperaba que con la implementación de la ley las víctimas encontrasen un entorno más seguro para denunciar la agresión sin temor a represalias, considerando las medidas de protección de emergencia que eran brindadas. Así, uno de los objetivos de la ley era contribuir al aumento del costo esperado del castigo para el agresor, en función de la mayor probabilidad de condena y del aumento de severidad del castigo. Como destacan Cerqueira *et al.* (2015), todo ello debería contribuir a disminuir la tasa de homicidios de mujeres causados por cuestiones de género.

14.2.2 Metodología de la evaluación

Para medir el efecto de la LMP sobre los homicidios de mujeres relacionados a cuestiones de género, los autores de esta evaluación

68 Error tipo I es la probabilidad aceptada de concluir que un resultado es significativo cuando no lo es en realidad.

69 Los test estadísticos suelen ser de dos colas, asumiendo que los dos grupos que son comparados podrían tener diferencias en una u otra dirección, esto es, el primer grupo podría acabar teniendo una media superior o inferior al segundo. Excepcionalmente, cuando se descarta desde el principio que las diferencias puedan ir en una determinada dirección (por ejemplo, el primer grupo podría tener una media superior, pero nunca inferior al segundo) se adoptan test de una sola cola.

70 Cerqueira, D.; Matos, M.; Martins, A. P.; Pinto Júnior, J. (2015): *Avaliando a efetividade da Lei Maria da Penha*. Ipea: Brasília/Rio de Janeiro.

utilizaron datos del Sistema de Información de Mortalidad⁷¹ del Ministerio de Salud, basado en los certificados de defunción proporcionados por los Institutos Médicos Legales.

Los datos disponibles sobre homicidios contra mujeres no identifican cuáles se refieren a la violencia doméstica, que serían los casos afectados por la LMP. Para identificar los homicidios relacionados a la violencia doméstica y distinguirlos de otros tipos de homicidio se comparó aquellos que sucedieron dentro y fuera de la residencia, considerando que la mayor parte de los homicidios de mujeres por razones de género tienen lugar dentro de sus casas.

Cabe mencionar que la intervención se aplicó en todo el territorio nacional en un solo momento, al ser una modificación legal de aplicación general y automática. Considerando este hecho y el tipo de intervención, el diseño más apropiado para la evaluación fue el modelo de diferencias en diferencias,⁷² en el que el grupo de tratamiento (homicidio de mujeres) se enfrentó con el grupo de control (homicidio de hombres). Esta elección metodológica se basó en la idea de que existen factores asociados con la violencia generalizada en la sociedad y, en particular, la violencia urbana, que afectan a los homicidios de hombres y mujeres. Sin embargo, hay otros factores relacionados con el género que influyen solo a los homicidios de mujeres. En este sentido, la hipótesis central era que la introducción de la LMP debería reducir los homicidios de mujeres en mayor medida que los de hombres, particularmente los homicidios ocurridos dentro de las residencias.

La variable dependiente utilizada en este estudio es el logaritmo de la tasa de homicidios por microrregión y año. En vista de la especificidad del fenómeno, se estimaron modelos separados para la tasa general de homicidios y la tasa de homicidios ocurridos en las residencias. Las variables independientes utilizadas en los modelos fueron: sexo (hombres - grupo control; mujeres - grupo de tratamiento); existencia de la LMP; logaritmo de la tasa de mortalidad por alcohol; logaritmo de la tasa de suicidios con armas de fuego; indicador de microrregión; y año. Las variables mencionadas sobre armas de fuego y consumo de alcohol se crearon como mediciones indirectas (*proxies*), para intentar controlar la influencia de la prevalencia de armas de fuego y del consumo de alcohol, que podrían afectar los homicidios entre hombres y mujeres de manera diferente. A partir de este modelo, se obtuvo la tasa esperada de homicidios para hombres y mujeres en los períodos anteriores y posteriores a la aprobación de la LMP (en 2006).

A pesar de las altas tasas de homicidios en Brasil, este fenómeno es un evento raro desde un punto de vista estadístico, especialmente

71 Los registros considerados homicidios fueron aquellos que tenían los siguientes códigos de causa básica de muerte de la Clasificación Internacional de Enfermedades - 10a Revisión: X85 a X99 e Y00 a Y05.

72 Para saber más sobre la técnica de diferencias en diferencias [ver el apartado 6.2.](#)

en lugares con baja población, como las microrregiones con menos de 50 o 100.000 habitantes. En estos casos, hay distorsiones en el cálculo de la tasa de homicidios. En función de ello, los autores utilizaron el método espacial bayesiano para estimar las tasas de homicidios locales. Este método ajusta las tasas para cada microrregión en función de las tasas de las microrregiones vecinas. Con ello, se evita también el valor 0, considerando que el riesgo real de homicidios nunca es igual a cero.

14.2.3 Resultados de la evaluación

Los autores realizaron un conjunto de regresiones utilizando el modelo de diferencias en diferencias con la variable dependiente del logaritmo de la tasa general de homicidios, calculada a través del método espacial bayesiano, comparando la evolución de las tasas masculinas con las femeninas. La reducción de las tasas femeninas, asociada al efecto de la LMP, fue significativamente superior a la de las tasas masculinas, lo que apunta a que ley consiguió contener la violencia letal de género en Brasil. El impacto de la ley habría sido de una reducción de aproximadamente 3,4%, en relación con la tasa original de homicidio femenino.

Específicamente, el efecto de la LMP fue aún más fuerte en los modelos cuya variable dependiente fueron apenas los homicidios dentro de las residencias, lo cual refuerza la validez discriminante de los resultados y fortalece la conclusión de un impacto específico en los homicidios por motivos de género. En el caso de los homicidios en residencias, la caída asociada a la LMP habría sido de aproximadamente 9,2%, en relación con la tasa original de homicidio femenino.

En otras palabras, este modelo confirmó la hipótesis del impacto de la LMP en la reducción de los homicidios entre mujeres relacionados con la violencia doméstica.

14.2.4 Limitaciones de la evaluación

En este tipo de evaluaciones de cambios repentinos que afectan a todo el universo, como una nueva ley, la amenaza a la validez interna radica, sobre todo, en la posibilidad de que se hayan dado otros fenómenos concomitantes a la aprobación de la ley (lo que se conoce como *la amenaza de historia*) que puedan afectar a los homicidios de uno y otro sexo de forma diferencial.

No hay duda de que el efecto más intenso sobre los homicidios ocurridos en las residencias, en comparación al efecto sobre los homicidios en general, fortalece la conclusión de un impacto positivo de la ley. Sin embargo, la evidencia sería aún más contundente si el modelo

estadístico hubiera comparado directamente los homicidios dentro y fuera de la residencia (por ejemplo, usando como variable dependiente la razón entre estos dos tipos de homicidio), pues la diferencia entre ambos debería ser afectada por la LMP.

Por otro lado, la elección de los suicidios por arma de fuego y de las muertes por alcohol como variables a ser controladas en el modelo es un tanto idiosincrática, en el sentido de que muchos otros fenómenos relevantes podrían también haber sido incluidos.

Por último, el cálculo de las tasas de homicidio a partir de un procedimiento espacial bayesiano significa que no se trabaja exactamente con los homicidios observados, sino con una estimativa.⁷³

14.3 Impacto de la prohibición temporal de porte de armas en Cali y Bogotá (1993-1997). Evaluación realizada por Villaveces, Cummings, Espitia, Koepsell, McKnight y Kellermann⁷⁴

14.3.1 Descripción del programa

En 1993, con una tasa de homicidios de 105 por cada 100.000 habitantes, el alcalde de Cali creó el Programa de Desarrollo, Seguridad y Paz, conocido como DESEPAZ, en un esfuerzo por contener la creciente tasa de asesinatos de la ciudad. En 1995, Bogotá, con una tasa de homicidios de 68 por cada 100.000 habitantes, también desarrolló un programa específico y sistemático contra la violencia.

Entre las muchas medidas que fueron tomadas en ambas ciudades, se decidió prohibir, por medio de una norma municipal, el porte de armas de fuego en espacios públicos durante algunos días de alto riesgo de violencia letal, generalmente en fines de semana. Estos días solían corresponder a feriados prolongados, a elecciones o a fines de semana en los que se daba el pago de salarios. Sin embargo, debido al esfuerzo demandado para fiscalizar el cumplimiento de la normativa, la intervención se aplicó solo en algunos de los días que cumplían las condiciones de riesgo. El público fue notificado de la prohibición a través de los medios de comunicación.

En Cali, la intervención se aplicó en 34 ocasiones, durante un total de 89 días; la intervención comenzó en noviembre de 1993 y se desarrolló de manera intermitente durante 1994. En Bogotá, la intervención se llevó a cabo en 22 ocasiones, durante un total de 67 días, en 3 intervalos: diciembre de 1995 a marzo de 1996, diciembre 1996 a febrero de 1997, y marzo y abril de 1997.

Estrategia de evaluación: Serie temporal interrumpida.

73 Específicamente, si existieran correlaciones espaciales en otras variables del modelo (como los suicidios por arma de fuego y las muertes por alcohol) y estas variables están relacionadas con los homicidios, esto podría en alguna medida sesgar los resultados.

74 Villaveces, A.; Cummings, P.; Espitia, V. E.; Koepsell, T. D.; McKnight, B. & Kellermann, A. L. (2000): "Effect of a ban on carrying firearms on homicide rates in 2 Colombian cities". *Jama*, 283(9), 1205-1209.

Durante los días de intervención, la policía estableció puntos de control en toda la ciudad y registró a las personas. También se realizaron búsquedas de armas durante las paradas de tráfico y otras actividades policiales. Los puntos de control fueron seleccionados por cada comandante de la policía local y generalmente estaban ubicados en áreas donde la policía creía que la actividad criminal era común. La policía también registró a personas que ingresaban a bares y clubes. Durante la prohibición, si se encontraba un arma de fuego en manos de un individuo que la adquirió legalmente, el arma era retirada temporalmente y la persona era multada. Las personas sin prueba de adquisición legal del arma eran arrestadas y sus armas eran confiscadas permanentemente.

14.3.2 Metodología de la evaluación

El diseño del estudio fue el de una serie temporal interrumpida con múltiples repeticiones. De hecho, la intervención se aplicó de forma intermitente en ambas ciudades, lo que permitió comparar períodos con intervención y períodos similares sin ella. Esto fue posible porque la intervención posee una temporalidad en su impacto ([ver apartado 5.4](#)), que corresponde en principio a un efecto inmediato y de corta duración, una vez que este efecto cesaría el día siguiente al fin de la prohibición.

El número observado de muertes y los datos de la población total se obtuvieron de las bases de datos de vigilancia epidemiológica en Cali y Bogotá. Los homicidios ocurridos durante cada hora se clasificaron por sexo, edad y si se utilizó o no un arma de fuego. Las tasas se calcularon en función del lugar de la intervención: para Cali en 1993 y 1994; y para Bogotá en 1995, 1996 y hasta agosto de 1997.

Para llevar a cabo el estudio, los casos se agruparon por categorías temporales: a) día de la semana; b) fin de semana con día de pago; c) fin de semana sin día de pago. Además, los casos fueron codificados según la época del año (agrupada en intervalos de 4 meses) y según el tiempo de la semana (agrupado en 28 intervalos iguales de 6 horas cada uno). Estas categorías fueron usadas como estratos en el proceso de estandarización de las tasas. El número total de muertes esperado en los días de prohibición del porte, bajo el supuesto de que la prohibición no hubiera existido, se obtuvo aplicando las tasas específicas para cada estrato (obtenidas en el período de no intervención) a la distribución de los días de intervención. El número realmente observado de muertes en los días de intervención se dividió entre ese número esperado, para conocer el posible impacto de la intervención y para poder calcular tasas de homicidios estandarizadas que equiparaban las condiciones entre los

períodos de intervención (como si fuera un grupo experimental) y los períodos de no intervención (como si fuera un grupo de control).

Este proceso es esencial, porque si se comparasen las tasas brutas de homicidios de los períodos de intervención con las del resto, se llegaría a la conclusión, equivocada y falaz, de que la prohibición de porte provocó muchas muertes, justamente porque la prohibición se daba en períodos de alto riesgo de violencia. El objetivo de la estandarización, por tanto, era poder comparar lo que sucedió en los períodos de prohibición de porte con lo que sucedía en períodos de riesgo similar en los que el porte de armas no se prohibió.

Además, modelos de regresión fueron también usados como otra estrategia para controlar el posible impacto de diversas variables en las tasas de homicidio. Fueron estimados modelos para los homicidios por arma de fuego y para el resto de los homicidios, en el supuesto de que el programa debería impactar básicamente en los primeros.

Por otro lado, si una persona planeaba matar a alguien con un arma de fuego, podría decidir que era demasiado arriesgado llevarlo a cabo durante el período de intervención y, por lo tanto, posponer el ataque a un momento en el que la prohibición de porte no estuviera en vigor. De ser así, cualquier reducción aparente en el homicidio durante los períodos de intervención podría ser seguida por un aumento compensatorio en los homicidios poco después de la intervención, o sea, un desplazamiento criminal.⁷⁵ Para verificar esta hipótesis, los autores del estudio analizaron la incidencia de homicidios durante un período de 7 días inmediatamente posterior a cada período de intervención, para ver si eran más elevados de lo esperado.

14.3.3 Resultados de la evaluación

La tasa estandarizada de homicidios en Cali en el período de intervención fue de 89 por cada 100.000 habitantes, mientras que en el período sin intervención fue de 107,5 para los mismos 100.000 habitantes. En consecuencia, esa reducción de 18,5 homicidios por 100.000 habitantes sería entonces atribuible a la prohibición del porte de armas. En Bogotá, la tasa estandarizada de homicidios también fue más baja en el período en el que la intervención estuvo en vigencia (54,2 por 100.000 habitantes, comparada con 59,3 por 100.000 habitantes en el período sin la prohibición de porte). En este caso, la reducción que habría sido causada por el programa sería de 5,1 homicidios por cada 100.000 habitantes.

Como ya se explicó, el análisis se dividió en tres categorías: días laborables, fines de semana con pago de salarios y fines de semana sin

⁷⁵ Para saber más sobre desplazamiento criminal [ver el apartado 11.3.](#)

pago de salarios. Durante la semana no hubo diferencias significativas en Cali entre períodos sin y con intervención. La diferencia sí fue significativa durante los fines de semana, tanto en aquellos con pago de salarios como en los otros sin pago. En Bogotá la intervención solo fue significativa en los fines de semana de pago.

De la misma forma, los modelos de regresión, que consideraban los períodos de fin de semana, feriados, días hábiles, períodos de 6 horas cada semana, mes del año y la tendencia general en las tasas de homicidios, confirmaron que la incidencia de homicidios durante los períodos de vigencia de la prohibición fue significativamente menor que en los períodos de no intervención, tanto en Bogotá como en Cali.

En los 7 días posteriores a los períodos de intervención, la incidencia de homicidios no fue estadísticamente diferente a la de intervalos comparables en que el programa también estaba ausente. En otras palabras, aparentemente no hubo un desplazamiento temporal de los homicidios, al menos para períodos inmediatamente posteriores a la prohibición.

En resumen, la evaluación encontró que el impacto de la intervención en Bogotá fue significativo solo en los fines de semana de pago, mientras que en Cali generó reducciones significativas en todos los fines de semana.

14.3.4 Limitaciones de la evaluación

Uno de los problemas de este tipo de evaluaciones es que no se conocen con exactitud los criterios de selección de los fines de semana en los que se prohibió el porte. El estudio supone que todos los fines de semana de pago, por ejemplo, eran igualmente peligrosos, lo que permite la comparación entre aquellos con y sin intervención. Sin embargo, no se sabe si había otros criterios para elegir los fines de semana y si en realidad eran más o menos peligrosos que los demás. De cualquier forma, cabe suponer que los gestores públicos elegirían los momentos que consideraban de mayor riesgo. En la medida en que esto sea cierto y que los gestores estuvieran en lo correcto, el estudio estaría subestimando el impacto potencial del programa, ya que habría sido empleado en momentos de mayor riesgo que aquellos con los que está siendo comparados.

Por otro lado, un resultado inquietante que la evaluación encontró es que, mientras en Bogotá el impacto era significativo apenas en los homicidios por arma de fuego como cabría esperar, en Cali la reducción era significativa para los “otros homicidios”, pero no para los homicidios

por arma de fuego. Este dato sorprendente es totalmente contrario a la lógica del modelo y levanta dudas sobre los resultados en general.

14.4 Impacto de un nuevo procedimiento de investigación de homicidios introducido en Bogotá (2016). Evaluación realizada por Collazos, Fergusson, La Rota, Mejía y Ortega⁷⁶

14.4.1 Descripción del programa

En Bogotá, en 2015, la tasa de homicidios era de 18 por cada 100.000 personas, la tasa de acusación del 22,4% y la tasa de condena del 74,6%. Esto significa que solo el 17% de los casos de homicidio producían una condena penal.

El procedimiento tradicional de investigación disponía a varios oficiales de policía operando en diferentes etapas bajo el liderazgo de fiscales que iban cambiando durante el proceso. De acuerdo con diagnósticos elaborados por la propia Fiscalía, esto generaba retrasos y dificultaba la continuidad entre las etapas del proceso penal. En este contexto, la Oficina de Asuntos de Política Pública del Procurador General desarrolló un nuevo procedimiento para la investigación de homicidios, a fin de evitar la pérdida de información entre los funcionarios e intentar optimizar el uso del tiempo, los recursos humanos y físicos.

La nueva organización buscaba superar la desconexión entre la investigación inicial, las acciones que tuvieron lugar antes del juicio y en el juicio. El cambio principal consistía en la estructuración de equipos que investigasen juntos los casos de homicidio, desde la asignación inicial hasta el momento del juicio. Cada caso estaba dirigido por un fiscal de distrito y era investigado por un equipo fijo compuesto por tres o cuatro investigadores técnicos y dos detectives, además del fiscal. La intervención se probó durante dos semanas en diciembre de 2015 y comenzó oficialmente el 20 de enero de 2016, continuando hasta el 4 de diciembre de 2016.

En teoría, la introducción de estos equipos fijos de investigación cambiaba el proceso de investigación de homicidios existente en cuatro dimensiones. En primer lugar, desde el momento en que los investigadores técnicos y los detectives eran enviados a la escena del crimen, se encontraban ya con el fiscal a cargo de la investigación, lo que les permitía comunicarse, recibir solicitudes directas y aclarar posibles preguntas. En segundo lugar, el mismo fiscal pasaba a ser responsable desde los actos iniciales hasta el enjuiciamiento. En tercer lugar, el fiscal podía reunirse con su equipo para discutir la estrategia de investigación y los supuestos criminales. Por fin, después de la denuncia el caso era

Estrategia de evaluación: Diseño experimental.

76 Collazos, D.; Fergusson, L.; La Rota, M.; Mejía, D. Ortega, D. (2017): *Impact Evaluation Report: A new homicide investigation policy in Bogotá*, Colombia. Mimeo.

transferido a un nuevo fiscal que se encargaba del juicio, de modo que el fiscal a cargo de las investigaciones no perdiera el tiempo con las audiencias.

En resumen, el nuevo procedimiento buscaba una continuidad del equipo a lo largo de toda la investigación de un caso, mejorando la comunicación entre sus integrantes y aumentando la estabilidad durante la conducción de los trabajos.

14.4.2 Metodología de la evaluación

Los investigadores utilizaron datos administrativos sobre homicidios reportados a la Oficina del Fiscal General, referidos a acciones ocurridas antes del 31 de diciembre de 2016. Las acciones y decisiones tomadas por investigadores técnicos, detectives y fiscales generaron los principales resultados utilizados en la evaluación. Los detectives e investigadores técnicos a menudo realizan actividades en la escena del crimen y procedimientos de investigación adicionales para encontrar y analizar evidencia física. Las acciones del fiscal incluyen instruir a los detectives para llevar a cabo acciones de investigación adicionales y tomar decisiones como presentar cargos contra una persona. Algunas acciones están condicionadas a las anteriores y solo pueden ocurrir en ciertas etapas del proceso. Por lo tanto, los autores de la evaluación estimaron los efectos de la intervención en las acciones de cada etapa del proceso investigativo.

La base de datos principal es el Sistema Penal Acusatorio Oral, un sistema por el cual las acciones de los casos procesados por la Fiscalía deben ser reportadas por los detectives y los fiscales. Otra fuente de información fue el Sistema de Gestión Integrada de la Policía Judicial, utilizado específicamente para informar a los detectives de las órdenes emitidas por los fiscales y que contiene información más completa sobre las acciones de investigación. Además, los evaluadores tuvieron acceso a informes escritos de los equipos de investigadores técnicos, detectives y fiscales para cada caso. Con el análisis de texto de estos datos, se realizaron ejercicios adicionales para explorar la naturaleza del trabajo de investigación llevado a cabo por los equipos, identificando y contando palabras claves y actividades específicas.

Posteriormente, los evaluadores crearon índices resumidos para la primera etapa del proceso (acciones de investigación). Además, se realizó una encuesta a investigadores técnicos, detectives y fiscales para ayudar a evaluar los mecanismos que podrían explicar el efecto de la intervención. A partir de esta encuesta, se crearon cuatro índices: motivación y *feedback*; funciones, efectividad y calidad de los miembros del equipo; trabajo en equipo; y coordinación y progreso.

En total fueron creados 6 equipos compuestos cada uno de ellos, como ya fue explicado, por un fiscal, investigadores técnicos y detectives. Un equipo asumía todos los casos que llegaban durante el período de 12 horas en que estaba de guardia. De los 6 equipos 4 fueron asignados a la nueva metodología de investigación (grupo experimental) y 2 permanecieron con la estrategia tradicional (grupo de control). Se llevó a cabo un esfuerzo sistemático para que los 6 equipos asumiesen, en la misma proporción, los 14 turnos en los que la semana estaba dividida, de forma que no hubiera diferencias temporales entre ellos.

La unidad de análisis en este estudio fue el caso penal. Los casos de homicidio fueron direccionados aleatoriamente al grupo de tratamiento y al grupo de control en función de cuál equipo estaba de guardia en el momento en que eran registrados.

Los efectos de la intervención sobre las acciones tomadas durante cada etapa de la investigación se estimaron mediante una regresión de mínimos cuadrados.

14.4.3 Resultados de la evaluación

El análisis de las estadísticas descriptivas de los datos administrativos sugiere un aumento en las acciones relacionadas con la investigación en el grupo de tratamiento en relación con el de control, medidas por los índices resumidos. Sin embargo, las estadísticas descriptivas mostraron que el tiempo transcurrido hasta que se presenta una denuncia es más alto en el grupo de tratamiento, lo cual iría en sentido contrario de la mayor agilidad buscada por el programa. Los autores aducen que la mayor demora se podría deber a que los equipos del grupo experimental se habrían dedicado a investigar de forma más concienzuda para garantizar una futura condenación, pero esto no deja de ser una hipótesis *ad-hoc*.

La proporción de casos que eran enviados al perito forense para que este confirmase si se trataba o no de un homicidio, es decir, casos en los que había dudas de si la muerte podía ser fruto de un accidente o un suicidio, era menor en el grupo experimental que en el grupo de control. Los autores interpretan este dato en el sentido de que los investigadores del grupo experimental prestarían más atención a los detalles de las escenas del crimen y a la evidencia.

Por otro lado, los pedidos realizados en el contexto del crimen y el análisis de la escena del crimen fueron evaluados utilizando tres variables claves: a) realización de interrogatorios (número de solicitudes de interrogatorios enviadas); b) localización de “personas de interés”

(testigos o sospechosos); y c) órdenes para verificar la identidad de los sospechosos. De hecho, el grupo de tratamiento registró un número mayor en cada una de estas tres variables.

Además, se estudiaron las peticiones de prórroga para la tramitación de los casos, las cuales resultaron ser menos frecuentes en el grupo experimental a pesar de que, como ya fue mencionado, el tiempo hasta la denuncia es más largo en este mismo grupo.

El análisis de los informes escritos de los equipos de investigadores técnicos, detectives y fiscales generó resultados no siempre convergentes. Por ejemplo, la extensión del texto es más corta en el grupo de tratamiento. La hipótesis de los autores en este punto es que los equipos bien coordinados se comunican con más fluidez y necesitan menos papeleo. En la frecuencia de los temas, hay resultados mixtos. Por un lado, hay más discusión sobre la primera persona interrogada y más mención a fiscales e investigadores de distrito. Ambos elementos sugieren que se prestó más atención a la información principal de la investigación, y la última, en particular, es consistente con un mejor flujo de información dentro del equipo, un objetivo primario de la intervención. Sin embargo, fue sorprendente constatar que había menos menciones a las evidencias en el grupo de tratamiento, cuando esto debía ser central en el trabajo investigativo.

Finalmente, los datos de la encuesta de referencia fueron analizados en relación con dimensiones como la motivación para el trabajo, la retroalimentación sobre su desempeño, la pertinencia y efectividad de las funciones de los miembros del equipo, y la calidad y la coordinación del equipo. La calidad y la coordinación del trabajo en equipo fue mayor en el caso de los funcionarios que trabajaron bajo el nuevo procedimiento de investigación.

14.4.4 Limitaciones de la evaluación

En relación con el mecanismo de aleatorización, aunque la unidad de análisis sea el caso, lo que realmente se aleatoriza es la asignación de cada uno de los 6 grupos de investigación, 4 atribuidos al nuevo procedimiento y 2 al antiguo. Este pequeño número de grupos torna los resultados frágiles, pues la composición de los integrantes de estos equipos puede tener un gran impacto en el resultado. Imaginemos, por ejemplo, que haya uno o dos fiscales entre los seis que sean especialmente activos o eficientes. El hecho de que estos fiscales sean asignados a uno u otro grupo puede sesgar todas las estimativas. Además, como los equipos del grupo de control son solo dos, la propia variación de la composición de

los equipos dentro del grupo, que debería ser la diferencia fundamental con el grupo experimental, es limitada, pues en principio solo habría 18 personas disponibles para componer estos equipos de 9.

Por otro lado, el hecho de trabajar con un número pequeño de personas en ambos grupos (aproximadamente 50), presumiblemente en contacto frecuente entre ellas, que debían ser muy conscientes de estar siendo evaluadas, podría afectar de alguna forma su desempeño. Esta conciencia de ser evaluado puede suponer una amenaza a lo que en la literatura se denomina *validez de constructo* de las conclusiones (Cook & Campbell, 1979).

Las encuestas realizadas a un número tan reducido de personas adolecen de los mismos problemas de composición individual de los grupos que el procedimiento de aleatorización antes descrito. Además, el cuestionario aplicado después de la intervención (en el postest) mostró una pérdida considerable de casos en el grupo de control, lo que no ocurrió con el grupo de tratamiento. En consecuencia, la composición que estaba equilibrada en el número de cuestionarios entre el grupo control y el grupo de tratamiento en el pretest, se volvió muy sesgada a favor del grupo de tratamiento (70% versus 30%) en postest. Esto implica que los resultados de la encuesta deben interpretarse con mucha cautela.

Por otro lado, la variable dependiente central, que debería ser la tasa de condenación de los homicidios atribuidos a cada grupo, no está disponible porque es necesario mucho más tiempo hasta que los casos lleguen a producir condenas. En su ausencia, los evaluadores recurren a una multiplicidad de variables dependientes (número de acciones registradas y tiempos transcurridos en los sistemas administrativos, una encuesta a los integrantes de los equipos, análisis textuales de los informes). Algunas de ellas parecen tener una validez o una interpretación cuestionable, como el número de menciones de determinadas palabras o la extensión de los textos, a pesar del notable esfuerzo realizado por los evaluadores.

En general, los resultados obtenidos son en ocasiones contradictorios y difíciles de reconciliar, como por ejemplo que el grupo experimental tarda más tiempo hasta las denuncias, pero pide menos extensiones de los plazos y sus informes son, además, más cortos. De la misma forma, el grupo experimental demora más hasta la denuncia, pero no hasta el momento de formalización del acta de la denuncia. Y sus informes hablan más de interrogatorios y de fiscales e investigadores, pero menos de evidencias. Por último, las hipótesis de sus informes parecen más semejantes unas de otras, como si las copiaran de unos casos a otros,

lo que no encaja con la idea de una investigación más profunda de los casos.

RESUMEN

Ejemplos presentados en el capítulo

Programa de prevención de homicidios	Estrategia de evaluación
14.1 Creación del Metrocable en Medellín	Diseño cuasiexperimental con grupo de control no equivalente
14.2 Ley “Maria da Penha” para inhibir la violencia contra las mujeres en Brasil	Modelo de diferencias en diferencias
14.3 Prohibición temporal de porte de armas en Cali y Bogotá	Serie temporal interrumpida
14.4 Nuevo procedimiento de investigación de homicidios en la Fiscalía de Bogotá	Diseño experimental

REFERENCIAS BIBLIOGRÁFICAS

- Abadie, A.; Diamond, A.; Hainmueller, J. (2014): "Comparative Politics and the Synthetic Control Method". *American Journal of Political Science*, 59(2), 495-510. Disponible en: <<https://doi.org/10.1111/ajps.12116>>.
- Aebi, M. (2008): *Temas de criminología*, Dickinson, Madrid.
- Aguirre, K.; Muggah, R. (2017): "Las agencias multilaterales y el enfoque de seguridad ciudadana en América Latina". *Revista CIDOB d'Afers Internacionals*, 116, 25-52. Disponible en: <https://www.cidob.org/es/articulos/revista_cidob_d_afers_internacionals/116/las_agencias_multilaterales_y_el_enfoque_de_seguridad_ciudadana_en_america_latina>.
- Australasian Evaluation Society (2013): Guidelines for the Ethical Conduct of Evaluations. Disponible en: <<https://www.aes.asn.au/>>.
- Berk-Seligson, S.; Orcés, D.; Pizzolitto, G.; Seligson, M. A.; Wilson, C. J. (2014): *Evaluación de impacto de la estrategia basada en la comunidad de USAID para la prevención de la delincuencia y la violencia en América Central: Informe regional para El Salvador, Guatemala, Honduras y Panamá*. Proyecto de Opinión Pública de América Latina (LAPOP) Vanderbilt University. Disponible en: <https://www.vanderbilt.edu/lapop/carsi/Regional_Report_Spanish_Final_W_121714.pdf>.
- Braga, A. A.; Papachristos, A. V.; Hureau, D. M. (2010): "The Concentration and Stability of Gun Violence at Micro Places in Boston, 1980–2008". *Journal of Quantitative Criminology*, 26, 33-53. Disponible en: <<https://doi.org/10.1007/s10940-009-9082-x>>.
- Bumpus, J.; Meyers, J.; Sanchez, P.; Alinikof, E.; Kizuka, K.; Yu, S.; Archundia, J.; Hua, A.; Johnson, K.; Lau, C.; Puncy, H. (2014): *Best Practices in Reducing Violent Homicide Rates: Honduras, El Salvador, México*. Woodrow Wilson School, Princeton University. Disponible en: <https://www.princeton.edu/sites/default/files/content/591g%20Homicide%20Reduction%20in%20Honduras_1.pdf>.
- Bush, K.; Duggan, C. (2015): *Evaluation in the extreme: research, impact and politics in violently divided societies*. International Development Research Centre, Sage, Nueva Delhi. Disponible en: <<https://www.idrc.ca/en/book/evaluation-extreme-research-impact-and-politics-violently-divided-societies>>.

- Campbell, D. T.; Stanley, J. C. (1963): *Experimental and Quasi-Experimental Designs for Research*. Houghton Mifflin.
- Cano, I. (2007): *¿Vivir sin armas? Evaluación del proyecto Municipios Libres de Armas, una experiencia arriesgada en un contexto de riesgo*. PNUD, San Salvador. Disponible en: <https://www.sv.undp.org/content/el_salvador/es/home/library/democratic_governance/-vivir-sin-armas--evaluacion-del-proyecto-municipios-libres-de-a.html>.
- Cano, I. (2007b). *Evaluación preliminar del Proyecto Barrio Seguro*. Laboratorio de Análisis de la Violencia, Universidad del Estado de Rio de Janeiro. Informe de acceso restringido.
- Cano, I.; Borges, D.; Ribeiro, E. (orgs.) (2012): *Os Donos do morro: uma avaliação exploratória do impacto das Unidades de Polícia Pacificadora (UPPs) no Rio de Janeiro*. Fórum Brasileiro de Segurança Pública. LAV/UERJ, Rio de Janeiro. Disponible en: <<http://www.lav.uerj.br/docs/rel/2012/RelatUPP.pdf>>.
- Cano, I.; Rojido, E. (2017): "Introducción: la singularidad de la violencia letal en América Latina". *Revista CIDOB d'Afers Internacionals*, 116, 7-24. Disponible en: <https://www.cidob.org/es/articulos/revista_cidob_d_afers_internacionals/116/introduccion_la_singularidad_de_la_violencia_letal_en_america_latina>.
- Cano, I.; Rojido, E. (2016): *Mapeo de programas de prevención de homicidios en América Latina y el Caribe* [Informe final]. Río de Janeiro: Laboratório de Análise da Violência, Fórum Brasileiro de Segurança Pública. Disponible en: <http://www.lav.uerj.br/docs/rel/2016/Relatorio_Final.pdf>.
- Cano, I.; Rojido, E.; Aguilar, J. (coords.) (2018): *Protocolo Técnico para Investigaciones de Campo en Contextos de Violencia*. Red de Conocimiento sobre Seguridad Ciudadana (CONOSE), San Salvador. Disponible en: <<http://conose.org/>>.
- Cano, I.; Rojido, E.; Rodrigues, R. (2017): *Evaluación del proyecto Paz y Justicia desarrollado por la Asociación por una Sociedad Más Justa en Honduras*. Laboratorio de Análisis de la Violencia, Universidad del Estado de Rio de Janeiro. Informe de acceso restringido.
- Cepal (2015): *Informe anual 2013-2014. El enfrentamiento de la violencia contra las mujeres en América Latina y el Caribe* (LC/G.2626), Santiago de Chile. Disponible en: <<https://www.cepal.org/es/publicaciones/observatorio-de-igualdad-de-genero-de-america-latina-y-el-caribe-oig-informe-anual-1>>.
- Cepal (2014): *Documento conceptual de los sistemas estadísticos de seguridad pública y justicia en los países de América Latina y el*

- Caribe*, p. 68. Disponible en: <<https://rtc-cea.cepal.org/sites/default/files/document/files/GT-Seguridad-SE-seguridad-publicaALC.pdf>>.
- Cerdá, M.; Morenoff, J. D.; Hansen, B. B.; Tessari Hicks, K. J.; Duque, L. F.; Restrepo, A.; Diez-Roux, A. V. (2012): "Reducing Violence by Transforming Neighborhoods: A Natural Experiment in Medellín, Colombia". *American Journal of Epidemiology*, 175(10).
- Cerqueira, D.; Vieira Martins Matos, M.; Antunes Martins, A. P.; Pinto, J. (2015): "Avaliando a efetividade da Lei Maria da Penha. Instituto de Pesquisa Econômica Aplicada (IPEA)". Texto para Discussão 2048, Brasília. Disponible en: <https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=24606&Itemid=383>.
- Cohen, L. E.; Felson, M. (1979): "Social Change and Crime Rate Trends: A Routine Activity Approach". *American Sociological Review*, 44(4), 588-608.
- Collazos, D.; Fergusson, L.; La Rota, M.; Mejía, D.; Ortega, D. (2017): *Impact Evaluation Report: A New Homicide Investigation Policy in Bogotá, Colombia*. Mimeo.
- Cook, T. D.; Campbell, D. T. (1979): *Quasi-experimentation: Design & Analysis Issues for Field Settings*. Houghton-Mifflin, Boston.
- Corradi, C.; Stöckl, H. (2014): "Intimate partner homicide in 10 European countries: Statistical data and policy development in a cross-national perspective". *European Journal of Criminology*, 11(5), 601-618. Disponible en: <<https://doi.org/10.1177/1477370814539438>>.
- Corradi, C.; Marcuello-Servós, C.; Boira, S. & Weil, S. (2016): "Theories of femicide and their significance for social research". *Current Sociology*, 64 (7), 975-995. Disponible en: <<https://doi.org/10.1177/0011392115622256>>.
- Dammert, L.; Salazar, F.; Montt, C.; González, P. (2010): *Crimen e inseguridad: indicadores para las Américas*. FLACSO Chile/Banco Interamericano de Desarrollo (BID), Santiago de Chile, p. 109. Disponible en: <<http://www.flacsochile.org/publicaciones/crimen-e-inseguridad-indicadores-para-las-americas/>>.
- Deus, A.; González, D. (2018): *Análisis de legislación sobre femicidio/feminicidio en América Latina y el Caribe e insumos para una ley modelo*. Informe de consultoría para MESECVI y ONU-Mujeres.
- DFID (2016): *Review of Ethics Principles and Guidance in Evaluation and Research*. Disponible en: <<https://www.gov.uk/government/publications/review-of-ethics-principles-and-guidance-in-dfid-evaluation-and-research>>.

- Dobrin, A. (2016): *Homicide Data Sources An Interdisciplinary Overview for Researchers*. Springer.
- Duailibi, S.; Ponicki, W.; Grube, J.; Pinsky, I.; Laranjeira, R.; Raw, M. (2004): *Prevenção de homicídios em Diadema, Brasil. A influência de novas políticas do Álcool*. Pacific Institute for Research and Evaluation. San Pablo.
- Duailibi, S.; Ponicki, W.; Grube, J.; Pinsky, I.; Laranjeira, R.; Raw, M. (2007): "The effect of restricting opening hours on Alcohol-Related Violence". *American Journal of Public Health*, 97(12). Disponible en: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2089093/>>.
- Dugan, L.; Nagin, D. S.; Rosenfeld, R. (2003): "Exposure reduction or retaliation? The effects of domestic violence resources on intimate partner homicide". *Law & Society Review*, 37(1), 169-198.
- Eckhardt, K.; Pridemore, W. A. (2009): "Differences in female and male involvement in lethal violence in Russia". *Journal of Criminal Justice*, 37(1), 55-64. Disponible en: <<https://doi.org/10.1016/j.jcrimjus.2008.12.009>>.
- Economic and Social Research Council (2015): *Framework for Research Ethics*. Updated January 2015. Disponible en: <<https://esrc.ukri.org/>>.
- Eisner, M. (2003): "Long-Term Historical Trends in Violent Crime". *Crime and Justice*, 30, 83-142.
- Eisner, M. (2009): "No effects in independent prevention trials: can we reject the cynical view?". *Journal of Experimental Criminology*, 5(2), 163-183. Disponible en: <<https://doi.org/10.1007/s11292-009-9071-y>>.
- Eisner, M. (2013): "What Causes Large-scale Variation in Homicide Rates?". *Aggression in Humans and Other Primates: Biology, Psychology, Sociology*. Walter de Gruyter.
- Eisner, M. (2015): *Cómo reducir los homicidios en un 50% en los próximos 30 años*. Notas de Homicidios 1. Instituto Igarapé, Rio de Janeiro. Disponible en: <<https://igarape.org.br/en/como-reducir-los-homicidios-en-un-50-en-los-proximos-30-anos/>>.
- Eisner, M. & Nivette, A. (2012): "How to reduce the global homicide rate to 2 per 100,000 by 2060", in: Welsh, B. & Loeber, R. (eds.): *The Future of Criminology*, New York: Oxford University Press, 219-228.
- Farrington, D. P.; Gottfredson, D. C.; Sherman, L. W.; Welsh, B. C. (2002): "Maryland Scientific Methods Scale", in: Farrington, D. P.; MacKenzie, D. L.; Sherman, L. W.; Welsh, B. C.: *Evidence-Based Crime Prevention*. Routledge, London.

- Fetterman, (1994): "Empowerment Evaluation". *Evaluation Practice*, 15(1). Jai Press.
- Fluehr-Lobban, C. (1994): "Informed Consent in Anthropological Research: We Are Not Exempt". *Human Organization*, 53(1), 1-10.
- Foguel, M. N. (2016): "Diferenças em Diferenças", en Menezes Filho, N. A.; de Xavier Pinto, C. C.: *Avaliação econômica de projetos sociais*. Fundação Itaú Social, São Paulo.
- Fundación Ideas para la Paz (2012): *Evaluación de impacto del Plan Nacional de Vigilancia Comunitaria por Cuadrantes*. Disponible en: <<http://www.ideaspaz.org/publications/posts/73>>.
- García, M. A. (2012): *Planificación urbana y seguridad ciudadana. Impacto de la creación de espacios públicos en la violencia*. Instituto Latinoamericano de Investigaciones Sociales (ILDIS), Caracas. Disponible en: <<http://www.ildis.org.ve/website/administrador/uploads/DocumentoEspaciosPublicosyPrevencionDelito.pdf>>.
- Gertler, P. J.; Martínez, S.; Premand, P.; Rawlings, L. B.; Vermeersch, C. M. J. (2011): *Impact evaluation in practice*. World Bank, Washington D.C. Disponible en: <<https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice>>.
- Gertler, P. J.; Martínez, S.; Premand, P.; Rawlings, L. B.; Vermeersch, C. M. J. (2017): *La evaluación de impacto en la práctica*, Segunda edición. Banco Interamericano de Desarrollo y Banco Mundial, Washington D.C. doi:10.1596/978-1-4648-0888-3. Disponible en: <<http://documentos.bancomundial.org/curated/es/478231485455706362/Impact-evaluation-in-practice>>.
- Guerette, R. T.; Bowers, K. J. (2009): "Assessing the Extent of Crime Displacement and Diffusion of Benefits: a Review of Situational Crime Prevention Evaluations". *Criminology*, 47(4), 1331-1368. Disponible en: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-9125.2009.00177.x>>.
- Guo, S.; Fraser, M. W. (2010): "Propensity Score Analysis. Statistical Methods and Applications". *Sage Advanced Quantitative Techniques in the Social Sciences*, 11, London.
- Hewlett Foundation (2014): *Benchmarks for Spending on Evaluation*. Disponible en: <https://hewlett.org/wp-content/uploads/2016/08/Benchmarks%20for%20Spending%20on%20Evaluation_2014.pdf>.
- Bellasio, J.; Sim, M.; Van Gorp, A.; Cox, K.; Ito, K.; Baruch, B.; Guerin, B.; Martin, A.; Grand-Clement, S.; Janta, B.; Saunders, C. L.; Smith, E.; Cochrane, G.; Henham, M. L. (2018): *Evaluating interventions that prevent or counter violent extremism: A practical guide*.

- Katz, C.; Amaya, L. E. (2015): *La tregua entre las pandillas como una forma de intervención sobre la violencia. Implicaciones en políticas y prácticas*. FUNDE, San Salvador. Disponible en: <<http://www.repo.funde.org/1042/>>.
- Kelling, G.; Pate, A.; Dickman, D.; Brown, C. (1974): *The Kansas City Preventive Patrol Experiment: A Technical Report*. Police Foundation. Disponible en: <<https://www.policefoundation.org/wp-content/uploads/2015/07/Kelling-et-al.-1974-THE-KANSAS-CITY-PREVENTIVE-PATROL-EXPERIMENT.pdf>>.
- Kivivuori, J.; Suonpää, K.; Lehti, M. (2014): "Patterns and theories of European homicide research". *European Journal of Criminology*, 11(5), 530-551. Disponible en: <<https://doi.org/10.1177/1477370814536833>>.
- Krisch, M.; Eisner, M.; Mikton, C.; Butchart, A (2015): *Global Strategies to Reduce Violence by 50% in 30 Years: Findings from the WHO and University of Cambridge Global Violence Reduction Conference 2014*. Cambridge, University of Cambridge. Disponible en: <<https://www.vrc.crim.cam.ac.uk/system/files/documents/violencereductionreport.pdf>>.
- Ioannou, M.; Hammond, L. (2015): "The changing face of homicide research: the shift in empirical focus and emerging research trends". *Journal of Criminal Psychology*, 5(3), 157-162. Disponible en: <<https://doi.org/10.1108/JCP-06-2015-0019>>.
- IPEA-FBSP: Instituto de Pesquisa Econômica Aplicada & Fórum Brasileiro de Segurança Pública (2019): *Atlas da violência 2019*. Brasília: Rio de Janeiro: São Paulo: Instituto de Pesquisa Econômica Aplicada, Fórum Brasileiro de Segurança Pública. Disponible en: <https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=34784&Itemid=432>.
- Lagarde, M. (2008): "Antropología, feminismo y política: violencia feminicida y derechos humanos de las mujeres", en M. Bullen y C. Diez (coords.): *Retos teóricos y nuevas prácticas*. Disponible en: <<https://www.ankulegi.org/wp-content/uploads/2012/03/0008Lagarde.pdf>>.
- Lenis, D.; Ronconi, L.; Schargrodsky, E. (2010): *The Effect of the Argentine Gun Buy-Back Program on Crime and Violence*. Disponible en: <<https://pdfs.semanticscholar.org/d061/dd214c9abacf80abb8e7cfcf71ec40c50835.pdf>>.
- Levy, H. (2012): *Youth Violence and Organized Crime in Jamaica: Causes and Counter- Measures. An Examination of the Linkages and Disconnections. Final Technical Report*. The University of the

- West Indies (UWI) - Institute of Criminal Justice and Security (ICJS), Kingston. Disponible en: <<https://idl-bnc-idrc.dspacedirect.org/bitstream/handle/10625/51348/IDL-51348.pdf>>.
- Maguire, E. R.; Oakley, M. T.; Corsaro, N. (2018): *Evaluating Cure Violence in Trinidad and Tobago*. Arizona State University - Inter-American Development Bank. Disponible en: <<https://publications.iadb.org/publications/english/document/Evaluating-Cure-Violence-in-Trinidad-and-Tobago.pdf>>.
- McDowall, D.; McCleary, R.; Meidinger, E.; Hay, R. A. (1980): *Interrupted Time Series Analysis*, SAGE. Disponible en: <<http://methods.sagepub.com/book/interrupted-time-series-analysis>>.
- McLean, A.; Lobban, S. B (2009): *Assessment of Community Security and Transformation: Programmes in Jamaica*. Report Comissioned by the Government of Jamaica. Disponible en: <<http://www.undp.org/content/dam/jamaica/docs/researchpublications/governance/CommunitySecurityAssessmentReport.pdf>>.
- MéxicoEvalúa(2018):*HomicidiosCDMX.Análisisespacialparalareducción de la violencia letal*. Disponible en: <<https://www.mexicoevalua.org/homicidioscdmx/documentos/5013HomicidiosCDMX.pdf>>.
- Miethe, T.; Regoeczi, W. (2004): *Rethinking Homicide Exploring the Structure and Process Underlying Deadly Situations*. Cambridge University Press, New York. Disponible en: <<https://www.cambridge.org/core/books/rethinking-homicide/AB686DF65352745D9FB1F0EE238CDEAC>>.
- Mill, J. S. (1888): *A System of Logic, Ratiocinative and Inductive*, 8th ed. Harper and Brothers, New York.
- ONU-ÚNETE (2013): *La regulación del delito de feminicidio/femicidio en América Latina y el Caribe*. Secretariado de la Campaña del Secretario General de las Naciones Unidas ÚNETE para poner fin a la violencia contra las mujeres Ciudad de Panamá, Panamá. Disponible en: <https://assets-compromissoeatitude-ipg.sfo2.digitaloceanspaces.com/2013/09/UN2013_regulacionfeminicidio_AnalsabelGaritaVilchez.pdf>.
- OMS-Organización Mundial de la Salud (2009): *Global Status Report on Road Safety: Summary*. Disponible en: <https://www.who.int/violence_injury_prevention/road_safety_status/2009/en/>.
- OMS-Organización Mundial de la Salud (2014): «67a Asamblea Mundial de la Salud. Resoluciones y decisiones». OMS, Ginebra (19 al 24 de mayo de 2014).

- OMS-Organización Mundial de la Salud (2015): *Clasificación estadística internacional de enfermedades y problemas relacionados con la salud*. 10.^a revisión. Washington, D.C. Disponible en: <http://www.paho.org/hq/index.php?option=com_tabs&view=article&id=3555&Itemid=3877&lang=es>.
- Patton, M. (1986): *Utilization-Focused Evaluation*. Second Edition. Newbury Park, CA, Sage.
- Patton, Michael Quinn (2008): *Utilization-Focused Evaluation*. Thousand Oaks, CA, Sage Publications.
- Peixoto, B.; Andrade, M.; Azevedo, J. P. (2008a): *Avaliação econômica do programa Fica Vivo: o caso piloto. Texto para discussão n.º 336*. UFMG/Cedeplar, Belo Horizonte.
- Peixoto, B.; Andrade, M.; Azevedo, J. P. (2008b): *Prevenção e controle de homicídios: uma avaliação de impacto no Brasil. Texto para discussão n.º 337*. UFMG/Cedeplar, Belo Horizonte.
- Radford, J.; Russell, D. (1992): *Femicide: The Politics of Woman Killing*. Twayne Publishers Inc., US.
- Ribeiro, E.; Borges, D.; Cano, I. (2015): *Calidad de los datos de homicidio en América Latina*. Rio de Janeiro. Disponible en: <http://homicidioslatam.org/wp-content/uploads/2018/03/Ribeiro-et-al-2015_esp-1.pdf>.
- Ribeiro, E.; Cano, I. (2016): “Vitimização letal e desigualdade no Brasil: Evidências em nível municipal”. *Civitas*, Porto Alegre, 16(2), 285-305, abr.-jun. 2016.
- Rojido, E.; Cano, I. (2016): “En el punto de mira: desafíos éticos y metodológicos de la investigación de campo en contextos de violencia”, en Gottsbacher, M.; de Boer, J. (coords.): *Vulnerabilidad y violencia en América Latina*. Editorial Siglo XXI, México DF. Disponible en: <<http://www.revistas.unam.mx/index.php/ras/article/view/64815>>.
- Sandoval, E. A. (2017): “Agendas criminales y procesos de paz en Colombia: El caso de la ‘donbernabilidad’ en Medellín”. *Revista Política y Estrategia*, (130), 57-100. Disponible en: <<https://www.politicayestrategia.cl/index.php/rpye/article/view/129/245>>.
- Shadish, W.; Cook, T.; Campbell, D. (2002): *Experimental and Quase-Experimental Designs for Generalised Causal Inference*. Houghton Mifflin, Boston.
- Shilan, C. (2017): *Intimate Partner Homicide Rates and Characteristics* [tesis doctoral]. Karolinska Institutet, Department of Clinical Neuroscience.

- Silveira, A. M.; Peixoto, B. (2010): “Manual de avaliação de programas de prevenção a violência”. Organização Pan-Americana da Saúde; Universidade Federal de Minas Gerais. Organização Pan-Americana da Saúde, Brasília.
- Spence, T.; Sinclair, G.; Hayman, A. (2006): Peace Management Initiative. PMI: a Review of the PMI. Final Report.
- Smit, P. R.; de Jong, R. R.; Bijleveld, C. C. J. H. (2012): “Homicide Data in Europe: Definitions, Sources, and Statistics”. In M. C. A. Liem; W. A. Pridemore (eds.): *Handbook of European Homicide Research: Patterns, Explanations, and Country Studies* (pp. 5-23).
- Tilley, N. (2009): *Crime Prevention*. Willan, Portland.
- Trochim, William M. K. (1984): *Research Design for Program Evaluation: The Regression-Discontinuity Approach*. Sage Publications, Beverly Hills.
- UNEG (2008a). *UNEG Ethical Guidelines for Evaluation*. Disponible en: <<http://www.unevaluation.org/document/detail/102>>.
- UNEG (2008b). *UNEG Code of Conduct for Evaluation in the UN System*. Disponible en: <<http://www.unevaluation.org/document/detail/100>>.
- UNODC-United Nations Office on Drugs and Crime (2014): *Global Study on Homicide: Trends, Contexts, Data. 2013*. Disponible en: <https://www.unodc.org/documents/gsh/pdfs/2014_GLOBAL_HOMICIDE_BOOK_web.pdf>.
- UNODC-United Nations Office on Drugs and Crime (2015): *Clasificación internacional de delitos con fines estadísticos*. Disponible en: <https://www.unodc.org/documents/data-and-analysis/statistics/crime/ICCS/ICCS_SPANISH_2016_web.pdf>.
- UNODC-United Nations Office on Drugs and Crime (2017): LABORATORY; SCIENTIFIC SECTION. *Evaluation Handbook: Guidance for Designing Independent Evaluation at UNODC*. United Nations Publications. Disponible en: <https://www.unodc.org/documents/evaluation/Evaluation_Handbook_new/UNODC_Evaluation_Handbook.pdf>.
- UNODC-United Nations Office on Drugs and Crime (2018). *Global Study on Homicide. Gender-Related Killing of Women and Girls*. Disponible en: <https://www.unodc.org/documents/data-and-analysis/GSH2018/GSH18_Gender-related_killing_of_women_and_girls.pdf>.
- UNODC-United Nations Office on Drugs and Crime (2019): *Global Study on Homicide. Executive Summary*. Disponible en: <<https://www.unodc.org/unodc/en/data-and-analysis/global-study-on-homicide.html>>.

- UNODC & HEUNI (2015): *Homicide and Gender*. Disponible en: <http://www.heuni.fi/material/attachments/heuni/projects/wd2vDSKcZ/Homicide_and_Gender.pdf>.
- Villaveces, A.; Cummings, P.; Espitia, V. E.; Koepsell, T. D.; McKnight B.; Kellermann, A. L. (2000): "Effect on a ban on carrying firearms on homicide rates in 2 Colombian cities". *JAMA*. 283(9), 1205-120.
- Walby, S.; Heiskanen, M.; Francis, B.; Strid, S.; Stöckl, H.; Helweg-Larsen, K.; Palmer, E.; Towers, J.; Corradi, C.; Mergaert, L.; Olive, P.; Balderston, S. (2017): *The Concept and Measurement of Violence Against Women and Men*. Policy Press. Disponible en: <https://doi.org/10.26530/OAPEN_623150>.
- Weisburd, D.; Farrington, D. P.; Gill, CH. (editors) (2016): *What Works in Crime Prevention and Rehabilitation Lessons from Systematic Reviews*. Springer Series on Evidence-Based Crime Policy, New York.
- Weisburd, D.; Wyckoff, L. A.; Ready, J.; Eck, J. E.; Hinkle, J. C.; Gajewski, F. (2006): "Does crime just move around the corner? A controlled study of spatial displacement and diffusion of crime control benefits". *Criminology*, 44(3), 549-592. Disponible en: <<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1745-9125.2006.00057.x>>.
- Weiss, C. H. (1995): "Nothing as Practical as Good Theory: Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families." In J. Connell; A. Kubisch; L. B. Schorr; C. H. Weiss (eds.): *New Approaches to Evaluating Community Initiatives*. Aspen Institute, Washington DC.
- Welsh, B.; Farrington, D. (2007): *Preventing Crime: What Works for Children, Offenders, Victims and Places*. Springer, New York.
- Wholey, J. S. (1994): "Assessing the Feasibility and Likely Usefulness of Evaluation", in J. S. Wholey; H. P. Hatry; K. E. Newcomer: *Handbook of Practical Program Evaluation*. Jossey-Bass, San Francisco.

La *Guía de evaluación para programas de prevención de homicidios en América Latina y el Caribe* tiene el propósito central de incentivar la realización de evaluaciones de calidad de los programas de prevención de homicidios.

Sin ser un manual técnico de evaluación ni un texto académico, la *Guía* aspira a facilitar la comprensión de los principales aspectos que componen la evaluación de programas de prevención al homicidio y a orientar a todos los actores relevantes sobre qué medidas deben tomarse para garantizar una evaluación cualificada y transparente.

El texto está pensado para cuatro tipos de público: formuladores y ejecutores de proyectos de prevención de homicidios; personas en cargos públicos y electivos que deben tomar decisiones sobre aprobación de proyectos y distribución de recursos; entidades financiadoras de programas de prevención de homicidios; y sociedad civil en general.

G uía de evaluación para programas de prevención de homicidios en América Latina y el Caribe



LABORATÓRIO DE
ANÁLISE DA VIOLÊNCIA