



USAID
DEL PUEBLO DE LOS ESTADOS
UNIDOS DE AMÉRICA

REVISANDO LAS RESPUESTAS A LOS PROBLEMAS:

¿FUNCIONÓ? UNA INTRODUCCIÓN
PARA SOLUCIONADORES
DE PROBLEMAS POLICIALES,
2DA. EDICIÓN

JOHN E. ECK



**GUÍAS PARA LA POLICÍA ORIENTADAS
A LA SOLUCIÓN DE PROBLEMAS**

De la serie de: **Herramientas para la
Atención de Problemas N.º 1**

Este proyecto fue apoyado por el acuerdo n.º 2013-DG-BX-K002 otorgado por la Oficina de Asistencia Judicial, Departamento de Justicia de EE.UU. a través de CNA, una organización de investigación y análisis sin fines de lucro ubicada en Arlington, Virginia, EE.UU. Las opiniones contenidas en el presente documento son las de su(s) autor(es) y no representan necesariamente la posición oficial o las políticas del Departamento de Justicia de los Estados Unidos. Las referencias a organismos, empresas, productos o servicios específicos no deben considerarse como una aprobación del producto por parte del autor o autores o del Departamento de Justicia de los Estados Unidos. Más bien, las referencias son ilustraciones para complementar la discusión de los temas.

La primera edición de esta publicación contó con el apoyo del acuerdo de cooperación n.º 99-CK-WX-K004 de la Oficina de Servicios Policiales Orientados a la Comunidad, del Departamento de Justicia de los Estados Unidos.

Las referencias de internet citadas en esta publicación eran válidas en la fecha de su publicación. Dado que las URL y los sitios web cambian constantemente, ni el autor ni la Oficina de Asistencia Judicial pueden garantizar su validez actual.

© 2017 Arizona Board of Regents. El Departamento de Justicia de los Estados Unidos se reserva una licencia libre de regalías, no exclusiva e irrevocable para reproducir, publicar o utilizar de cualquier otra forma, y autorizar a otros a utilizar esta publicación para fines del Gobierno Federal. Esta publicación puede distribuirse y utilizarse libremente con fines no comerciales y educativos.

Diciembre de 2017

Este documento es una traducción de inglés a español, por lo que los autores, ni el Departamento de Justicia de EE.UU., ni el Center for Problem-Oriented Policing son responsables de la exactitud de esta traducción.

Esta Guía es posible gracias al apoyo del pueblo de los Estados Unidos, a través de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID). Su contenido es responsabilidad de los autores y no refleja necesariamente el punto de vista de USAID o del Gobierno de los Estados Unidos de América.

06 Acerca de esta guía

- 06 Acerca de la serie de herramientas para la resolución de problemas

09 Agradecimientos

10 Introducción

- 10 De qué trata esta guía
11 Guías relacionadas de la serie de herramientas para la resolución de problemas
12 Cómo ayuda la evaluación a la toma de decisiones policiales

17 El papel de la evaluación en la resolución de problemas

20 Tipos de evaluaciones

- 20 Evaluaciones del proceso
21 Evaluaciones de impacto
22 Interpretación de las evaluaciones del proceso y de impacto

24 Realización de evaluaciones de impacto

24 Mediciones

- 25 *Mediciones cuantitativas*
25 *Mediciones cualitativas*
25 *Mapas*
27 *Validez de medición*
28 *Selección de mediciones válidas*

29 Criterios para alegar la causalidad

- 31 *Una explicación plausible de cómo la respuesta reduce el problema*
31 *La magnitud del problema y el nivel de la respuesta están relacionados*
32 *La respuesta al problema es anterior a su disminución*

35 Diseños

- 35 *Diseños pre-post*
40 *Diseños de series temporales*

CONTENIDO

- 43 *Combinación y selección de diseños*
- 44 *Examinar cómo funciona la respuesta*
- 47 *Desplazamiento y difusión de beneficios*

48	Conclusiones
50	Apéndice A
50	Los efectos del número de períodos en la validez de las conclusiones de la evaluación
54	Apéndice B: diseños con y sin grupos de control
55	Diseño de comparación estática
57	Diseño pre-post sin grupo de control
57	Diseño pre-post con grupo de control
60	Diseño de series temporales
62	Diseños de series temporales múltiples
64	Apéndice C: lista de verificación para la evaluación de la resolución de problemas
72	Apéndice D: resumen de los puntos fuertes y débiles de los diseños de evaluación
74	Referencias
76	Sobre el autor
77	Lista de lecturas recomendadas
80	Notas finales

Tablas

- 23 Tabla 1: interpretación de los resultados de las evaluaciones del proceso y de impacto
- 35 Tabla 2: tipos de diseño de evaluación
- 45 Tabla 3: la respuesta puede haber desencadenado uno o varios de esos mecanismos para reducir la actividad de prostitución
- 60 Tabla B1: calculo de la eficacia con diseño pre-post con control
- 67 Tabla C1: ¿qué diseño de evaluación tiene más sentido?
- 70 Tabla C2: interpretación de los resultados de las evaluaciones del proceso
- 71 Tabla C3: interpretación de los resultados de las evaluaciones del proceso y de impacto (otros diseños)

Figuras

- 13 Figura 1: cómo ayuda la evaluación a la toma de decisiones policiales
- 18 Figura 2: resolución de problemas y planificación de la evaluación
- 29 Figura 3: trazado de la calle antes y después de una respuesta a la prostitución
- 33 Figura 4: explicaciones alternativas
- 36 Figura 5: ejemplo de medición del impacto en un diseño pre-post
- 38 Figura 6: problemas con un diseño pre-post
- 42 Figura 7: medición del impacto en un diseño de series temporales interrumpidas
- 51 Figura A1: diseño pre-post de dos periodos
- 51 Figura A2: diseño de series temporales de nueve periodos (con trayectoria proyectada de problemas)
- 52 Figura A3: diseño de series temporales de dieciséis periodos (con trayectoria proyectada del problema)
- 53 Figura A4: diseño de series temporales de cuarenta periodos (con el número promedio de sucesos por periodo)
- 56 Figura B1: diseño de comparación estática
- 58 Figura B2: diseño pre-post
- 59 Figura B3: diseño pre-post con control
- 61 Figura B4: diseño de series temporales
- 63 Figura B5: diseño de series temporales múltiples

01 ACERCA DE ESTA GUÍA

ACERCA DE LA SERIE DE HERRAMIENTAS PARA LA RESOLUCIÓN DE PROBLEMAS

Las Herramientas para la solución de problemas son una de las tres series de las Guías policiales orientadas a solución de problemas. Las otras dos son las Guías sobre problemas específicos y las Guías de respuesta.

Las *Guías policiales orientadas a la solución de problemas* resumen los conocimientos sobre cómo puede la policía reducir los daños causados por problemas concretos de delincuencia y desórdenes públicos. Son guías para prevenir problemas y mejorar la respuesta global a los incidentes, no para investigar delitos o gestionar incidentes concretos. Tampoco cubren todos los detalles técnicos sobre cómo poner en práctica respuestas específicas. Las guías están escritas para policías de cualquier rango o puesto que deban abordar los problemas específicos que cubren las guías. Las guías serán más útiles para los oficiales que:

- **Comprenden los principios y métodos policiales básicos orientados a la solución de problemas.**
- **Pueden analizar los problemas a profundidad.**
- **Están dispuestos a considerar nuevas formas de llevar a cabo los asuntos policiales.**
- **Entienden el valor y los límites del conocimiento de la investigación.**
- **Están dispuestos a trabajar con otras agencias comunitarias y gubernamentales para encontrar soluciones efectivas a los problemas.**

Una literatura amplia de índole técnico y científico cubre cada una de las técnicas abordadas en las *Herramientas para la solución de problemas*. El objetivo de las guías es proporcionar sólo la información suficiente sobre cada técnica para que la policía y otras personas puedan utilizarla en el proceso de la resolución de problemas.

En la mayoría de los casos, la información recopilada durante un proyecto de resolución de problemas no tiene por qué someterse a un riguroso escrutinio científico. Cuando la policía necesita tener una mayor confianza en los datos, puede requerir la ayuda de expertos para utilizar la técnica. Los departamentos universitarios de sociología, psicología y justicia penal suelen ofrecer este tipo de ayuda.

Las necesidades de información de un mismo proyecto pueden ser muy diversas, y con frecuencia será necesario utilizar diversas técnicas de recopilación de datos para satisfacerlas. Del mismo modo, podría necesitarse una variedad de técnicas analíticas diferentes para analizar los datos. La policía y los analistas de la delincuencia podrían no estar familiarizados con algunas de las técnicas, pero el esfuerzo invertido en aprender a utilizarlas puede marcar la diferencia en el éxito de un proyecto.

	<p>Las Guías de Respuesta resumen el conocimiento sobre si la policía debe usar ciertas respuestas para abordar varios problemas de delincuencia y desorden, y sobre los efectos que pueden esperar. Cada guía incluye los siguientes pasos:</p>
	<p>Describe la respuesta</p>
	<p>Habla sobre las diversas formas en que la policía podría aplicar la respuesta</p>
	<p>Explica como la respuesta está diseñada para reducir el crimen y el desorden</p>
	<p>Examina el conocimiento de la investigación sobre la respuesta</p>
	<p>Aborda las posibles críticas y las consecuencias negativas que podrían derivarse de la respuesta</p>
	<p>Describe como ha policía a aplicado la respuesta a problemas específicos de delincuencia y desorden, y con qué efectos</p>

Estas guías enfatizan la efectividad y la imparcialidad como las principales consideraciones que la policía debe tener en cuenta al elegir las respuestas, pero reconocen que no son las únicas. La policía usa métodos particulares por diferentes razones además de —o independiente a— su efectividad e imparcialidad. Las actitudes y valores de la comunidad y las personalidades de los tomadores de decisiones clave a veces exigen diversos enfoques para abordar los problemas de delincuencia y desorden. Algunos individuos y comunidades prefieren respuestas orientadas a la aplicación, mientras que otros optan por enfoques colaborativos, orientados a la comunidad o a la reducción de daños. Estas guías no alterarán necesariamente esas preferencias, pero están destinadas a brindar mejor información.

Estas guías se han basado en resultados de investigaciones y prácticas policiales en los Estados Unidos, el Reino Unido, Canadá, Australia, Nueva Zelanda, los Países Bajos y Escandinavia. Aunque las leyes, costumbres y prácticas policiales varían de un país a otro, es evidente que la policía en todas partes experimenta problemas comunes. En un mundo cada vez más interconectado, es importante que la policía esté al tanto de las investigaciones y de las prácticas exitosas más allá de sus propios países.

Cada guía se basa en una revisión exhaustiva de la literatura de investigación y de la práctica policial ya documentada. Cada guía es revisada, antes de la publicación, por sus similares de manera anónima, por un oficial de policía de primera línea, un director de policía y un investigador. La Organización de Análisis e Investigación (CNA, por sus siglas en inglés) solicita las revisiones y gestiona el proceso de forma independiente.

Para obtener más información, visite el Centro para el Trabajo Policial Orientado a los Problemas en línea en www.popcenter.org. Este sitio web ofrece acceso gratuito a los siguientes recursos:

- La serie Guías para Problemas Específicos.
- La serie Guías de Respuesta Complementaria y Herramientas para Resolver Problemas.
- Publicaciones especiales sobre análisis criminal y vigilancia del terrorismo.
- Información instructiva sobre trabajo policial orientado a problemas y temas relacionados.
- Un ejercicio interactivo de formación policial orientado a la resolución de problemas.
- Módulo Interactivo de Análisis de Problemas.
- Acceso en línea a importantes investigaciones y prácticas policiales.
- Información sobre conferencias policiales orientadas a problemas y programa de premios.

02 AGRADECIMIENTOS

Las *Guías para la policía orientadas a la resolución de problemas* son elaboradas por el *Center for Problem-Oriented Policing* de la Universidad Estatal de Arizona bajo la dirección de Michael S. Scott.

Aunque cada guía tiene un autor principal, otros miembros del equipo del proyecto, personal de CNA y BJA y revisores anónimos contribuyeron a cada guía proponiendo textos, recomendando investigaciones y ofreciendo sugerencias sobre formato y estilo.

El equipo del proyecto que elaboró la serie de guías se conformó por Herman Goldstein, Ronald V. Clarke, John E. Eck, Michael S. Scott, Rana Sampson y Deborah Lamm Weisel. Miembros de los departamentos de policía de San Diego, California, National City, California y Savannah, Georgia aportaron comentarios sobre el formato y el estilo de las guías en las primeras fases del proyecto.

Vivian Elliott supervisó el proyecto para CNA. Phyllis Schultze realizó la investigación para la guía en la Biblioteca de Justicia Penal de la Universidad de Rutgers. Maurine Dahlberg de CNA editó esta guía.

03 INTRODUCCIÓN

El objetivo de la evaluación de los esfuerzos de solución de problemas es ayudar a los responsables de la policía a tomar mejores decisiones. Las evaluaciones responden a dos preguntas concretas: ¿Ha disminuido el problema? En caso afirmativo, ¿fue la respuesta prevista lo que provocó esta disminución? La respuesta a la primera pregunta ayuda a los responsables de la toma de decisiones a determinar si se puede poner fin al esfuerzo de solución de un problema y si los recursos pueden reasignarse a otros problemas. La respuesta a la segunda pregunta ayuda a los responsables de la toma de decisiones a determinar si la respuesta debería utilizarse de nuevo para abordar otros problemas similares.

DE QUÉ TRATA ESTA GUÍA

Esta guía pretende ayudar al lector a diseñar evaluaciones que puedan responder a estas dos preguntas. Esta se ha escrito para los oficiales de policía y otros responsables de evaluar la eficacia de las respuestas a los problemas. Se presupone que el lector tiene un conocimiento básico de la actuación policial orientada a la solución de problemas y del proceso de solución de problemas, incluyendo el proceso SARA (acrónimo de Exploración, Análisis, Respuesta y Evaluación, [Scanning, Analysis, Response, and Assessement, en inglés]). Está diseñado para ser útil a los lectores que no tienen ni experiencia en evaluación ni formación en métodos de evaluación e investigación académica.

Esta guía también presupone que el lector no cuenta con ayuda externa. No obstante, el lector debería buscar el consejo y la ayuda de investigadores académicos con formación y experiencia en evaluación, sobre todo si el problema que se aborda es grande y complejo. Un evaluador externo independiente puede ser especialmente útil si existe controversia sobre la utilidad de la respuesta.

A lo largo de esta guía, se hace referencia a la importancia de distinguir entre estas dos preguntas:

- ¿Ha disminuido el problema tras la respuesta?
- ¿Fue la respuesta lo que provocó la disminución?

Es probable que sea más importante para usted responder a la primera pregunta que a la segunda.

Esta guía complementa las guías de las series *Guías de problemas específicos* y *Guías de respuestas* de la serie de *Guías para la policía orientadas a la atención de problemas*. Cada guía de problemas específicos describe respuestas a un problema concreto y sugiere formas de medirlo. Cada guía de respuesta describe cómo y si esa respuesta funciona a la hora de abordar diversos tipos de problemas. Aunque esta guía está diseñada para trabajar con estas *Guías de problemas específicos* y *de respuestas*, los lectores deberían ser capaces de aplicar los principios de evaluación en cualquier proyecto de solución de problemas. Dado que esta guía es una introducción a un tema complejo, omite mucho de lo que se encontraría en un texto avanzado sobre evaluación.^a Los lectores que deseen profundizar en el tema de la evaluación deben consultar la lista de lecturas recomendadas que figura al final de esta guía.

GUÍAS RELACIONADAS DE LA SERIE DE HERRAMIENTAS PARA LA RESOLUCIÓN DE PROBLEMAS

Esta guía complementa otras de la serie Herramientas para la resolución de problemas. Estas guías abordan diversos aspectos de las cuatro fases de la resolución de problemas.^b

Fase de exploración:

- Identificación y definición de los problemas policiales (Guía n.º 13)

Fase de análisis:

- Investigación de un problema (Guía n.º 2)
- Utilización de las entrevistas con delinquentes para fundamentar la resolución de problemas policiales (Guía n.º 3)
- Análisis de la victimización reiterada (Guía n.º 4)

^a Quedan específicamente excluidas de esta discusión las menciones a la teoría de la medición, las pruebas de significación y la estimación estadística. Una monografía de esta extensión no puede describir esas cuestiones con suficiente detalle para que resulten útiles al lector.

^b Algunas guías abordan aspectos de más de una fase del modelo de resolución de problemas.

- Asociación con empresas para abordar problemas de seguridad pública (Guía n.º 5)
- Comprender las instalaciones de riesgo (Guía n.º 6)
- Utilización de la prevención de la delincuencia mediante el diseño ambiental en la resolución de problemas (Guía n.º 8)
- Mejora de la capacidad de resolución de problemas de las unidades de análisis de la delincuencia (Guía n.º 9)
- Análisis y respuesta a la reincidencia (Guía n.º 11)
- Comprender el robo de "productos llamativos" (Guía n.º 12)

Fase de respuesta:

- Análisis de la victimización reiterada (Guía n.º 4)
- Asociación con empresas para abordar problemas de seguridad pública Guía n.º 5)
- Comprender las instalaciones de riesgo (Guía n.º 6)
- Implementación de respuestas a los problemas (Guía n.º 7)
- Utilización de la prevención de la delincuencia mediante el diseño ambiental en la resolución de problemas (Guía n.º 8)
- Análisis y respuesta a la reincidencia (Guía n.º 11)
- Comprender el robo de "productos atractivos" (Guía n.º 12)

Fase de evaluación:

- Análisis de la victimización reiterada (Guía n.º 4)
- Utilización de la prevención de la delincuencia mediante el diseño ambiental en la resolución de problemas (Guía n.º 8)
- Análisis del desplazamiento y la difusión de la delincuencia (Guía n.º 10)
- Análisis y respuesta a la reincidencia (Guía n.º 11)
- Comprender el robo de "productos atractivos" (Guía n.º 12)

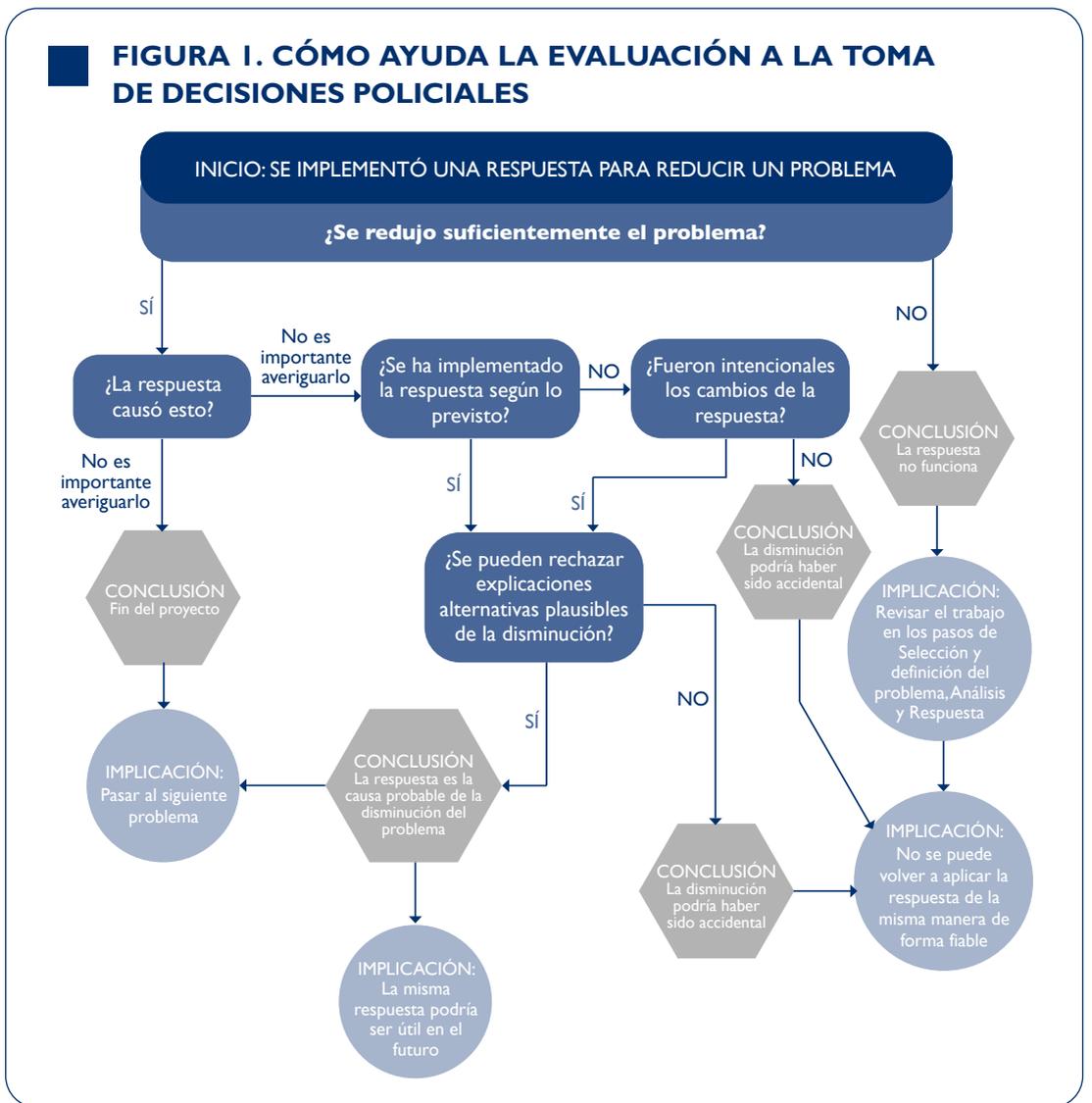
CÓMO AYUDA LA EVALUACIÓN A LA TOMA DE DECISIONES POLICIALES

En cualquier esfuerzo de resolución de problemas deben tomarse dos decisiones clave. En primer lugar, ¿ha disminuido el problema lo suficiente como para que se pueda reducir el esfuerzo de resolución del problema y aplicar los recursos policiales en otro lugar?

Si el problema no ha disminuido sustancialmente, el trabajo no ha terminado. En tal caso, la decisión más apropiada puede ser volver a analizar el problema y desarrollar una nueva respuesta.

Además, hay que alertar a los futuros encargados de la resolución de problemas para que puedan desarrollar mejores respuestas a problemas similares. Si el problema ha disminuido sustancialmente, puede existir una necesidad limitada de continuar con el esfuerzo de resolución de problemas más allá de la supervisión del problema y el seguimiento de cualquier respuesta de mantenimiento que puedan ser necesarios. Esta primera decisión de determinar cuándo se ha terminado de resolver el problema es el objetivo principal de esta guía.

FIGURA I. CÓMO AYUDA LA EVALUACIÓN A LA TOMA DE DECISIONES POLICIALES



En segundo lugar, si el problema disminuyó sustancialmente, ¿fue la respuesta prevista lo que provocó la disminución? Si esta disminución se debe, al menos en parte, a la respuesta, podría ser útil aplicar una respuesta similar a problemas similares. Si no puede comprobar de forma convincente que la respuesta causó la disminución del problema, podría no ser útil volver a aplicar la respuesta a problemas similares.

Entonces, las decisiones futuras sobre si aplicar o no la respuesta dependen, en parte, de la información de la evaluación. En este sentido, la evaluación es esencial en el aprendizaje organizativo de la policía. Sin evaluaciones, los responsables de la resolución de problemas pueden repetir sus errores o los de otros, o no beneficiarse de sus éxitos o los de otros.

El proceso comienza con la aplicación de alguna respuesta destinada a reducir sustancialmente el problema. El significado de "sustancialmente" depende de la naturaleza del problema y de los objetivos de los responsables de la toma de decisiones. La primera pregunta es si el problema ha disminuido sustancialmente. Si la respuesta es negativa, está claro que la respuesta no está funcionando lo suficientemente bien. Suponiendo que haya transcurrido el tiempo necesario para confiar en que la respuesta ha tenido tiempo de surtir efecto, la implicación es clara: hay que volver a etapas anteriores del proceso SARA y hacer revisiones. Además, ahora dispone de información que indica que esta respuesta no debería recomendarse en el futuro. Supongamos que el problema ha disminuido sustancialmente. Si no necesita determinar que la respuesta ha sido la responsable de esta disminución, puede dar por terminado el proyecto y pasar al siguiente problema. Pero si sigue este camino, no sabrá si la respuesta es útil para utilizarla en el futuro.

Si es importante averiguar si la respuesta ha provocado la disminución del problema, la siguiente pregunta es si la respuesta se aplicó según lo previsto. Si no fue así y las alteraciones de la respuesta no fueron intencionadas, no se sabe por qué disminuyó el problema. Las alteraciones no intencionadas incluyen el fracaso de una corporación de policía a la hora de llevar a cabo la función asignada debido a una supervisión deficiente, desviaciones del plan por parte de los organismos asociados que se basan en circunstancias imprevistas (recortes presupuestarios, el nombramiento de un líder que no simpatiza con la respuesta o la ineptitud administrativa). El éxito en estas circunstancias es bienvenido, pero no puede llevarse el mérito y no tiene motivos suficientes para recomendar la respuesta ante problemas similares.

Si la respuesta se aplicó según lo previsto o las revisiones de la respuesta fueron deliberadas, la siguiente pregunta es si algo ajeno a la respuesta podría haber causado la disminución del problema. En principio, nunca se puede estar seguro. En la práctica, esta pregunta se reduce a: "¿Puedo rechazar de forma creíble todas las explicaciones alternativas específicas plausibles de la disminución del problema?".

Si la respuesta es "no", es posible que algo distinto de la respuesta sea responsable de la disminución. Entonces, no puede recomendar definitivamente el uso de la respuesta para futuros problemas. Cualquier recomendación de este tipo debe hacerse con cautela. Si la respuesta es "sí", tiene pruebas razonablemente sólidas de que esta respuesta podría volver a funcionar.

Tenga en cuenta las salvedades en el lenguaje. Nunca podemos estar seguros de que una respuesta vaya a funcionar igual en el futuro. En cambio, podemos pensar que las recomendaciones son como las apuestas: Si apostara dinero por el resultado de mi recomendación, ¿tengo más probabilidades de ganar la apuesta o de perderla? Las evidencias le ayudan a ganar más de estas apuestas de las que pierde, pero nunca las ganará todas.

Para llegar a conclusiones sensatas, es necesario comprender en detalle tres cosas: la naturaleza del problema, la forma en que se supone que la respuesta reducirá el problema (véase más adelante) y el contexto en el que se ha aplicado la respuesta. Por este motivo, el proceso de evaluación comienza en cuanto se identifica el problema por primera vez durante la fase de selección y definición.

En esta guía se analizan dos diseños sencillos: pre-post y series temporales interrumpidas. El diseño pre-post sólo es útil en el primer tipo de decisión, es decir, si hay que poner fin a la resolución de un problema. El diseño de series temporales puede ayudar en ambos tipos de decisiones. Los diseños que implican grupos de comparación (o "control") se describen en un apéndice de esta guía y no en el texto principal. Puede resultar difícil para un encargado de la solución de problemas aplicarlos con éxito sin recibir más asesoramiento del que puede proporcionarse en esta guía. No obstante, estos diseños pueden suministrar la información necesaria para ayudar a tomar el segundo tipo de decisión: si se volverá a utilizar la respuesta en circunstancias similares.

Esta guía está organizada del siguiente modo.

- El cuerpo de la guía describe cuestiones fundamentales para desarrollar evaluaciones sencillas pero útiles.
- La lista de lecturas recomendadas enlaza esta guía con libros más técnicos sobre evaluación. Muchos de ellos aclaran la terminología.
- Los apéndices amplían el material presentado en el texto y sólo deben examinarse una vez leído este.
- El Apéndice A utiliza un ejemplo ampliado para mostrar por qué la evaluación de las respuestas a lo largo de periodos más prolongados permite comprender mejor la eficacia de la respuesta.

- El Apéndice B describe cinco diseños que utilizan datos de una evaluación rigurosa de un proyecto de resolución de problemas. (Incluye tres diseños que no se tratan en el cuerpo de la guía: uno que nunca debe utilizar y dos diseños más avanzados).
- El Apéndice C contiene una lista de verificación diseñada para guiar al encargado de la resolución de problemas a través del proceso de evaluación, ayudar a seleccionar el diseño más aplicable y extraer interpretaciones razonables de los resultados de la evaluación.
- El Apéndice D ofrece un resumen de los puntos fuertes y débiles de los diseños.

En resumen, esta guía explica en lenguaje ordinario los aspectos de los métodos de evaluación más importantes para la policía a la hora de abordar los problemas. En la siguiente sección, examinaremos cómo encaja la evaluación en el proceso SARA de resolución de problemas. A continuación, examinaremos los dos tipos principales de evaluación: de proceso y de impacto.

04 EL PAPEL DE LA EVALUACIÓN EN LA RESOLUCIÓN DE PROBLEMAS

La evaluación es la fase final del proceso SARA de resolución de problemas.² Es un intento sistemático de determinar si un problema disminuyó después de que se hiciera un esfuerzo por reducirlo. Aunque la evaluación es la fase final del proceso de resolución de problemas, a lo largo de todo el proceso SARA se tomarán decisiones críticas sobre la evaluación. La parte izquierda de la Figura 2 muestra el proceso SARA estándar y algunas de las preguntas más básicas que se plantean en cada etapa. También llama la atención sobre el hecho de que la etapa de evaluación puede producir información que obligue al encargado de la resolución de problemas a volver a etapas anteriores para hacer modificaciones. Esto ocurre sobre todo si la respuesta no ha tenido el éxito esperado.

En la parte derecha de la Figura 2 se muestran las cuestiones críticas que hay que abordar para llevar a cabo una evaluación. Durante la fase de identificación, debe definir el problema con suficiente precisión como para poder medirlo. Aquí es donde se determina "cómo se ve el éxito". ¿Cuál es la reducción mínima necesaria del problema que resulta aceptable? En esta fase, recopilará datos de referencia sobre la naturaleza y el alcance del problema. En la fase de análisis, se recopilarán datos que describan los pormenores del problema: quién se ve afectado, cuándo, dónde y en qué medida. Prácticamente todas las cuestiones importantes que deben abordarse durante el análisis serán relevantes en la fase de evaluación.

FIGURA 2: RESOLUCIÓN DE PROBLEMAS Y PLANIFICACIÓN DE LA EVALUACIÓN

EL PROCESO SARA	PREGUNTAS DE EVALUACIÓN
EXPLORACIÓN ¿Cuál es el problema?	¿Cómo debe medirse el problema? ¿Qué tendría que disminuir para que se viera el éxito?
ANÁLISIS ¿Cuánto problema hay? ¿Quién está involucrado y cómo? ¿Dónde está el problema y por qué?	¿“Cuánto” problema hay? ¿Quién está involucrado y cómo? ¿Dónde está el problema y por qué?
RESPUESTA ¿Qué hay que hacer para resolver el problema? ¿Quién debe hacerlo y cómo? ¿Se está haciendo?	¿Cómo se determinará la rendición de cuentas? ¿Cómo se medirá la reducción del problema? ¿Cómo se medirán el desplazamiento y la difusión?
EVALUACIÓN ¿Se produjo la respuesta según lo previsto? ¿Existe menos el problema? ¿Qué debería hacerse a continuación?	¿Se implementó la respuesta (evaluación de proceso)? ¿Ha cambiado el problema? ¿Se pueden eliminar explicaciones alternativas para los cambios?

Esto se debe a que durante la evaluación se desea saber si el problema ha cambiado: la información descubierta durante la fase de análisis se convierte en información de referencia vital (línea base o mediciones "pre-respuesta") para la fase de evaluación. Durante la fase de respuesta, a la vez que se desarrolla una estrategia para reducir el problema, también hay que desarrollar un mecanismo de rendición de cuentas para asegurarse de que los distintos participantes en la respuesta llevan a cabo lo que deberían estar haciendo. Como veremos más adelante, un tipo de evaluación —la evaluación de procesos— está estrechamente relacionada con la rendición de cuentas. Además, el tipo de respuesta que se utilice influirá mucho en cómo se diseñe el otro tipo de evaluación: la evaluación de impacto.

Todas estas decisiones previas se reúnen durante la fase de evaluación para responder a las preguntas: ¿La respuesta se implementó según lo previsto? ¿Cambió (disminuyó) el problema? ¿Hay buenas razones para creer que la respuesta es la explicación más importante de los cambios en el problema?

En resumen, se empieza a planificar una evaluación cuando se aborda un problema; la evaluación se desarrolla a lo largo del proceso SARA, culmina durante la fase de evaluación y proporciona conclusiones que ayudan a determinar si se debe volver atrás y revisar las fases anteriores para mejorar la respuesta. El Apéndice C contiene una lista de verificación que puede utilizarse como guía general para la evaluación a lo largo del proceso SARA.

05 TIPOS DE EVALUACIONES

Como ya hemos dicho, hay dos tipos de evaluación: la evaluación de proceso y la evaluación del impacto. Se complementan mutuamente.

EVALUACIONES DE PROCESO

En una evaluación de proceso se pregunta: ¿Se implementó la respuesta según lo previsto? ¿Funcionaron todos los componentes de la respuesta? O, dicho más claramente, ¿se hizo lo que se dijo que se haría? Se trata de una cuestión de rendición de cuentas. Empecemos con un ejemplo hipotético. Aunque ficticio, este ejemplo se basa en un esfuerzo real contra la prostitución en Londres (Matthews, 1992). Regresaremos a este ejemplo repetidamente en esta guía para ilustrar numerosos puntos.

Tras un minucioso análisis, un equipo de resolución de problemas determina que, para controlar un problema de prostitución callejera, pedirá al departamento de ingeniería de tráfico de la ciudad que convierta en unidireccional una vía principal y cree varias calles sin salida para impedir el paso de los "clientes". Esto se llevará a cabo inmediatamente después de una amplio operativo contra las trabajadoras sexuales de la zona. Las trabajadoras sexuales detenidas, si son condenadas, quedarán en libertad condicional, con la condición de que no puedan estar en la zona objetivo durante un año. Por último, una organización sin ánimo de lucro ayudará a las mujeres que quieran abandonar el trabajo sexual a adquirir las capacidades necesarias para conseguir un empleo legítimo. El escuadrón antivicio^c, los patrulleros del distrito, el fiscal, los jueces locales, la oficina de libertad condicional, el departamento del jefe de policía, el departamento de ingeniería de tráfico y la organización sin ánimo de lucro aceptaron este plan.

En una evaluación de proceso se examinaría si se aplicó la ofensiva, cuántas detenciones se efectuaron durante la implementación, si se alteraron los patrones callejeros según

^c En contextos anglosajones un escuadrón antivicio es una división policial que se enfoca en los delitos definidos en las leyes como juegos de azar, narcóticos, pornografía, comercio sexual y venta ilegal de alcohol

lo previsto, cuántas trabajadoras sexuales pidieron ayuda para adquirir nuevas capacidades laborales y cuántas trabajadoras sexuales pudieron encontrar un empleo legítimo. La evaluación de proceso también examinaría si todo esto ocurrió en la secuencia planeada. La evaluación de proceso reveló lo siguiente: La ofensiva no ocurrió hasta después de las reformas de las calles. Sólo se detuvo a una parte de las trabajadoras sexuales que operaban en la zona, y ninguna de ellas buscó obtener capacidades laborales. Basándose en esto, se podría sospechar que el plan no se llevó a cabo en su totalidad o no se realizó en la secuencia especificada. Se podría concluir que la respuesta fue un fracaso colosal. Sin embargo, el hecho es que esta evaluación no nos da ninguna prueba de éxito o fracaso, porque una evaluación de procesos sólo responde a la pregunta: "¿Qué medidas se tomaron?". Esta evaluación no responde a la pregunta: "¿Qué pasó con el problema?".

EVALUACIONES DE IMPACTO

Para determinar qué ha pasado con el problema, se necesita una evaluación de impacto. Una evaluación de impacto plantea las siguientes preguntas: ¿Ha disminuido sustancialmente el problema? En caso afirmativo, ¿fue la respuesta lo que provocó esta disminución? Siguiendo con el ejemplo anterior, veamos cómo podría funcionar. Durante la fase de análisis del proceso de resolución del problema, los agentes de patrulla y los detectives de antivicio llevaron a cabo un censo de las trabajadoras sexuales que operaban en la zona objetivo. También pidieron al departamento de ingeniería de tráfico que instalara contadores de tráfico en la vía principal y en las calles laterales críticas para medir el flujo de tráfico. Estos se utilizaron para determinar cómo se movían los clientes por la zona. La brigada antivicio realizó grabaciones de video encubiertas de la zona objetivo para documentar los métodos con los que las trabajadoras sexuales interactuaban con los clientes potenciales. Todo esto se hizo antes de seleccionar una respuesta; la información obtenida ayudó al equipo a crearla.

Una vez aplicada la respuesta (aunque no la prevista, como hemos visto), el equipo repitió estas medidas. Descubrieron que, en lugar de las veintitrés trabajadoras sexuales contabilizadas en el primer censo, sólo se podían encontrar diez. También descubrieron que había un ligero descenso del tráfico en la vía principal los viernes y sábados por la noche, pero no a otras horas. Sin embargo, el tráfico en las calles laterales disminuía considerablemente los viernes y sábados por la noche. Nuevas grabaciones de video encubiertas mostraron que las trabajadoras sexuales de la zona habían modificado su forma de acercarse a los vehículos y actuaban con más precaución. En resumen, el equipo tenía evidencia de que el problema había disminuido con respecto a antes de la implementación de la respuesta.

Entonces, ¿cuál ha sido la causa de la disminución del problema? Esta pregunta puede no ser tan importante como parece a primera vista.

Al fin y al cabo, si el objetivo era reducir o eliminar el problema y se ha conseguido, ¿qué importa cuál haya sido la causa? No importa, a menos que le interese utilizar la misma forma de respuesta en situaciones similares en el futuro. Si no tiene interés en volver a utilizar la respuesta, lo único que importa es que se ha logrado el objetivo. Entonces, los recursos dedicados a abordar el problema pueden emplearse en una preocupación más apremiante. Pero si cree que la respuesta puede volver a utilizarse, es muy importante determinar si la respuesta fue responsable de la disminución del problema.

Supongamos que el equipo de resolución de problemas de prostitución creyó que la respuesta podría ser útil para abordar problemas similares. La respuesta, aunque no se aplicara según lo previsto, podría haber causado la disminución, pero también era posible que algo más lo causara. Hay dos razones por las que el equipo se tomó en serio esta segunda posibilidad. En primer lugar, la respuesta real se apartó de la respuesta prevista, la cual se había diseñado para adaptarse al problema. Pero la naturaleza imprevista de la respuesta real hace que la explicación para la disminución sea menos plausible. En segundo lugar, la evaluación de impacto no fue especialmente sólida. Más adelante, analizaremos por qué fue una evaluación deficiente y qué puede hacerse para reforzarla.

INTERPRETACIÓN DE LAS EVALUACIONES DE PROCESO Y DE IMPACTO

Las evaluaciones de proceso y de impacto responden a preguntas diferentes y sus resultados combinados suelen ser muy informativos. La Tabla I resume la información que puede extraerse de ambas evaluaciones. Como se verá en el Apéndice B, la interpretación de esta tabla depende del tipo de diseño utilizado para la evaluación de impacto.

Cuando una respuesta se aplica según lo previsto (o casi), las conclusiones son mucho más fáciles de interpretar (casillas A y B de la Tabla I). Cuando la respuesta no se ejecuta según lo previsto, tenemos más dificultades para determinar qué ha pasado y qué hacer a continuación (casillas C y D). La casilla D es especialmente problemática porque todo lo que realmente se sabe es que "no lo hicimos y no funcionó". ¿Debería intentar implementar su plan original o debería empezar de cero?

Los resultados de la casilla C merecen una discusión más profunda. La disminución del problema significa que podría poner fin a este proceso de resolución de problemas y pasar a algo diferente. Si el problema ha disminuido sustancialmente, esto podría ser satisfactorio.

TABLA I: INTERPRETACIÓN DE LOS RESULTADOS DE LAS EVALUACIONES DEL PROCESO Y DE IMPACTO

RESULTADOS DE LA EVALUACIÓN DE PROCESO			
RESULTADOS DE LA EVALUACIÓN DE IMPACTO		Se implementó casi según lo previsto	No se implementó o se implementó de forma radicalmente distinta a la prevista
	El problema disminuyó	A. Hay pruebas de que la respuesta causó una disminución del problema	C. Sugiere que otros factores podrían haber causado la disminución del problema, o que la respuesta fue accidentalmente eficaz
	El problema no disminuyó	B. Hay pruebas de que la respuesta fue ineficaz y que debería intentarse una respuesta diferente	D. Se aprendió poco; tal vez se habrían observado mejores resultados si la respuesta se hubiera implementado según lo previsto, pero esto es especulativo

06 REALIZACIÓN DE EVALUACIONES DE IMPACTO

Una evaluación de impacto consta de dos partes. La primera consiste en medir el problema: ¿cuál es su magnitud? La segunda consiste en las formas de comparar sistemáticamente los cambios en el problema para descubrir si se redujo después de la respuesta o si se redujo más que otros problemas similares, pero no tratados.

La segunda parte se denomina diseño de la evaluación. Los diseños de evaluación se crean para aportar el máximo de pruebas de que la respuesta aplicada fue la causa principal del cambio en la medida. Los diseños endebles ofrecen poca confianza en que la respuesta haya causado el cambio. Los diseños sólidos ofrecen mucha más confianza en la conclusión de que la respuesta fue la causa de la desaparición del problema.

MEDICIONES

Las evaluaciones de impacto requieren mediciones del problema antes y después de implementar la respuesta. (En el Apéndice B se describe un mal diseño comúnmente utilizado que no tiene una medición del antes). Las decisiones sobre cómo medir el problema deben comenzar en la fase de selección y definición e y estar resueltas en el momento en que se haya completado el análisis del problema. Esto permitirá que la información recopilada durante la etapa de análisis se utilice para describir cómo era el problema antes de la respuesta.

En la fase de evaluación, las medidas se toman después de aplicar la respuesta. El problema se mide de la misma manera antes y después de la respuesta.

Mediciones cuantitativas

Las mediciones pueden ser cualitativas o cuantitativas. Las mediciones cuantitativas implican números. El número de robos en un complejo de apartamentos es una medición cuantitativa. Se pueden contar antes y después de la respuesta y calcular la diferencia. Las mediciones cuantitativas permiten utilizar las matemáticas para estimar el impacto de la respuesta. Por ejemplo, los robos a casa de habitación se redujeron un 10% después de la implementación de las respuestas.

En el ejemplo anterior, tanto el recuento de las personas activas que ofrecen servicios sexuales como las cifras de volumen de tráfico son mediciones cuantitativas.

Mediciones cualitativas

Las mediciones cualitativas permiten hacer comparaciones, pero no se les pueden aplicar las matemáticas. En el ejemplo, la observación de cómo interactúan las personas que ofrecen servicios sexuales con los clientes es una medición cualitativa. Aunque la mayoría de las evaluaciones utilizan mediciones cuantitativas, las mediciones cualitativas pueden ser extremadamente útiles. El hecho de que las mediciones cualitativas no se puedan sumar, restar, multiplicar o dividir no significa que sean inútiles. Lo importante es que estas mediciones se recojan sistemáticamente, y antes y después de la intervención, para que sean comparables. Las fotos de la limpieza de una zona antes y después de un esfuerzo de resolución de problemas pueden ser útiles si se toman en los mismos lugares en las mismas condiciones de iluminación, desde el mismo ángulo y desde las mismas distancias. Un conjunto arbitrario de instantáneas de antes y después de la respuesta tiene poco valor para evaluarlo.

Mapas

Los mapas ofrecen otro método de medición cualitativa. Los mapas son muy útiles para mostrar patrones de delincuencia y desorden. Aunque el número de delitos es una medición cuantitativa y el tamaño y la forma de los patrones delictivos suelen dibujarse mediante un algoritmo informático, cuando comparamos patrones cartográficos se suelen utilizar comparaciones cualitativas.

Validez de medición

Tanto para las mediciones cualitativas como para las cuantitativas, hay que asegurarse de que éstas registran el problema y no otra cosa. Por ejemplo, el recuento de detenciones por drogas suele ser una mejor medida de la actividad policial que los cambios en el problema de las drogas. Sólo se deben utilizar los datos de detenciones como medición del problema si se puede estar seguro de que los esfuerzos y las técnicas policiales se han mantenido constantes. Por el contrario, la vigilancia encubierta sistemática de un área crítica de narcotráfico antes y después de la respuesta podría ser una medición válida, si la forma de vigilancia no cambia y sigue sin ser detectada por los narcotraficantes. Las mediciones rara vez son válidas o inválidas, sino más —o menos— válidas que las mediciones alternativas. En resumen, querrá asegurarse de que el cambio en el problema que mide se debe a cambios en el problema y no a cambios en la forma de tomar las mediciones. Una forma de pensar en esto es compararlo con las pruebas físicas recolectadas en el lugar de un delito. La razón por la que existen protocolos estrictos para la recolección y manejo de pruebas es que

no queremos confundir las actividades del delincuente con las de los encargados de recabar las pruebas. Lo mismo ocurre en las evaluaciones.

Cuanto menos directa sea la medición, menos validez tendrá. Por ejemplo, si se quiere medir el tráfico de drogas, la vigilancia de los lugares donde se trafica con drogas proporciona observaciones directas del tráfico de drogas. Las estadísticas de arrestos son mediciones indirectas porque implican las actividades de los narcotraficantes y los clientes (los aspectos del problema que más le puede interesar), así como las decisiones por parte de los ciudadanos de reportarlo a la policía y las decisiones policiales sobre intervenir y de cómo va a hacerlo. Estas decisiones de los ciudadanos y de la policía pueden no estar siempre relacionadas con la realidad subyacente al problema. Por ejemplo, los cambios en la política de tiempo extra de la policía o la presencia de brigadas especiales antidroga pueden modificar el número de detenciones, aunque el problema de la droga permanezca constante. Por este motivo, el número de arrestos de narcotraficantes es una medición menos directa y con frecuencia deficiente de un problema de drogas.

A veces, sin embargo, es imposible obtener una medición directa del problema y hay que recurrir a una medición indirecta. En 2004, veintitrés inmigrantes chinos murieron ahogados mientras recolectaban marisco en el Reino Unido. Se emprendió un esfuerzo de resolución del problema para reducir las posibilidades de que esto volviera a ocurrir. Evaluar el éxito de la respuesta resultó difícil porque las muertes por ahogamiento eran —afortunadamente— poco frecuentes y las muertes múltiples por ahogamiento eran aún menos frecuentes. Por lo tanto, la contabilización del número de muertes por ahogamiento antes y después de la intervención sobrestimaría el éxito del proyecto, ya que se había producido un número inusualmente alto de muertes por ahogamiento en el único incidente anterior a la intervención y, aunque la policía no hiciera nada, probablemente se produciría un número muy bajo de muertes por ahogamiento en el futuro. Los evaluadores policiales, en cambio, contabilizaron las llamadas de rescate al servicio de salvamento costero. Las pruebas mostraron que estas llamadas disminuyeron sustancialmente, lo que demuestra que la respuesta fue un éxito.³

Volvamos al problema de la prostitución para ver otro ejemplo de medición indirecta y directa. En este ejemplo, el significado de *directo* e *indirecto* depende de cómo se defina el problema. Los hombres conducen dentro de un barrio los viernes y los sábados por la noche en busca de trabajadoras sexuales. Esto molesta a los vecinos. Llamen a la policía para que haga algo. Se puede elegir entre dos mediciones para este problema. La primera es una medición cuantitativa tomada a partir de contadores automáticos de tráfico colocados estratégicamente en las calles críticas tres meses antes de la intervención y dejados allí hasta tres meses después de finalizada la respuesta. Estos dispositivos miden el flujo de tráfico.

La diferencia entre el volumen medio de tráfico de los viernes y sábados por la noche y el volumen medio durante el resto de la semana se utiliza como estimación del tráfico derivado de la prostitución.

Su segunda medición se basa en entrevistas a residentes locales realizadas tres meses antes de la respuesta y tres meses después. Se pregunta a los residentes sobre su percepción del problema de la prostitución utilizando una escala numérica (0 = ninguna, 1 = menor, 2 = moderada, 3 = fuerte).

Si ha definido su problema como tráfico relacionado con el trabajo sexual, la primera medición es más directa que la segunda. No toda la diferencia entre el nivel de tráfico los viernes y sábados y el nivel durante el resto de la semana se debe a esta actividad, pero una gran parte probablemente sí. Se trata, pues, de un enfoque razonable. Preguntar a los ciudadanos por sus percepciones, sin embargo, está plagado de dificultades. Su percepción actual de la prostitución puede estar influida por observaciones pasadas. Es posible que no vean mucho del tráfico en razón del trabajo sexual, en particular, si están resguardados en casa para evitar el problema. Pueden percibir erróneamente otras actividades como relacionadas con la prostitución. Si, por el contrario, ha definido el problema como la molestia de los residentes por el tráfico relacionado a esta actividad, las entrevistas son una medición más directa que los recuentos de tráfico. Puede ser que el tráfico relacionado con la oferta de servicios sexuales no haya cambiado, pero los ciudadanos piensan que es así. Según esta medición, la respuesta ha sido un éxito. Pero si el tráfico relacionado con la prostitución (medido por los contadores) ha disminuido precipitadamente y los ciudadanos no son conscientes de ello, entonces, según esta medición, la respuesta no ha funcionado.

Por supuesto, se pueden utilizar múltiples mediciones. En este ejemplo, se podría medir tanto la reducción del tráfico relacionado con la prostitución como la percepción de este. Sólo si ambos disminuyeran, el éxito sería inequívoco. Si los contadores indicaran un descenso del tráfico, pero las encuestas a los ciudadanos mostraran que los residentes no eran conscientes del descenso, la respuesta podría modificarse para tomar en cuenta las percepciones.

Selección de mediciones válidas

¿Cómo seleccionar mediciones específicas para su problema? No existe una respuesta única a esta pregunta que pueda aplicarse a cualquier tarea de resolución de problemas. Si está trabajando en un problema para el cual se ha elaborado una guía específica para resolverlo, en ella encontrará algunas ideas de mediciones específicas para abordarlo. Si está trabajando en otro tipo de problema, el enfoque más sencillo es utilizar uno o varios de los indicadores que se utilizaron para identificar y analizar el problema. Sin embargo, es importante reflexionar detenidamente sobre la definición del problema.

Como vimos en el ejemplo del trabajo sexual, cambios aparentemente menores en cómo se define el problema pueden tener implicaciones significativas para la medición. Lo que sí está claro es que hay que pensar en las mediciones de evaluación al momento en que se inicia el ciclo SARA.

Una forma de aclarar las mediciones que deben utilizarse es plantear la pregunta: "¿Por qué a nosotros, la policía, nos preocupa este problema?". Las respuestas conducen a las mediciones de resultado. Entre otras razones, la policía se preocupa porque: (1) los ciudadanos están molestos (o asustados); (2) la gente está resultando herida; (3) le está costando demasiado dinero a la ciudad, o (4) le está haciendo perder tiempo a la policía. Tenga en cuenta que estas respuestas de ejemplo no son genéricamente válidas. Una actividad de prostitución en una zona industrial-almacén debería producir respuestas diferentes que la misma actividad en una zona residencial. Tenga en cuenta también que "va contra la ley" no es una respuesta válida. La ley es una herramienta que ayuda a reducir los problemas, por lo que su cumplimiento no es un resultado. El resultado es la reducción del problema.

CRITERIOS PARA DETERMINAR LA CAUSALIDAD

Como ya hemos comentado, una evaluación de la resolución de problemas tiene dos objetivos: determinar si el problema ha cambiado y determinar si la respuesta ha provocado el cambio en el problema. Nos interesa especialmente el primer objetivo. El segundo objetivo sólo es importante si (1) el problema ha cambiado y (2) una respuesta similar puede utilizarse para abordar otros problemas. Si no se cumple ninguna de estas dos condiciones, no hay que preocuparse por la causa del cambio y la evaluación es relativamente sencilla.

Sin embargo, si el problema ha cambiado y es probable que la respuesta se utilice de nuevo, es importante determinar si la respuesta fue de hecho la causa del cambio. Si el problema ha disminuido por motivos ajenos a la respuesta, es poco probable que, al volver a utilizar la respuesta en circunstancias similares, produzca resultados útiles.

¿Qué tal si el problema ha empeorado tras la respuesta? La respuesta puede o no ser responsable de ello. Si puede determinar que el cambio en el problema no se debió a la respuesta, esta podría ser útil para otros problemas. Si la respuesta provocó el aumento del problema, está claro que no quiere volver a utilizarla y debería advertir a los demás de que no lo hagan.

El concepto de causa puede parecer bastante sencillo, pero no lo es. Antes de poder proclamar con seguridad que una respuesta ha provocado la disminución del problema, hay que cumplir cuatro criterios.

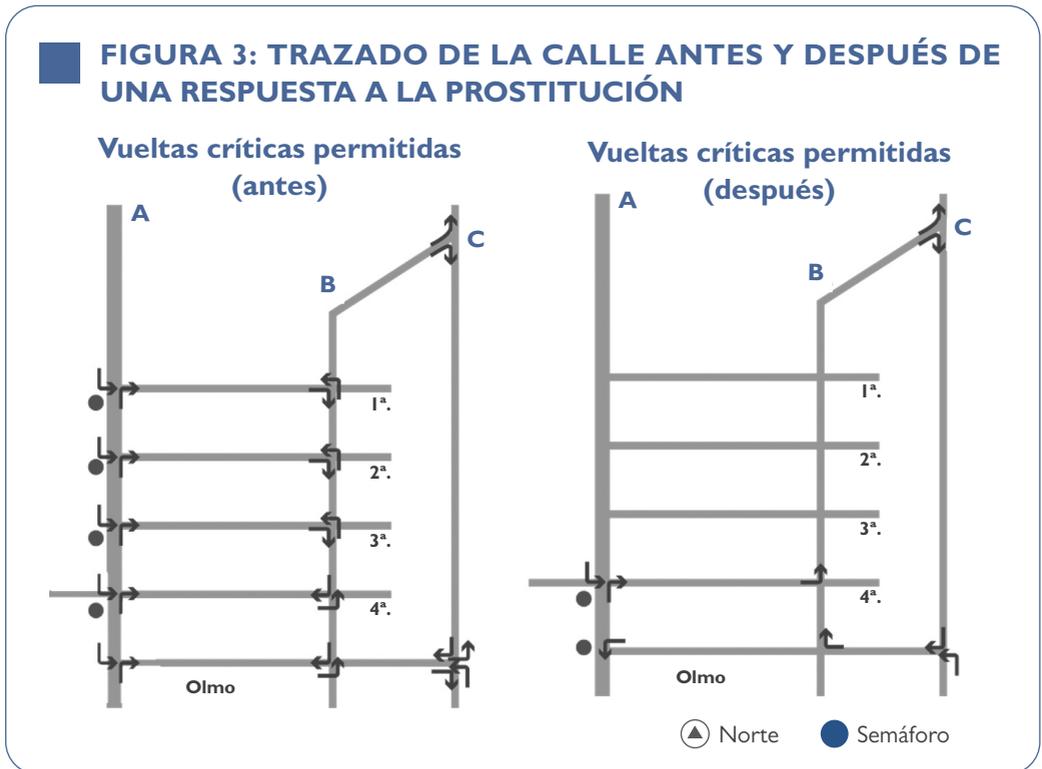
Los tres primeros son relativamente sencillos y suelen cumplirse. El cuarto criterio no puede cumplirse con absoluta certeza. Los analizamos a continuación.

Una explicación plausible de cómo la respuesta reduce el problema

El primer criterio es que debe haber un argumento convincente que demuestre cómo se supone que la respuesta aborda el problema.^d Esta explicación debe basarse en un análisis detallado del problema, preferiblemente ampliado con investigaciones y teorías previas.

El hecho de que otros hayan utilizado una respuesta similar y hayan podido reducir su problema no es una explicación. Esa información es útil, pero sigue siendo necesario explicar cómo ocurrió. En ausencia de una explicación convincente, no sabrá si esta experiencia anterior tuvo éxito por accidente, si su éxito fue exclusivo de la situación en la que se aplicó por primera vez (y no funcionará en su problema concreto) o si se trata de una respuesta útil en general.

Volviendo al ejemplo de la prostitución, podemos ilustrar lo que se entiende por una explicación plausible.



^d El término técnico para este criterio es "mecanismo".

¿Por qué creen los residentes que funcionará? Esperamos que su respuesta sea una explicación plausible, es decir, que sea lógica y que tenga en cuenta los hechos conocidos. Los residentes afirman que esta zona es un hervidero de actividad de prostitución en gran parte porque las calles facilitan el comportamiento de compra de los clientes y la exhibición de las sexo servidoras. Los clientes pueden dar una vuelta rápida a la manzana en busca de trabajadoras sexuales. Al cambiar el patrón de la calle de la forma descrita, recorrer el circuito se vuelve más lento. Si los clientes no consiguen un contacto en la primera pasada, tomarán mucho más tiempo en el viaje de vuelta. Al reducirse la comodidad de los clientes, serán menos los que acudan a la zona y se reducirá el problema. Además, una vez agilizado el flujo de tráfico, será más fácil para la policía detectar actividades relacionadas con la prostitución, lo que aumentará el riesgo de detección. Al observar a los clientes y a las trabajadoras sexuales, podremos verificar el comportamiento de los cruceros. Si esta explicación es lógicamente coherente con la información disponible y no existe información contradictoria clara y evidente, los residentes han superado el primer obstáculo para establecer una relación causal.

Muchas ideas plausibles no funcionan cuando se ponen a prueba; por lo que una explicación lógica y razonable por sí misma no garantiza que la respuesta vaya a funcionar, pero sí hace que la respuesta sea la candidata más probable para una solución exitosa que aquellas explicaciones que no se basan en la lógica, los hechos y la experiencia. La investigación previa es importante para establecer la plausibilidad. El éxito de la respuesta utilizada en el ejemplo se hace plausible por el hecho de que la investigación previa describe la relación entre la prostitución y los patrones circulares de conducción⁴ y demuestra que la reducción de la facilidad de movimiento del tráfico a través de los barrios a veces reduce el delito.⁵ Además, esta intervención es coherente con la teoría de la Prevención Situacional del Delito, en particular, con la estrategia de aumentar el esfuerzo de los delincuentes^e.⁶ Con demasiada frecuencia, la policía, los funcionarios y el público se detienen en la noción de verosimilitud y asumen que, si suena razonable, debe ser cierto. Y con la misma frecuencia, las pruebas demuestran que esta impresión inicial era errónea.

En resumen, el primer paso para demostrar que una respuesta ha reducido el problema es una explicación plausible de (1) cómo funciona el problema y (2) cómo se supone que la respuesta va a interrumpir esta operación. Esta explicación debe decir cómo, dónde, cuándo y por qué funciona la respuesta. Si tal explicación se prepara cuando se está elaborando la respuesta, puede ayudar a orientar la planificación y ejecución de esta. Entre más específica sea esta explicación, mejor será la respuesta y más informativa será la evaluación.

^e Nota de traducción: Es importante resaltar que en Estados Unidos la prostitución es un delito, por ello cabe lugar una estrategia para "aumentar el esfuerzo de los delincuentes", que en este caso pudiera ser tanto los usuarios como las trabajadoras del sexo. En México, este tipo de estrategia no tendría sentido, porque la prostitución o trabajo sexual es legal, siempre y cuando sea ejercida por una persona mayor de edad y por voluntad propia.

Lo ideal sería que esta explicación también describiera las circunstancias bajo las cuales es poco probable que la respuesta funcione. Esto puede ayudar tanto en la evaluación de proceso como en la del impacto.

La magnitud del problema y el nivel de la respuesta están relacionados

El segundo criterio para afirmar que una respuesta ha provocado una disminución del problema es que exista una relación entre la presencia de la respuesta y una disminución del problema (y la ausencia de la respuesta y un aumento del problema).^f

Volvamos al problema de la prostitución. ¿Cómo demostraríamos aquí una relación? ¿Existen barrios similares que podamos comparar con los Olmos? Justo al norte de los Olmos, hay un barrio como los Olmos (también está entre las calles A y C, con una zona industrial ligera deteriorada al oeste y la próspera urbanización de la calle C al este), pero las calles no permiten manejar con facilidad de manera circular. Ahora bien, si la facilidad de la conducción circular está asociada a la prostitución, deberíamos ver poca o ninguna oferta de prostitución en este otro barrio. Esto implicaría que cambiar el patrón de las calles en los Olmos podría ser útil. Sin embargo, si también hay prostitución en esta zona, no existe una fuerte relación entre la prostitución y la facilidad para circular y esto sugiere que cambiar el patrón callejero puede no ser efectivo. En cualquier caso, las pruebas no serían sólidas, pero los resultados podrían ser útiles.

También podríamos intentar demostrar una relación midiendo el problema antes y después de los cambios en las calles. Si observamos altos niveles de prostitución (o altos niveles de percepción de la prostitución por parte de los residentes) antes de los cambios, pero bajos niveles en estas mediciones después de los cambios en las calles, tendremos pruebas de una relación.

Para superar el segundo obstáculo para identificar las causas, debemos demostrar que la situación más crítica cuando no hay una respuesta que cuando esta está en marcha.

Si es así, resulta tentador declarar la victoria en esta fase; sin embargo, hay otros dos obstáculos que deben superarse antes de que podamos estar seguros de que la solución fue responsable de la disminución del problema. Esto nos lleva al tercer criterio para demostrar una conexión causal.

La respuesta al problema es anterior a su disminución

El tercer criterio es que la disminución del problema se produzca después de la respuesta;^g lógicamente, una respuesta no tendría efecto antes de implementarse.

^f El término técnico para este criterio es "asociación". Normalmente, la asociación se mide por la correlación entre la respuesta y el nivel del problema.

^g El término técnico para este criterio es "orden temporal".

Aquí hay una advertencia importante: por respuesta, incluimos la publicidad —intencionada o accidental— de la respuesta. Una ofensiva contra los conductores ebrios puede ir precedida de una amplia campaña en los medios de comunicación; si es así, los posibles conductores ebrios pueden modificar su comportamiento incluso antes de la intervención. En este caso, la campaña mediática forma parte de la respuesta.^h Una disminución en la conducción bajo los efectos del alcohol tras el inicio de la campaña mediática, pero antes de las medidas enérgicas, podría atribuirse a la respuesta. Sin embargo, una disminución en la conducción bajo los efectos del alcohol antes de la campaña mediática sería una prueba de que algo distinto a la respuesta ha provocado la disipación del problema.

A pesar de su evidente simplicidad, es sorprendentemente frecuente observar violaciones de este criterio. A lo largo de la década de 1990, los homicidios disminuyeron en las grandes ciudades de Estados Unidos. A mediados de la década, un par de años después de iniciarse la tendencia a la baja, varias ciudades estadounidenses pusieron en marcha estrategias de reducción de delito y adquirieron gran notoriedad. A medida que los homicidios seguían disminuyendo en estas ciudades, los defensores afirmaban que estas reducciones se debían a las nuevas estrategias. De hecho, los homicidios habían estado disminuyendo antes de los cambios. Dado que la tendencia de los homicidios era descendente antes de los cambios, es difícil atribuir la disminución a los cambios en las estrategias policiales.ⁱ En resumen, la supuesta causa de la disminución se produjo después de que esta comenzara. Si estos mismos cambios se hubieran aplicado en 1990, la afirmación de que causaron el descenso de los homicidios sería más plausible.

Para demostrar que la respuesta precedió a la disminución del problema, hay que saber cuándo empezó la respuesta (incluida la publicidad sobre ella) y disponer de mediciones del problema antes de ese momento y después de ese momento. Esto se denomina diseño de evaluación antes-después (o pre-post). Es el diseño de evaluación más común, pero no es un diseño especialmente sólido. Es decir, un sencillo diseño pre-post puede mostrar una disminución, pero es insuficiente para establecer qué causó la disminución.

A pesar de su simplicidad superficial, este criterio puede ser difícil de demostrar. Pero incluso si se puede demostrar que la disminución en el problema surgió después de la respuesta, necesita cumplir un criterio más antes de poder afirmar definitivamente que la respuesta causó la disminución: debe eliminar las explicaciones alternativas.

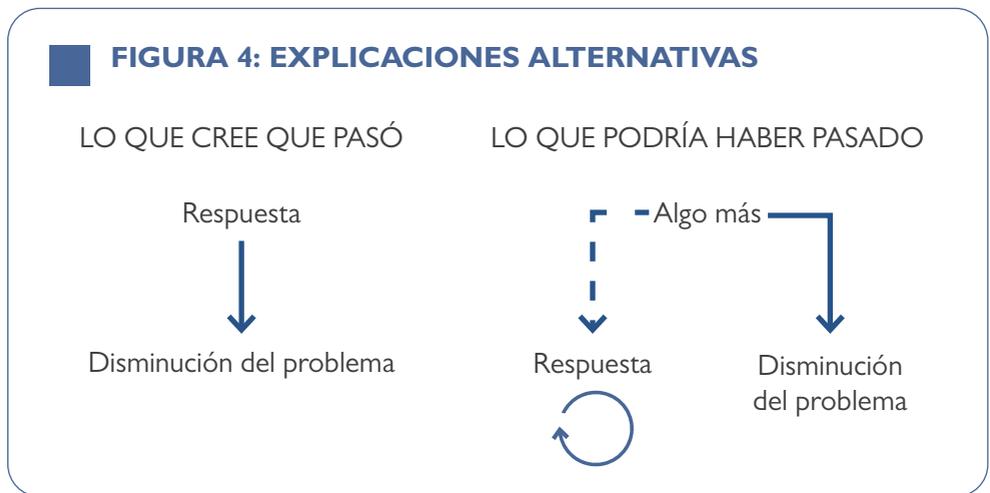
Eliminación de explicaciones alternativas

Sigamos con el problema de la prostitución. Ya tenemos una explicación, hemos demostrado que existe una relación y que la respuesta fue anterior a la disminución del problema.

^h El término técnico para este fenómeno es "beneficio anticipado".

Ahora debe asegurarse de que ninguna otra cosa ha podido causar la disminución de la prostitución.¹ Recuerde que el corredor de la calle C y los Olmos están experimentando una serie de cambios. Nuevas personas se están instalando en la zona y se están aliando con los residentes de más edad que se quedaron para limpiar la zona. Una de las cosas que hicieron fue pedir ayuda a la policía. ¿Hicieron algo más? Supongamos que la Asociación de Vecinos de los Olmos (ENA) y la Asociación de Comerciantes del Corredor de la Calle C (CSCBA) identificaron a los propietarios del inmueble abandonado y baldío y los presionaron para que lo limpiaran. Esto impidió a las trabajadoras sexuales acceder al inmueble. Y supongamos que estos cambios se pusieron en marcha más o menos al mismo tiempo que los cambios en las calles. Así que se podría pensar en la ENA y la CSCBA como la causa de los cambios en las calles y de los cambios en el uso del suelo. Si los cambios en el uso del suelo fueran la causa real de la reducción de la prostitución, y los cambios en las calles fueran irrelevantes, se seguiría observando una relación entre el cierre de calles y la reducción de la prostitución, y se seguiría observando la respuesta antes de la reducción. Sin embargo, algo más sería responsable de la disminución del problema.

La Figura 4 muestra la noción de una explicación alternativa. El panel izquierdo muestra lo que usted cree: que la respuesta (mostrada por la flecha) causó la disminución del problema. Esta creencia puede proceder de diversas fuentes válidas. Sin embargo, algo más ha causado tanto la respuesta como la reducción del problema (panel derecho). En este caso, "algo más" produjo más respuesta y, al mismo tiempo, causó una reducción del problema.



¹ El término técnico para este criterio es "carácter no espurio". Una relación espuria es una relación falsa: parece que la respuesta está causando la disminución del problema, pero en realidad algún otro factor es la causa de la disminución y posiblemente también la respuesta.

La ausencia de una flecha entre la respuesta y la disminución del problema significa que, en realidad, la respuesta era irrelevante para el problema. Una persona ajena que observara más respuesta y menos problema al mismo tiempo podría llegar a la conclusión errónea de que la respuesta y el problema están relacionados causalmente. En situaciones como esta, la relación observada entre la respuesta y la disminución del problema es engañosa. La posibilidad de una relación engañosa entre una respuesta y un problema es una amenaza para la validez de las conclusiones de una evaluación. Nótese que se trata de una posibilidad, no de una certeza demostrada. Una amenaza para la validez de las conclusiones no significa que la respuesta haya sido un fracaso. Significa que no podemos estar seguros de que la respuesta haya funcionado. Existe una duda sustancial porque hay una explicación alternativa plausible. De nuevo, un juicio con jurado es un ejemplo útil. Si el fiscal no consigue descartar a otras personas que podrían haber cometido el delito (y la defensa llama la atención del jurado sobre ello), el jurado debe tener algunas dudas sobre la culpabilidad del acusado. La absolución, en este caso, no significa que el fiscal esté equivocado. Significa que el fiscal no ha eliminado con éxito las explicaciones alternativas.

Hay una preocupación relacionada: El "algo más" podría haber ocurrido por casualidad más o menos al mismo tiempo que la respuesta. En la práctica, puede que no importe si el "algo más" se produjo accidentalmente al mismo tiempo que su respuesta o si el "algo más" causó tanto la respuesta como la disminución del problema. En ninguno de los dos casos la respuesta causó la disminución del problema.

Para demostrar una conexión causal entre el problema y la respuesta, un evaluador necesita aportar pruebas sólidas de que no hay "algo más". Para ello, el evaluador debe demostrar que no existen explicaciones razonables de la disminución del problema distintas de la respuesta. Eliminar todas las explicaciones alternativas es difícil. Nunca se puede hacer de forma definitiva, porque hay muchas causas posibles de las fluctuaciones del problema. Lo único que se puede hacer es eliminar las explicaciones alternativas conocidas más obvias para la disminución del problema. Nunca podemos demostrar que una respuesta haya causado la disminución de un problema, porque no podemos eliminar todas las posibles explicaciones alternas de la disminución. Sin embargo, podemos presentar mejores o peores argumentos para tales afirmaciones. Y aquí es donde entra en juego el diseño de la evaluación. Algunos diseños permiten afirmaciones de causalidad más sólidas que otros, del mismo modo que algunos procesos penales son más plausibles para un jurado que otros.

DISEÑOS^k

Un diseño de evaluación es una estrategia sistemática, coordinada con la respuesta para organizar cuándo y dónde tendrá lugar la recopilación de datos. Si desarrolla el diseño de la evaluación junto con la respuesta, será más probable que la evaluación produzca información útil. Si espera a que la respuesta se haya aplicado para decidir cómo se evaluará, tendrá más dificultades para determinar si la respuesta ha sido eficaz.

Existen muchos tipos de diseños de evaluación que pueden utilizarse (véanse las lecturas recomendadas). Analizaremos dos diseños prácticos comunes: el diseño pre-post (que ya hemos abordado en cierta medida anteriormente) y el diseño de series temporales. Ninguno de los dos tiene grupos de control o comparación. En el Apéndice B se tratan los diseños de grupos de comparación y los de series temporales múltiples (la fila inferior de la Tabla 2) y se describe cuándo puede querer utilizar un grupo o área de control.

Diseños pre-post

El diseño pre-post más sencillo implica una única medición del problema antes de la respuesta y una única medición después de la respuesta. La medición posterior se compara con la medida anterior. Si después de la respuesta el problema es menor que antes, esto demuestra que el problema ha disminuido. Como veremos, este diseño a veces es adecuado para determinar si el problema disminuyó, pero es insuficiente para determinar que la respuesta causó la disminución.

TABLA 2: TIPOS DE DISEÑOS DE EVALUACIÓN

	Medición única (antes y después)	Mediciones múltiples (antes y después)
Grupo sin comparación (control)	Diseño pre-post	Diseño de series de tiempo
Grupo de comparación (control)	Pre-post diseño con grupo de control	Diseño de series temporales múltiples

^k Los diseños no experimentales no se abordan en esta monografía porque a menudo no pueden demostrar que la respuesta se produjera antes de la disminución del problema y porque son especialmente deficientes para eliminar explicaciones alternativas de la disminución. Tampoco se abordan los diseños aleatorios. Aunque son útiles para estudiar intervenciones genéricas para su aplicación en una clase de problemas, generalmente no son adecuados para la resolución de problemas operativos en los que el interés principal es la reducción de un problema específico en lugar de probar una solución genérica. Puede encontrar información sobre estos y otros tipos de diseños no descritos en esta guía en la lista de lecturas recomendadas.

La Figura 5 muestra los resultados de un diseño pre-post. El Departamento de Policía de Houston, en colaboración con la autoridad de salud mental del condado, aplicó un enfoque para mejorar el trato a los enfermos mentales. El proyecto de resolución de problemas fue finalista del Premio Herman Goldstein 2010 a la Excelencia en la Actuación Policial Orientada a la Resolución de Problemas. En el marco de la evaluación de la intervención, el Departamento de Policía de Houston examinó si el número de sucesos relacionados con la salud mental había disminuido del año anterior al año posterior a la intervención. El par de barras marcado como "total" muestra que hubo una reducción sustancial: alrededor del 47%. La policía también examinó los tipos de sucesos, los cuales se muestran en los otros tres pares de barras, y observó que se habían producido disminuciones notables en las tres categorías.¹

FIGURA 5: EJEMPLO DE MEDICIÓN DEL IMPACTO EN UN DISEÑO PRE-POST

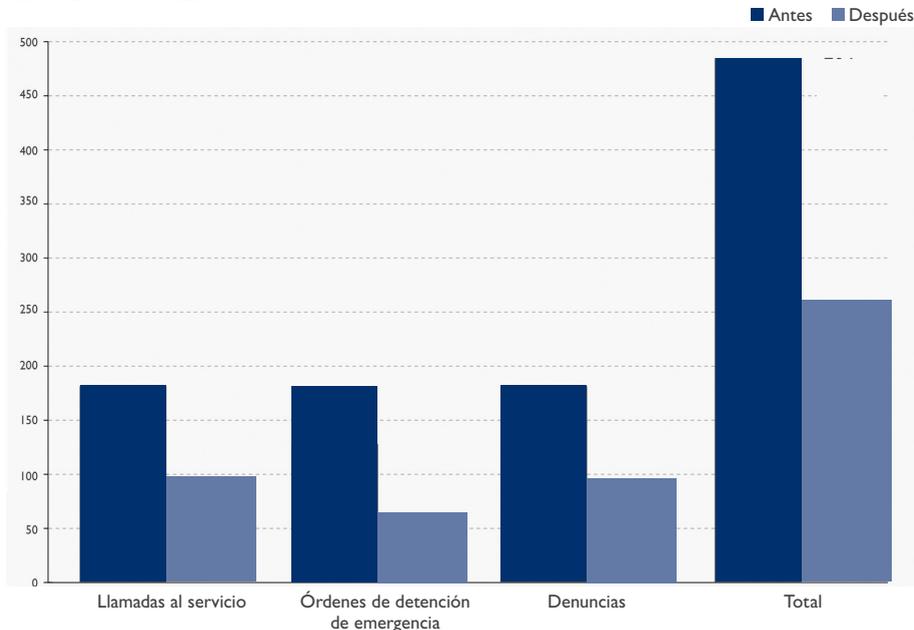


Gráfico derivado del Departamento de Policía de Houston y Autoridad de Salud Mental/Retraso Mental del Condado de Harris (2011). Chronic Consumer Stabilization Initiative: A Multi-Agency Collaboration [Iniciativa de Estabilización de Consumidores Crónicos: Una Colaboración de Múltiples Agencias]. Presentación en la Problem-Oriented Policing Conference, 2010 Award Finalist. Miami, Florida. Octubre. Accesible en <http://www.popcenter.org/conference/conferencepapers/2011/WebLeeChronicConsumers.pdf>.

¹ En la mayoría de las investigaciones de evaluación, se utiliza una prueba de significación estadística para determinar si la diferencia entre la respuesta previa y la respuesta posterior se debe probablemente al azar. En otras palabras, una explicación alternativa es que las fluctuaciones aleatorias normales en el nivel del problema causaron la diferencia entre las mediciones del problema antes y después de la respuesta. Las pruebas de significación estadística son más útiles cuando las diferencias son pequeñas pero significativas y el número de sucesos problemáticos anteriores a la respuesta es pequeño. En tales circunstancias, las fluctuaciones aleatorias normales del problema son una causa potencial del cambio. Debido a la naturaleza altamente técnica de las pruebas de significación, no se tratarán en esta monografía. Los lectores interesados en las pruebas de significación pueden encontrar explicaciones en la mayoría de los textos introductorios de estadística, en la documentación que acompaña a los programas informáticos de estadística y entre los estadísticos y los científicos sociales de las universidades locales.

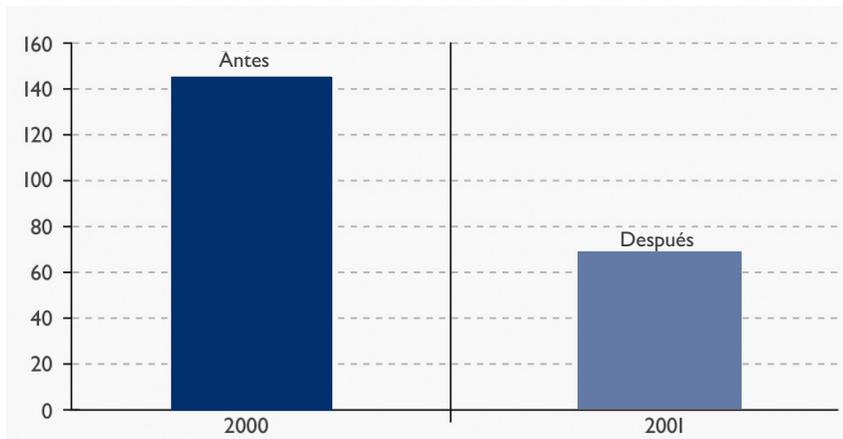
Como hemos visto, este diseño es débil a la hora de eliminar explicaciones alternativas para la disminución del problema. Esto se debe a que “algo más” podría haber causado tanto la respuesta como la disminución del problema, o porque otras cosas, las que ocurren en el momento de la respuesta, pueden ser responsables del cambio en el problema.

Para ver por qué un diseño pre-post es deficiente, considere el ejemplo de la Figura 6. Los datos de este ejemplo proceden de un informe sobre una iniciativa de resolución de problemas de robos en vehículos. En el gráfico superior de la Figura 6 vemos una simple comparación pre-post. La pregunta que se plantea es si la instalación de cámaras de videovigilancia en la zona objetivo ha provocado una reducción de los robos de vehículos. La respuesta parece ser “sí”. En el gráfico inferior, vemos otros dos años de datos sobre robos. Dos cosas son evidentes: la caída en picado de las denuncias de robos en vehículos comienza un año antes de que se instalara la videovigilancia. Esto pone en duda la validez de la conclusión de que la videovigilancia causó el descenso. Dado que los diseños pre-post no examinan tendencias a largo plazo, no pueden eliminar la explicación alternativa de que ya se estaba produciendo una disminución del problema antes de la intervención.

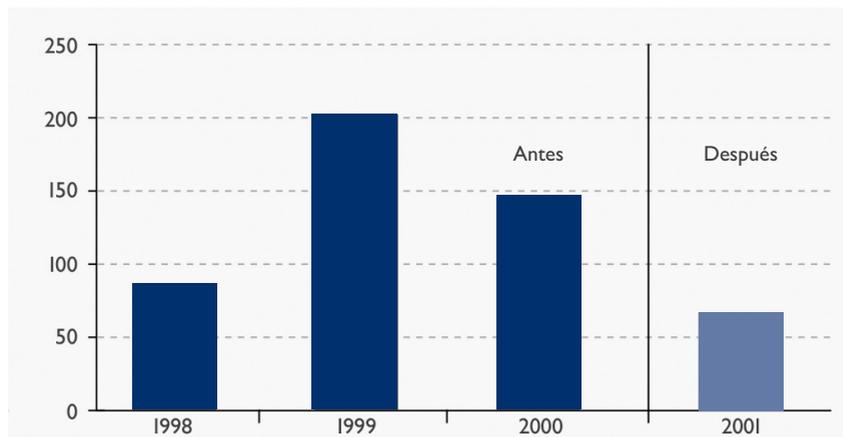
Aunque el diseño pre-post es popular, el ejemplo de la Figura 6 ilustra sus puntos débiles. Un repaso a los cuatro criterios de causalidad lo deja claro. En cuanto al primer criterio (que debemos ver una explicación plausible de cómo la respuesta podría reducir el problema), este diseño sencillo no es ni mejor ni peor que otros. En cuanto al segundo criterio (que debemos ver una relación entre la respuesta y la disminución del problema), no le va tan bien.

Si comparamos los dos paneles de la Figura 6, nuestra confianza en que existe una relación entre la respuesta de la videovigilancia y los robos en vehículos disminuye cuando se añaden otros dos periodos de tiempo.

FIGURA 6: PROBLEMAS CON UN DISEÑO PRE-POST



Este ejemplo procede de la descripción de un proyecto de resolución de problemas a largo plazo para reducir los robos en vehículos. Los autores señalan que la disminución de más del 50% en los robos de vehículos de 2000 a 2001 se debe posiblemente a la instalación de videovigilancia en la zona objetivo a principios de 2001.



Afortunadamente, los autores muestran datos de robos en vehículos de dos años antes. Esto ilustra uno de los obstáculos de un diseño pre-post: no muestra la tendencia. Los robos en vehículos alcanzaron su punto máximo en 1999 y disminuyeron un 28% en 2000. Así que es posible que el disminución de 2000 a 2001 se deba en parte o totalmente a algo que ocurrió mucho antes de que se instalara la videovigilancia.

Estos gráficos se crearon a partir de datos extraídos de la Tabla 5 (página 30) de Clarke, R.V., & Goldstein, H. (2003). *Theft From Cars in Center City Parking Facilities -A Case Study*. Washington D.C.: Office of Community Oriented Policing Services, U.S. Department of Justice. Esta tabla trataba de una pequeña faceta de un esfuerzo mucho mayor por analizar un problema.

Aunque estos robos disminuyeron tras la instalación de la videovigilancia, vemos que el número de robos subía y bajaba antes de la instalación de la videovigilancia. Los problemas suelen fluctuar, aunque no se haga nada al respecto. Esto significa que a los picos les siguen los descensos, seguidos de picos.

Por consiguiente, cualquier medida que se adopte en un periodo de máxima actividad tendrá prácticamente garantizado un buen resultado, ya que la trayectoria más probable del problema después de un pico es que disminuya.^m Este gráfico plantea la preocupación de que esto es lo que podría haber estado ocurriendo aquí. No podemos estar seguros sin más datos.

Los años de datos añadidos también sugieren que se ha incumplido el tercer criterio (que debemos estar seguros de que el problema no disminuyó hasta después de que se aplicara la respuesta): los robos empezaron a disminuir un año antes de que se instalara la videovigilancia. No disponemos de tecnología que nos permita retroceder en el tiempo, por tanto, siempre que veamos una tendencia a la baja que comienza antes de la respuesta, debemos sospechar que la respuesta tuvo poco o nada que ver con la disminución.

Por último, tampoco se ha cumplido el último criterio (que tenemos que estar seguros de que ninguna otra cosa podría haber causado la disminución del problema). Basándonos en los datos de la Figura 6, podemos imaginar al menos tres explicaciones alternativas plausibles: (1) que los robos suben y bajan aleatoriamente y que la videovigilancia se introdujo cuando los robos estaban disminuyendo; (2) que el 1999 fue un año anormalmente alto para los robos en vehículos, y que estos delitos simplemente disminuyeron a su nivel natural; y (3) que algún otro cambio en la ciudad entre 1999 y 2000 provocó la disminución (por ejemplo, una campaña intensiva de información para advertir a los conductores de que sacaran los objetos de los compartimentos de pasajeros de sus vehículos).

Los diseños pre-post también son difíciles de interpretar cuando los resultados no indican ningún cambio. Sin conocer la tendencia a largo plazo, no sabemos si el problema tendía a aumentar antes de la respuesta. Si era así y si el problema dejó de empeorar después de la respuesta, entonces, la respuesta tuvo éxito en evitar este aumento. En este caso, el diseño pre-post da la falsa impresión de que la respuesta fue ineficaz. Una última dificultad de un diseño pre-post es que no sabemos si la disminución del problema se mantiene. Imaginemos que disponemos de datos sobre robos en vehículos correspondientes a 2002. Si estos datos mostraran que en 2002 se produjeron tantos robos como en 1999 o en el año 2000, no estaríamos seguros de que la instalación de videovigilancia hubiera marcado la diferencia.

^m El término técnico para este "proceso automático" es "regresión a la media".

Si los datos mostraran niveles de robos no superiores a los de 2001, estaríamos más seguros. Entre más tiempo se mantenga la reducción después de la respuesta, más confianza tendremos en que la respuesta está funcionando bien y que los resultados del "después" no son fruto de la casualidad. No es infrecuente que los programas tengan éxito durante un breve periodo de tiempo y que luego el problema vuelva a surgir después de que la atención se desvíe hacia otros asuntos.

El ejemplo de Houston de la Figura 5 es notable porque el evaluador utilizó múltiples mediciones del problema. La coherencia del descenso del problema tras la respuesta, a través de varias mediciones diferentes, da mayor validez a las conclusiones. Aunque sigue siendo posible que algo distinto a la respuesta haya provocado las disminuciones, es menos probable que la disminución se deba a fluctuaciones aleatorias: no esperaríamos que todas las mediciones mostraran el mismo cambio si la causa fuera aleatoria.

Hemos ilustrado cómo funciona el diseño común pre-post y hemos descrito cuatro problemas que plantea la interpretación de los resultados de este tipo de diseños. Las cuatro preocupaciones se derivan del desconocimiento de la tendencia a largo plazo. A continuación, examinaremos los diseños que pueden subsanar estas preocupaciones.

Diseños de series de tiempo

El diseño de series temporales es muy superior al diseño pre-post porque puede abordar estas cuatro preocupaciones: existe una explicación plausible; la respuesta está asociada a una reducción del resultado; la respuesta es anterior al resultado, y se han eliminado las explicaciones alternativas más plausibles. Con este diseño, primero se toman muchas mediciones del problema antes de la respuesta. Esto permite observar la tendencia del problema antes de la respuesta. También permite determinar la fluctuación normal del problema antes de la respuesta. A continuación, se toman muchas mediciones del problema después de la respuesta. Esto permite determinar la tendencia a largo plazo del problema después de la respuesta. Puede observar si el problema se recupera o se mantiene a la baja. La comparación de la tendencia anterior con la posterior proporciona un indicador de la eficacia. Esto es factible si se utilizan los datos sobre incidencia reportados por la policía u otra información recopilada y archivada de forma rutinaria por organizaciones públicas y privadas. Es más difícil de llevar a cabo si hay que iniciar un esfuerzo especial de recopilación de datos, como encuestas a la población.

El enfoque básico consiste en utilizar mediciones repetidas del problema antes de la respuesta para prever el nivel probable del problema después de la respuesta. Si la diferencia entre las mediciones realizadas tras la respuesta y la previsión es significativa y negativa, indica que la respuesta ha sido eficaz (véase el Apéndice A).

Este tipo de diseño proporciona pruebas sólidas de que la respuesta se produjo antes de la disminución del problema, porque se pueden identificar tendencias preexistentes. Si el proceso de medición del problema no ha cambiado, este diseño elimina la mayoría de las explicaciones alternativas de la reducción del problema.

Tenga en cuenta que lo que importa es el número de periodos de medición, no la duración. Así, por ejemplo, es mucho menos útil tener tres años de datos anuales antes de la respuesta y tres años de datos anuales después de la respuesta que tener treinta y seis meses de mediciones mensuales antes y treinta y seis meses de mediciones mensuales después, aunque haya transcurrido la misma cantidad de tiempo.

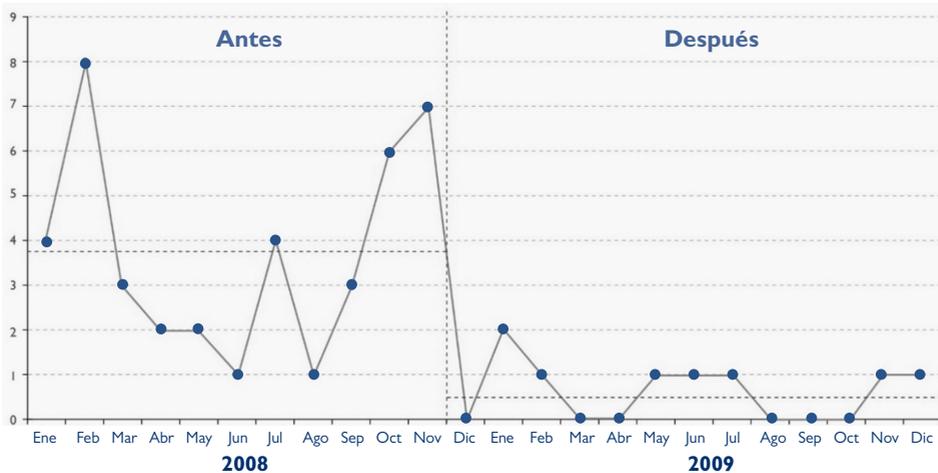
Se podría caer en la tentación de llevar esto al extremo: si los datos mensuales son mejores que los anuales, ¿por qué no los semanales, diarios o incluso datos por hora? La respuesta es que a medida que el intervalo de tiempo se acorta, el número de delitos por intervalo de tiempo se vuelve demasiado pequeño para poder extraer conclusiones significativas. Si el número de sucesos es extremadamente grande (como ocurre a veces cuando se utilizan datos de llamadas de servicio para grandes áreas), podrían ser útiles intervalos muy cortos. Pero si el número de sucesos es muy pequeño (como homicidios, violaciones entre extraños o muertes en accidentes de tráfico en una ciudad de tamaño modesto), puede que haya que utilizar intervalos grandes de tiempo.

La Figura 7 ilustra un diseño simple de series temporales y lo contrasta con un diseño pre-post. Este ejemplo procede de un informe sobre el esfuerzo de resolución de problemas de la policía de South Yorkshire para combatir el robo de metales. Una forma de robo de metales era el robo de calderas de calefacción en edificios residenciales. La figura muestra la frecuencia de estos robos de casa habitación a lo largo de veinticuatro meses (once antes de la respuesta y trece después). Obsérvese la gran variación de este tipo de robos antes de la intervención. Una simple comparación pre-post (a la derecha) no capta esto, por lo que la evaluación es vulnerable a los problemas señalados anteriormente. En el gráfico de series temporales, interrumpido por la línea que indica el inicio de la intervención, se observa claramente que tanto el número de estos robos y la fluctuación de su número disminuyeron considerablemente tras la respuesta. Se trata de una prueba más convincente de que es poco probable que la tendencia, la fluctuación natural o la falta de sostenibilidad sean responsables de la disminución.

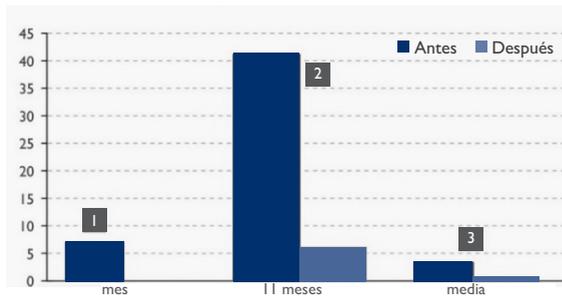
Una comparación del nivel promedio del problema antes y después muestra una disminución del problema tras la respuesta. Si la tendencia del problema hubiera sido ascendente, se habría utilizado una proyección ascendente y habría que calcular la pendiente (en el Apéndice A se muestra un ejemplo).

Entre más periodos de tiempo se examinen antes de la respuesta, más seguro se estará de conocer la trayectoria del problema antes de la respuesta. Entre más periodos de tiempo se examinen después de la respuesta, más seguro se estará de que la trayectoria ha cambiado. Los cálculos implicados en el análisis de un diseño de series temporales interrumpidas pueden llegar a ser bastante complicados; por lo tanto, si hay mucho en juego en el resultado de la evaluación, puede merecer la pena buscar la ayuda de un experto.

FIGURA 7: MEDICIÓN DEL IMPACTO EN UN DISEÑO DE SERIES TEMPORALES INTERRUPTIDAS



La policía de South Yorkshire optó por representar algunos de los resultados de su evaluación en un gráfico de series temporales interrumpidas (arriba). Este gráfico muestra 11 meses de robos en los que sustrajeron una caldera del sistema de calefacción antes de la intervención y 13 meses después de la respuesta. El impacto de la intervención es evidente, a pesar de las fluctuaciones mensuales. Antes de la intervención, sólo hubo dos meses con un solo robo y la mayoría de los meses tuvieron tres o más. Después de la intervención, más de la mitad de los meses no tuvieron ningún robo y solo un mes tuvo hasta dos. El número medio de robos, antes (3.7) y después (0.5), se muestra mediante líneas horizontales discontinuas. No todos los gráficos de series temporales muestran mejoras tan claras.



Si la policía de South Yorkshire hubiera utilizado un diseño pre-post, habría tenido al menos tres opciones (izquierda). 1) Podrían haber seleccionado los meses anteriores y posteriores a la respuesta, mostrando una disminución de 7 a 0. Esto exagera el impacto, ya que da a entender que acabaron con el problema, cuando no fue así. 2) Podrían haber mostrado el total de los 11 meses anteriores y los 11 meses posteriores. 3) O podrían comparar el número medio (durante 11 meses) de robos, antes y después. Las tres opciones ocultan las variaciones mensuales y ninguna muestra que la tendencia y la variación natural sean causas improbables de la reducción.

Gráficos extraídos de South Yorkshire Police (2010). *Shiregreen Alliance Partnership*. Goldstein Award Submission. Annual Problem-Oriented Policing Conference. Arlington, TX. Septiembre. Tabla 4. Accesible en <http://www.popcenter.org/library/awards/goldstein.cfm?browse=abstracts&year=2010>.

Incluso si sólo está interesado en determinar si el problema ha disminuido (y tiene poco interés en establecer la causa de la disminución), un diseño de series temporales interrumpidas sigue siendo superior a un diseño pre-post. Esto se debe a que un diseño de series temporales interrumpidas puede mostrar si el problema disminuyó y se mantuvo. Como ya se ha indicado, los problemas pueden fluctuar, por lo que conviene determinar la estabilidad de la disminución. Entre más larga sea la serie temporal tras la disminución, más seguro podrá estar de que el problema se ha eliminado o se mantiene estable en un nivel mucho más bajo.

Aunque los diseños de series temporales interrumpidas son superiores a los diseños pre-post, no siempre resultan prácticos. He aquí cinco razones comunes para ello:

- La medición es cara o difícil.
- Los datos de muchos periodos anteriores a la respuesta no están disponibles.
- Los responsables de la toma de decisiones no pueden esperar a que transcurra el tiempo suficiente tras la respuesta.
- Las prácticas de registro de datos han cambiado, lo que invalida las comparaciones entre periodos.
- Los sucesos problemáticos son raros para intervalos de tiempo cortos, lo que obliga a utilizar menos intervalos más largos.
- En estas condiciones, un diseño pre-post podría ser la alternativa más práctica.

Combinación y selección de diseños

Aunque hemos examinado estos diseños por separado (aquí y en el Apéndice B), en muchas circunstancias es posible utilizar dos o más diseños para probar la eficacia de una respuesta.

Esto resulta especialmente útil si se dispone de varias mediciones del problema (por ejemplo, datos de delitos denunciados e información de encuestas ciudadanas) recopiladas durante distintos periodos. Una combinación de diseños seleccionados para descartar explicaciones alternativas especialmente difíciles de refutar puede ser mucho más útil que la adhesión estricta a un único diseño.

El Apéndice C ofrece una lista de comprobación estructurada para seleccionar el diseño más apropiado para su problema. El Apéndice D resume los puntos fuertes y débiles de los diseños analizados aquí y en el Apéndice B.

Al considerar qué tipo de diseño o combinación de diseños podría seleccionar, es importante tener en cuenta que no se pueden eliminar todas las explicaciones alternativas de la reducción del problema. El tiempo, el dinero y la experiencia en evaluación aconsejan seleccionar el diseño más sencillo que elimine las explicaciones alternativas más obvias. En consecuencia, es útil anticiparse a las explicaciones alternas más creíbles antes de seleccionar un diseño de evaluación. Una vez más, su análisis del problema debería de darle alguna idea. También es útil escuchar a los críticos más elocuentes de la respuesta. Entonces, mientras planifica la evaluación, puede recopilar datos y desarrollar diseños que aborden las preocupaciones de los críticos.

Examinar cómo funciona la respuesta

Muchas respuestas de resolución de problemas constan de varias partes, cualquiera de las cuales puede ser eficaz, y algunas pueden no serlo. Además, como hemos señalado al hablar de las evaluaciones del proceso, a veces algunas partes de la respuesta no se aplican o se aplican mal. La recopilación y el examen de pruebas sobre la implementación, así como la plausibilidad de explicaciones alternativas, ayudan a determinar qué características de la respuesta (si las hubo) atendieron mínimamente el problema y cuáles fueron inútiles. Esto puede ilustrarse volviendo a nuestro ejemplo hipotético de un esfuerzo por reducir la prostitución callejera.

La Tabla 3 enumera una serie de explicaciones sobre cómo puede haber disminuido la actividad de la prostitución. Tales explicaciones se denominan "mecanismos". Como ya se ha explicado, los mecanismos son formas plausibles en que la respuesta podría reducir el problema.

Los mecanismos describen *cómo* funciona la respuesta. Los cinco primeros mecanismos de la tabla se basan en la respuesta prevista. Los dos últimos son mecanismos mediante los cuales explicaciones alternativas podrían haber reducido el problema. La segunda columna describe lo que ocurrió (o no ocurrió). Así, por ejemplo, en la primera fila, vemos que, a pesar de los planes, no se detuvo a muchas personas que ofrecen servicios sexuales. Y en la sexta fila, vemos que un subproducto no previsto de la reconstrucción de la calle fue la presencia de trabajadores de la carretera. La tercera columna muestra las pruebas que apoyan o contradicen la presencia del mecanismo. La última columna resume nuestras conclusiones sobre la probabilidad de que cada mecanismo influyera en el problema.

Esta tabla muestra que la mayor parte de la respuesta prevista probablemente no tuvo ningún impacto en el problema. Una parte puede haber sido responsable: la reconfiguración de la calle. Además, la tabla sugiere que hay que considerar otras dos explicaciones alternativas: la presencia de equipos de construcción y la presión de los vecinos. También podría ser una combinación de estas tres cosas.

El mecanismo de los equipos de construcción podría refutarse si la actividad de la prostitución se ha mantenido baja mucho después de que las cuadrillas se hayan ido. Sin embargo, si vuelve la prostitución, la presencia de los equipos podría ser la mejor explicación.

Examinar la respuesta descomponiéndola en los mecanismos que la componen y examinando los mecanismos desde explicaciones alternativas resulta útil si se piensa volver a utilizar la respuesta, o incluso, si sólo se quiere mantener la respuesta. En este caso, podríamos detener los esfuerzos para detener y desviar a las personas activas que ofrecen servicios sexuales. En su lugar, podríamos supervisar el impacto de la reconfiguración de las calles. Si utilizáramos esta respuesta en otros lugares, es posible que queramos movilizar proactivamente a los residentes del barrio como parte de las fases de análisis y respuesta.

La Tabla 3 ilustra un procedimiento sencillo para examinar los mecanismos. Las seis primeras respuestas se refieren a la respuesta prevista y su mecanismo. Esto ayuda a evaluar qué partes de la respuesta podrían haber funcionado y cuáles no. El séptimo punto es una alternativa a la respuesta que podría haber sido responsable del cambio en el problema. Esto ayuda a determinar si hubo razones para el cambio en el problema que no eran partes planificadas de la respuesta. Los objetivos que tiene son: (1) eliminar los mecanismos menos plausibles, (2) llegar a una conclusión razonable sobre si la respuesta pudiera haber causado el cambio y (3) determinar qué otros mecanismos alternativos pudieran haber actuado. De este modo obtiene una evaluación básica de lo que pudiera haber causado el cambio en el problema. También le dice algo acerca de si es probable que obtenga el mismo cambio en el problema si vuelve a utilizar la respuesta.

TABLA 3: LA RESPUESTA PUEDE HABER DESENCADENADO UNO O VARIOS DE ESTOS MECANISMOS PARA REDUCIR LA ACTIVIDAD DE PROSTITUCIÓN

RESPUESTA UTILIZADA	MECANISMO: CÓMO	ACCIONES PREVISTAS: QUÉ	EVIDENCIA DEL PROCESO	CONCLUSIÓN
Personas que ofrecen servicios sexuales arrestadas	Disuasión de personas que ofrecen servicios sexuales disuadidas	El operativo se produjo tarde	Pocas detenciones registradas	Poco probable

RESPUESTA UTILIZADA	MECANISMO: CÓMO	ACCIONES PREVISTAS: QUÉ	EVIDENCIA DEL PROCESO	CONCLUSIÓN
Prohibición de entrada a la zona a Personas que ofrecen servicios sexuales condenadas	Disuasión de personas que ofrecen servicios sexuales disuadidas	Pocas personas que ofrecen servicios sexuales procesadas	Pocas personas que ofrecen servicios sexuales recibieron órdenes de prohibición	Poco probable
Salida legítima del trabajo sexual	Reducción del número de personas que ofrecen servicios sexuales	Aplicación del plan de desvío	Pocas solicitudes de asistencia	Poco probable
Proporcionó a las personas que ofrecen servicios sexuales un empleo legítimo	Reducción del número de personas que ofrecen servicios sexuales	Aplicación del plan de desvío	Pocas personas que ofrecen servicios sexuales encontraron un empleo legítimo	Poco probable
Alteración de la red de carreteras y del flujo de tráfico	Dificultó el acceso de los clientes a las personas que ofrecen servicios sexuales	Cambios en el tráfico	Modificación de la frecuencia de tráfico en las calles laterales	Plausible
Reconstrucción de la calle	Obstaculizar la oferta de servicios sexuales	Se implementaron obras en la calle	Se observó que ocurrió mucha actividad de construcción	Plausible
Los vecinos presionaron a los propietarios	Denegación de lugares de trabajo a las personas que ofrecen servicios sexuales	No forma parte del proyecto, pero ocurrió más o menos al mismo tiempo	Las entrevistas con residentes y propietarios confirmaron que se ejerció presión	Plausible

Dado que normalmente no se puede depender de las explicaciones alternativas que operan de nuevo, entre más pesen las pruebas a favor de estas alternativas, menos probable será que el uso repetido de la respuesta sea eficaz.

Desplazamiento y difusión de beneficios

Una preocupación habitual sobre las respuestas de resolución de problemas es que provoquen un desplazamiento espacial, es decir, que la delincuencia o el desorden se desplacen de la zona objetivo a las zonas cercanas que no están siendo atendidas. Esta preocupación probablemente no sea tan grande como se suele imaginar.⁷ Aunque el desplazamiento dista mucho de ser inevitable, es una posibilidad que debe investigarse. Cada vez hay más pruebas de que algunas respuestas tienen efectos positivos que se extienden más allá de sus zonas objetivo.⁸ Esto se denomina "difusión espacial de los beneficios de la prevención de la delincuencia". Aunque no todas las respuestas generan beneficios más allá de los previstos, algunas sí lo hacen, y esta posibilidad también debe tenerse en cuenta en las evaluaciones. Si no tenemos en cuenta estas posibilidades, podríamos obtener resultados engañosos.

No examinaremos aquí el desplazamiento ni la difusión. Véase la *Guía de Herramientas para la Resolución de Problemas n.º 10, Análisis del desplazamiento y la difusión de delitos*, para profundizar en el desplazamiento y la difusión de beneficios, cómo detectarlos y cómo medir su impacto.

07 CONCLUSIONES

Esta guía ha introducido algunos principios básicos para evaluar la eficacia de las respuestas a los problemas. Todas las evaluaciones requieren mediciones válidas del problema que se tomen sistemáticamente antes y después de la implementación de una respuesta. Hay dos objetivos posibles para cualquier evaluación de resolución de problemas. El primero es demostrar que el problema disminuyó lo suficiente. Este es el requisito más básico de una evaluación. Para este objetivo, no nos preocupa si la reducción fue causada directamente por la respuesta o por algo totalmente distinto.

En muchas circunstancias, también es útil determinar si la disminución del problema se debió a la respuesta. Esto es un segundo objetivo. Si se prevé volver a utilizar la respuesta en problemas similares (o en el mismo problema si vuelve a aparecer), es importante tomar esta determinación. Esto requiere un diseño de evaluación que pueda eliminar las explicaciones alternativas más probables para la disminución del problema.

La eliminación de esas explicaciones requiere el uso de un diseño de series temporales interrumpidas o el uso de un grupo de control (Apéndice B). El grupo de control le indica cuál habría sido probablemente el nivel del problema en ausencia de este esfuerzo de resolución de problemas.

Los resultados de una evaluación de impacto deben compararse con los de una evaluación de proceso para formarse una imagen detallada de si la respuesta se aplicó según lo previsto y qué impacto tuvo en el problema. Esta información ayuda a mostrar si la respuesta fue la causa de la disminución del problema, y qué partes de la respuesta constituyen el "ingrediente activo".

Un tema recurrente en esta guía es que el diseño de una evaluación se basa en los conocimientos adquiridos durante el análisis del problema. Las evaluaciones competentes requieren que los evaluadores tengan un conocimiento detallado del problema para poder desarrollar mediciones útiles y anticipar posibles razones de la disminución del problema tras una respuesta.

La evaluación de las respuestas puede ser extremadamente compleja. Esta guía es sólo una introducción. Para la resolución de problemas a pequeña escala, donde los costes de conclusiones erróneas no son graves y las inferencias causales débiles son tolerables, la información aquí contenida debería ser suficiente. Sin embargo, si hay mucho en juego en el resultado, si es importante demostrar si la respuesta causó una disminución del problema o si hubiera consecuencias graves si se llegara a una conclusión errónea de la evaluación debe buscar ayuda profesional para desarrollar una evaluación rigurosa. La decisión de contar con la ayuda de un evaluador externo debe tomarse lo antes posible una vez que se haya identificado el problema, de modo que puedan tomarse mediciones adecuadas antes de la respuesta y pueda desarrollarse un diseño riguroso.

08 APÉNDICE A

LOS EFECTOS DEL NÚMERO DE PERIODOS EN LA VALIDEZ DE LAS CONCLUSIONES DE LA EVALUACIÓN

Para comprender la importancia de examinar un gran número de periodos de tiempo, considere el siguiente ejemplo hipotético. Los datos se crearon utilizando un generador de números aleatorios, por lo que ninguna de las fluctuaciones es sistemática. Esta serie ilustra cómo nos puede engañar la aleatoriedad, sobre todo si nos fijamos en intervalos de tiempo muy cortos. Todos los gráficos que siguen proceden de la misma serie.

La Figura A1 muestra los resultados de una evaluación pre-post en la que se toman mediciones del problema justo antes y justo después de una respuesta (periodos 19 y 20 de la serie). La conclusión que extraeríamos de este gráfico es que el problema experimentó una disminución moderada tras la respuesta.

El siguiente gráfico (Figura A2) muestra los periodos 12 a 20 de la serie, por lo que ahora hay ocho periodos antes de la respuesta y uno después de la respuesta. Los periodos adicionales proporcionan una oportunidad para examinar la tendencia del problema que conduce a la respuesta. La línea recta muestra esta trayectoria. La extensión de la trayectoria al periodo 20 permite comparar lo que cabría esperar si no se aplicara la respuesta (la trayectoria) con el nivel real del problema.

Podemos ver claramente que la tendencia del problema era descendente antes de la respuesta, por lo que no todo el descenso del problema tras la respuesta puede atribuirse a la respuesta.

No obstante, parece que la disminución del problema tras la respuesta fue mayor de lo que cabría esperar sólo por la tendencia.

Los periodos anteriores a la respuesta ayudan a establecer la trayectoria de la serie temporal del problema. Aquí nos hemos centrado exclusivamente en la tendencia general, pero también es posible buscar ciclos estacionales y otras fluctuaciones recurrentes.

Ampliar los datos a periodos posteriores a la respuesta ayuda a determinar la estabilidad de esta. ¿Sigue siendo eficaz la respuesta, reduciendo aún más el problema? ¿O la respuesta se desvanece, permitiendo que el problema rebote? Esto se muestra en la Figura A3, que representa otros siete periodos posteriores a la respuesta. Se utiliza la misma línea de tendencia basada en los datos anteriores a la respuesta, pero ahora proyectada a ocho periodos de tiempo después de la respuesta.

FIGURA A1: DISEÑO PRE-POST DE DOS PERIODOS

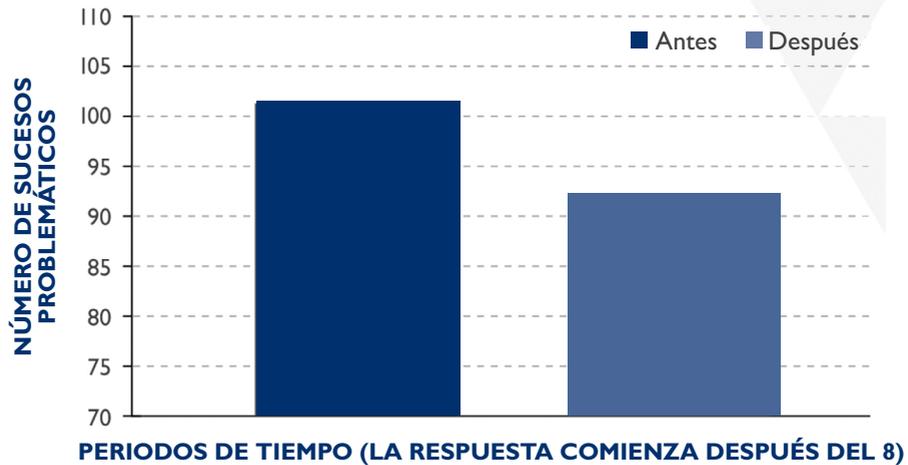
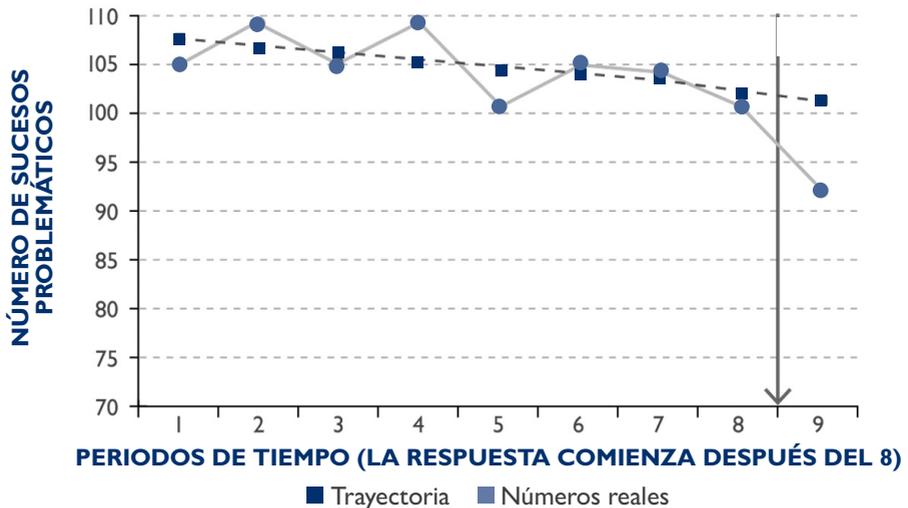


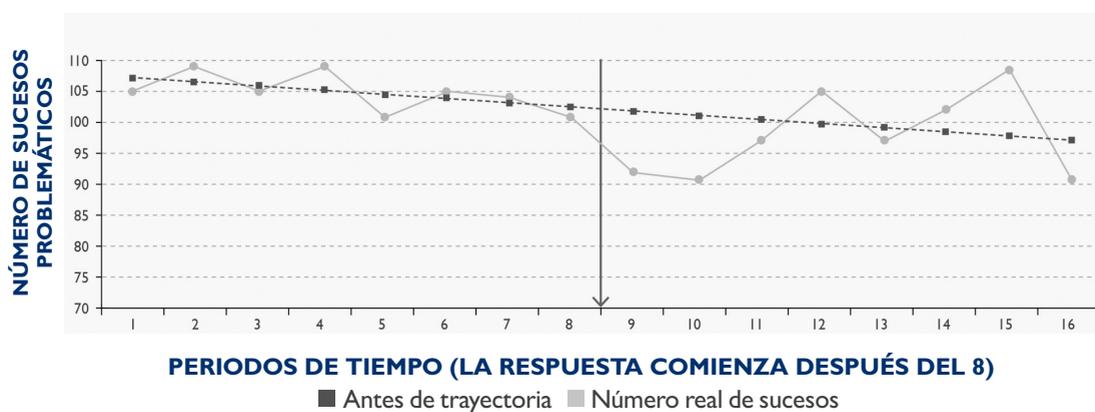
FIGURA A2: DISEÑO DE SERIES TEMPORALES DE NUEVE PERIODOS (CON TRAYECTORIA PROYECTADA DE PROBLEMAS)



Vemos que el problema repuntó y luego parece oscilar en torno a la misma línea de tendencia. Así que, en el mejor de los casos, la respuesta fue temporalmente útil.

No se puede exagerar el valor de una serie temporal muy larga. Con demasiada frecuencia, las corporaciones policiales muestran sólo unos pocos periodos de tiempo, aunque sus sistemas informáticos contengan datos para muchos más. Note cómo cambia nuestra interpretación de la tendencia cuando observamos la serie completa de 40 periodos de la que se extrajeron estos tres gráficos. Esto se muestra en la Figura A4.

FIGURA A3: DISEÑO DE SERIES TEMPORALES DE DIECISÉIS PERIODOS (CON TRAYECTORIA PROYECTADA DEL PROBLEMA)



Mientras que la tendencia de la Figura A3 sugiere una tendencia a la baja, la Figura A4 muestra que se trata de una ilusión; de hecho, la tendencia a largo plazo es plana. La aleatoriedad subyacente de los datos se hace mucho más evidente. Oscila en torno a 101 eventos por periodo (línea de puntos). Para minar aún más nuestra confianza en la respuesta, podemos ver que hay al menos otros dos intervalos con disminuciones como los que vemos después de la respuesta. Así que parece que lo que pensábamos que era una disminución debido a la respuesta puede muy bien ser una fluctuación temporal debida a la variación normal del problema.

La lección aquí es que es fácil dejarse engañar por la aleatoriedad, especialmente cuando se analiza la delincuencia en periodos de tiempo cortos.

La policía, el público, los medios de comunicación y los funcionarios son todos susceptibles a este engaño, porque comparar este mes con el mes pasado o este año con el año pasado es muy habitual.

FIGURA A 4: DISEÑO DE SERIES TEMPORALES DE CUARENTA PERIODOS (CON EL NÚMERO PROMEDIO DE SUCESOS POR PERIODO)



El uso de mediciones múltiples y de series temporales más largas es una protección razonable contra este tipo de engaño.

A diferencia de los datos reales, en los que nunca se está seguro de la causa, con estos datos intencionadamente aleatorios, sabemos con absoluta certeza que la variación en torno a la media de 101 sucesos por periodo es aleatoria.ⁿ Esto incluye los periodos justo antes y después de la respuesta. El ejemplo ilustra el hecho de que las fluctuaciones aleatorias de los datos pueden malinterpretarse fácilmente como cambios significativos. Merece la pena señalar que una prueba de significación para detectar aleatoriedad en un diseño pre-post podría sugerir en realidad que la caída *no* se debe a cambios aleatorios. Esto se debe a que la aleatoriedad afecta a toda la serie y el diseño pre-post sólo analiza una pequeña parte de la serie.

ⁿ Creamos esta serie de datos estableciendo un nivel constante para el problema y luego utilizamos un generador de números aleatorios para proporcionar las fluctuaciones en torno a este nivel. Colocamos el inicio de la respuesta hipotética en el centro de la serie.

09 APÉNDICE B

DISEÑOS CON Y SIN GRUPOS DE CONTROL

Los diseños del cuerpo principal del texto se centran en los datos del grupo de personas o la zona que recibe la respuesta. Para determinar si la respuesta es la causa de una disminución del problema, es útil utilizar un grupo de control. Además, los grupos de control son fundamentales para obtener estimaciones razonables de la cantidad de desplazamiento espacial y difusión de beneficios (véase la *Guía de Herramientas para la resolución de problemas n.º 10, Análisis del desplazamiento y la difusión de la delincuencia*). Pueden añadirse grupos de control al diseño pre-post o al diseño de series temporales.

En este apéndice, analizaremos cinco diseños, incluidos los dos examinados en el cuerpo de esta guía. Utilizaremos datos de una evaluación de una iniciativa de resolución de problemas para reducir los accidentes de tráfico fatales o con personas lesionadas en Cincinnati. La evaluación utilizó un diseño de series temporales múltiples y un proceso de análisis estadístico muy complejo para obtener una estimación precisa del número de vidas salvadas y lesiones evitadas mediante la implementación de una respuesta a los accidentes de tráfico con personas lesionadas. Los autores constataron que los accidentes se redujeron entre un 5.7% y un 10.3% en Cincinnati, en comparación con las zonas de comparación. Esta evaluación fue posible gracias a una larga colaboración entre el Departamento de Policía de Cincinnati y el Instituto de Ciencias del Delito de la Universidad de Cincinnati (con sede en la Facultad de Justicia Penal).^o

Aquí no reproduciremos el análisis realizado en el artículo publicado. En su lugar, utilizaremos los datos para ilustrar cómo pueden cambiar las conclusiones sobre la eficacia de la respuesta, dependiendo del diseño de evaluación utilizado. Empezaremos usando los datos sobre los accidentes de tráfico de Cincinnati para ilustrar un diseño que no debería utilizarse: un grupo de comparación estático.

^o Gracias a los doctores Nick Corsaro y Robin Engel, del Instituto de Ciencias del Delito, perteneciente a la Escuela de Justicia Penal de la Universidad de Cincinnati, por facilitarnos estos datos. El Instituto de Ciencias del Delito presta servicios de consultoría científica a la policía y a otros organismos encargados de hacer cumplir la ley, incluidas evaluaciones complejas.

Esta será nuestra línea de base. A continuación, mostraremos por qué el diseño pre-post supone una mejora. Luego mostraremos por qué es útil un grupo de control. Tras esto, volveremos al diseño de series temporales. Concluiremos mostrando un diseño de series temporales con un grupo de control. Este breve tutorial es una introducción a los diseños de evaluación, cuyo único objetivo es ilustrar su lógica básica.

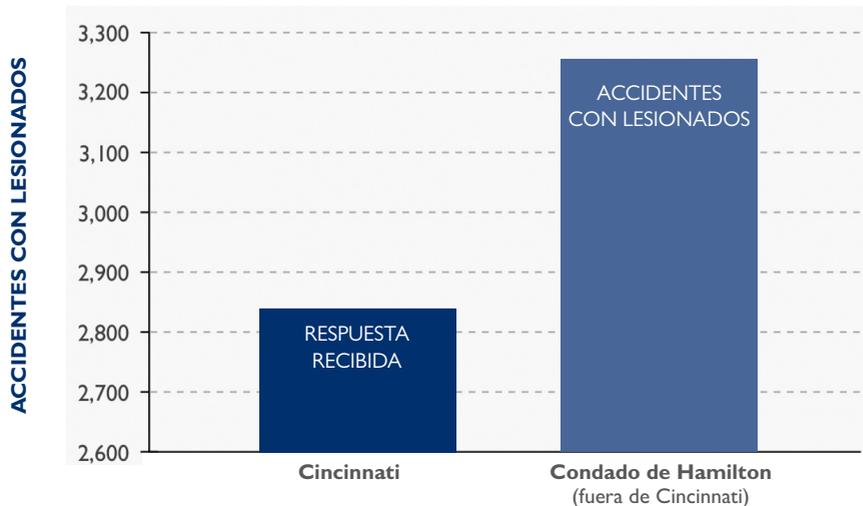
DISEÑO DE COMPARACIÓN ESTÁTICA

Supongamos que un año después de que la División de Tráfico del Departamento de Policía de Cincinnati pusiera en marcha su respuesta, cuyo objetivo era disminuir el número de accidentes de tráfico con lesionados, se le pide que determine si marcó la diferencia. Un diseño habitual, que no se recomienda, consiste en comparar el número de accidentes con lesionados en Cincinnati con las cifras de las jurisdicciones cercanas que no utilizaron la respuesta.

La lógica es que estas jurisdicciones cercanas estarían expuestas a las mismas condiciones de tráfico y a los mismos conductores, por lo que deberían de tener un nivel similar de accidentes. Es decir, usted está asumiendo que, si funcionó la respuesta, Cincinnati debería de tener menos accidentes que la zona de comparación, y que, si Cincinnati no hubiera utilizado la respuesta, su nivel de accidentes sería similar al de la zona de comparación.

La Figura B1 muestra los resultados. Cincinnati está dentro del condado de Hamilton, por lo que el condado de Hamilton (sin Cincinnati) es la comparación. Al dividir el número de accidentes durante un periodo de doce meses entre la población conductora de cada jurisdicción (o las millas conducidas en las zonas), se controlarían las diferencias de población. No lo hacemos aquí por una sencilla razón: el principal problema de este diseño es que el área de comparación es sistemáticamente diferente del área de respuesta (tienen diferentes poblaciones de conductores, hay más carreteras en un área que en la otra, la población es mayor en un área que en la otra, etc.). La población es sólo un área en la que puede haber muchas diferencias.

FIGURA B I: DISEÑO DE COMPARACIÓN ESTÁTICA



En un diseño de comparación estática, compare el problema en un área o grupo que recibió una respuesta con un área o grupo similar que no la recibió. El periodo de tiempo para ambos es posterior a la respuesta.

El área o grupo que no recibe una respuesta proporciona un indicador del nivel del problema en el área o grupo de respuesta, si no se hubiera aplicado la respuesta.

En este caso, la respuesta del Departamento de Policía de Cincinnati solo tiene lugar dentro de la ciudad. La comparación de los 12 meses siguientes a la respuesta en Cincinnati con los mismos 12 meses en el condado circundante muestra un menor número de accidentes con heridos.

Aunque parece demostrar que la respuesta funciona, se trata de una prueba débil, ya que en Cincinnati suele haber menos accidentes con lesionados.

Debe evitar utilizar este tipo de diseño de evaluación, ya que tiene un alto riesgo de producir resultados engañosos. Lo engañoso que resulta puede apreciarse comparando los resultados de la Figura B I con los de la siguiente serie de figuras, la cuales ilustran diseños mejores.

DISEÑO PRE-POST SIN GRUPO DE CONTROL

Ya hemos tratado este diseño en la parte principal de la guía, por lo que sólo lo revisaremos brevemente aquí. La Figura B2 muestra los resultados de la evaluación de la respuesta diseñada para reducir la incidencia de las lesiones por accidentes de tráfico en Cincinnati. La comparación es entre los doce meses anteriores a la respuesta y los doce meses posteriores. Utilizamos una comparación de todo el año porque controla los cambios estacionales en los accidentes. Un periodo más corto (por ejemplo, de septiembre anterior a la respuesta a septiembre posterior) es muy susceptible a los cambios aleatorios en los accidentes que una respuesta no puede tomar en cuenta. Con este diseño, actuamos como si la medición del antes fuera un indicador preciso del número de accidentes que habría tenido Cincinnati si no se hubiera aplicado ninguna respuesta. Por lo tanto, la diferencia entre las mediciones antes y después del problema es un indicador de la reducción debida a la respuesta.

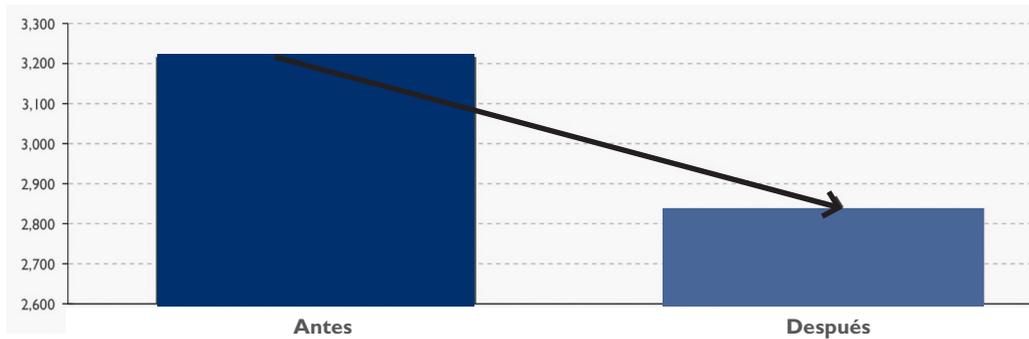
Los diseños pre-post son sencillos. Son más útiles si su interés principal es determinar si el problema disminuyó y no va a afirmar rotundamente que la respuesta fue una causa importante de la disminución.

DISEÑO PRE-POST CON GRUPO DE CONTROL^P

Si combinamos el uso de un control en el diseño de comparación estática con el uso de una medición del problema antes de la respuesta en el diseño pre-post podemos mejorar la evaluación. El área o grupo de control no recibe la respuesta, aunque tenga un problema similar al del área o grupo que recibe la respuesta. La finalidad del grupo de control es demostrar lo que habría ocurrido si no se hubiera adoptado ninguna respuesta. Saber esto puede ayudarle a eliminar algunas explicaciones alternativas para la disminución del problema.

Este diseño se ilustra en la Figura B3. Aquí vemos que, dentro del condado, pero fuera de Cincinnati, se experimentó una disminución en las colisiones de vehículos con lesionados desde antes hasta después de la respuesta dentro de Cincinnati. Esto indica que incluso sin una respuesta, Cincinnati podría haber experimentado una disminución similar. Sin embargo, la disminución de Cincinnati en accidentes con lesionados es mayor que el del condado de Hamilton (más de un 40% mayor). Esto indica que la respuesta en Cincinnati contribuyó al descenso general.

^P Este diseño suele denominarse "diseño de grupo de control no equivalente" para llamar la atención sobre el hecho de que los miembros del grupo de tratamiento (respuesta) y los miembros del grupo de control pueden ser diferentes en aspectos importantes que podrían afectar al resultado de la evaluación.

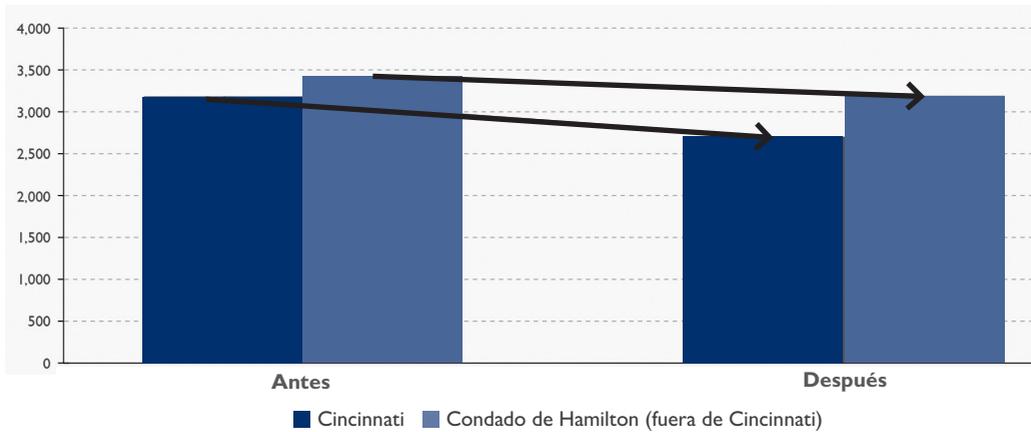
FIGURA B2: DISEÑO PRE-POST

En un diseño pre-post (también llamado diseño antes-después), se compara el nivel del problema después de la respuesta con el nivel del problema antes de la respuesta. No se utiliza un área de comparación o control. Se asume que la medición del problema antes es un indicador razonable de cómo sería el problema si se hubiera utilizado la respuesta. Por lo tanto, una disminución significativa del problema de antes a después es un indicador de éxito.

En este caso, el número de accidentes con lesionados es mayor antes de la respuesta del Departamento de Policía de Cincinnati que después. Esto es coherente con una respuesta eficaz.

Las limitaciones son obvias. En primer lugar, algo más podría haber ocurrido casi al mismo tiempo que la respuesta que podría haber causado la disminución de los accidentes. En segundo lugar, es posible que la tendencia de los accidentes con lesionados fuera a la baja antes de que se pusiera en marcha la respuesta.

FIGURA B3: DISEÑO PRE-POST CON CONTROL



Este diseño combina características de los diseños de comparación estática y pre-post. Se comparan las diferencias antes y después de los dos grupos. En este caso, los accidentes con heridos en Cincinnati disminuyeron un 1.6% desde antes de la respuesta hasta después. Los accidentes con heridos del condado de Hamilton disminuyeron sólo un 4.8%. Utilizamos el porcentaje de disminución porque las dos áreas tenían diferentes números de accidentes antes de la respuesta.

Se supone que la disminución en el condado de Hamilton (4.8%) es el que se habría producido en Cincinnati si no se hubiera llevado a cabo una respuesta de resolución de problemas en relación con los accidentes con lesionados. Para determinar el impacto de la respuesta, restamos 4.8 de 11.6 (ya que la disminución del 4.8% se habría producido presumiblemente tanto si se hubiera implementado una respuesta como si no). La respuesta de la policía de Cincinnati podría haber creado una disminución del 6.8% en los accidentes de tráfico con lesionados.

La razón por la que este diseño es mejor que la comparación estática y el diseño pre-post es que elimina de la consideración muchas causas alternativas para la disminución de Cincinnati. Por ejemplo, los cambios en la legislación estatal o en los precios de la gasolina afectarían tanto al condado como a la ciudad. Por tanto, no pueden explicar las diferentes tasas de disminución. Del mismo modo, si la tendencia de los accidentes con heridos fuera a la baja en todo el estado, esto también influiría en ambas jurisdicciones, por lo que no puede explicar la diferencia.

La principal limitación de este diseño es que podría haber habido algo (aparte de la respuesta) que ocurriera en Cincinnati y no en el condado (o viceversa) que empujara a la baja los accidentes con heridos en la ciudad.

Cincinnati tuvo una reducción de 375, lo que dividido por el número anterior (3215) supone un descenso del 11.6%. Si se resta el porcentaje de disminución en el control del porcentaje de disminución en la respuesta, se obtiene una reducción neta del 6.8% (dividiendo -6.8 entre -4.8 muestra que la caída de Cincinnati fue casi un 42% mayor que la del condado).

TABLA B1: CÁLCULO DE LA EFICACIA CON UN DISEÑO PRE-POST CON CONTROL

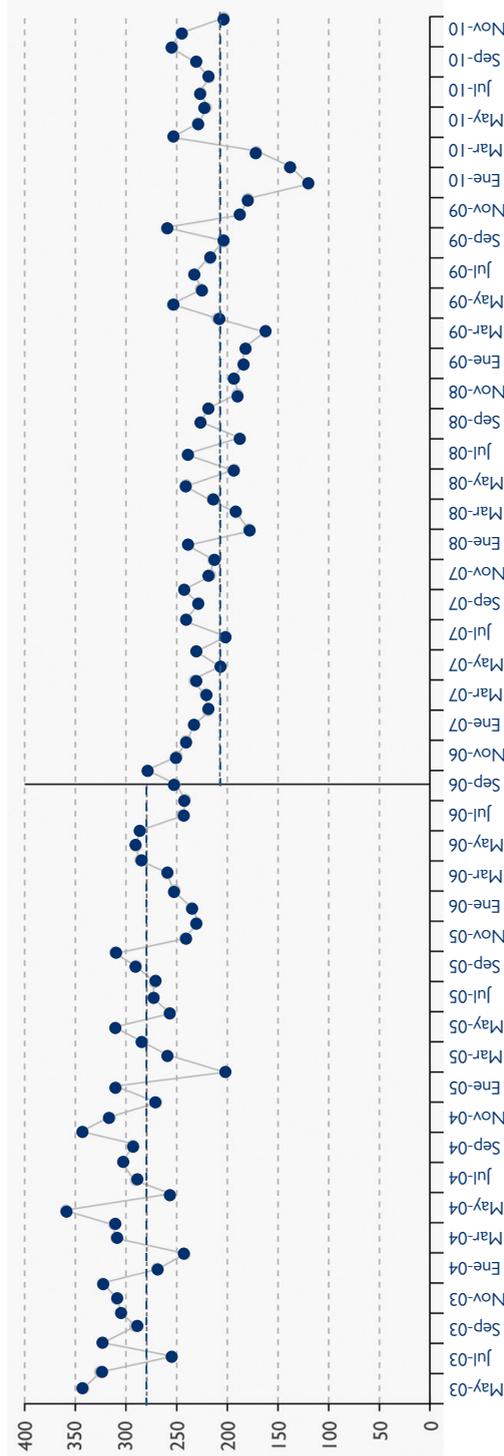
	ANTES	DESPUÉS	DIFERENCIA	% DIFERENCIA
Control (Condado)	3,421	3,257	-164	-4.8
Respuesta (Cincinnati)	3,215	2,840	-375	-11.6
Respuesta % diferencia – Control % diferencia				-6.8

DISEÑO DE SERIES TEMPORALES

Este diseño también se discutió en el cuerpo principal de la guía. La Figura B4 muestra la serie temporal de las colisiones de vehículos con lesionados en Cincinnati. Las líneas horizontales discontinuas indican el número medio de colisiones de este tipo por mes, antes y después de la respuesta. Los datos anteriores se utilizan para determinar lo que podría haber ocurrido en Cincinnati si no se hubiera implementado ninguna respuesta.

En este caso, una simple comparación de estas medias sugiere una respuesta eficaz. Normalmente, el analista utiliza un procedimiento estadístico más complejo para eliminar los efectos de las tendencias (en este caso a la baja) y los ciclos estacionales. Así se obtiene una estimación más precisa del impacto. Sin embargo, este tipo de análisis va mucho más allá de lo que puede explicarse en esta guía introductoria.

FIGURA B4: DISEÑO DE SERIES TEMPORALES



El diseño de series temporales amplía la medición anterior y posterior. En lugar de comparar dos periodos de doce meses, aquí observamos los cambios mensuales durante más de siete años. Esto nos permite ver tendencias y ciclos. Hay dos cosas que saltan a la vista. En primer lugar, el número de accidentes sube y baja mucho. Esto suele ocurrir con los problemas policiales, y es la razón por la que debemos desconfiar de las comparaciones mes a mes: las fluctuaciones aleatorias dominarán cualquier cambio sistemático. La prevención pretende producir un cambio sistemático, el cual la fluctuación aleatoria puede ocultar. La segunda cosa que podemos ver es que hay una lenta tendencia a la baja desde mayo de 2003 hasta aproximadamente mayo de 2009. Por tanto, aunque la policía de Cincinnati no hiciera nada nuevo, los heridos por accidentes habrían disminuido.

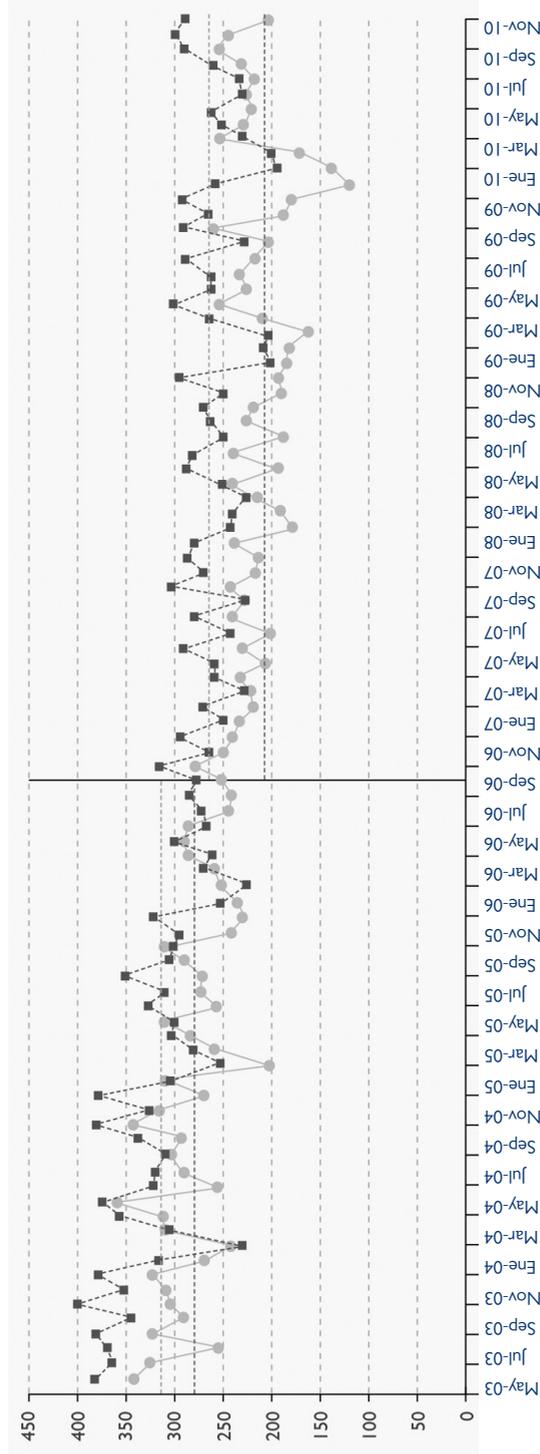
La línea vertical muestra cuándo comenzó la respuesta. Antes de la respuesta, había 287 accidentes en un mes medio (línea discontinua). Después de la intervención, solo hubo 220 accidentes de media al mes. Normalmente, se utiliza un complejo análisis estadístico de estos datos para desentrañar el impacto de la respuesta. El análisis de series temporales puede distinguir el efecto real de la respuesta de las tendencias y los ciclos estacionales, así como de otros factores mensurables (por ejemplo, cambios en los precios de la gasolina).

DISEÑOS DE SERIES TEMPORALES MÚLTIPLES

Cuando se utilizan dos o más series temporales, el diseño se denomina una serie temporal múltiple. Este diseño puede descartar la mayoría de las posibles explicaciones alternativas del cambio en el problema. La Figura B5 ilustra una serie temporal múltiple. Este ejemplo ilustra la utilidad de añadir una serie temporal de control. Si nos hubiéramos limitado a observar la Figura B4, podríamos haber supuesto legítimamente que gran parte de la disminución de las colisiones con lesionados en Cincinnati podría haberse debido a la tendencia descendente que precedió a la respuesta. En la Figura B5, vemos que la tendencia influyó tanto en el condado circundante como en la ciudad. Las diferencias en el número promedio de colisiones al mes entre la ciudad y el condado aumentaron tras la respuesta: antes, el condado tenía un promedio de 35 colisiones más que la ciudad en un mes; después, el condado tenía un promedio de 42 colisiones más que la ciudad en un mes. Al igual que el diseño simple de series temporales, los analistas utilizan técnicas estadísticas muy complejas. El estudio del que proceden estos datos ilustra parte de la complejidad involucrada.

La principal ventaja de utilizar un diseño de series temporales múltiples es que puede eliminar un gran número de explicaciones alternativas para una mejora del problema. La única alternativa posible a la afirmación de que la respuesta provocó la disminución es que algo ocurrió en Cincinnati aproximadamente al mismo tiempo que se implementó la respuesta, y esto se produjo en el condado (o se produjo en el condado, pero no en la ciudad). Así pues, los resultados de un diseño de series temporales múltiples, aunque sólidos, no son seguros. Sin embargo, a efectos prácticos, estos resultados probablemente superan el nivel de certeza que necesitamos para considerar que la respuesta ha tenido éxito.

FIGURA B5: DISEÑO DE SERIES TEMPORALES MÚLTIPLES



Un diseño de series temporales múltiples añade una o más áreas de control al análisis. Esto ayuda a eliminar posibles causas no relacionadas con la respuesta. En este caso, tanto la ciudad como el condado tenían tendencias muy similares antes de la respuesta, aunque la ciudad tendía a ser sistemáticamente inferior al condado en accidentes con lesionados (alrededor de 287 frente a 322 accidentes al mes). Tras la respuesta, la diferencia entre el condado y la ciudad aumentó (220 frente a 262 accidentes al mes). Ningún factor común a ambas jurisdicciones podría explicarlo, por lo que la única explicación es que algo diferente ocurrió en Cincinnati, en relación con el condado. La respuesta es una de esas explicaciones.

Al igual que el diseño de las series temporales, se necesita un complejo análisis estadístico para separar el impacto de la respuesta del impacto de otros factores y de la fluctuación aleatoria.

10 APÉNDICE C

LISTA DE VERIFICACIÓN PARA LA EVALUACIÓN DE LA RESOLUCIÓN DE PROBLEMAS

Esta lista de verificación ofrece un resumen de las cuestiones que deben tenerse en cuenta al evaluar una iniciativa de resolución de problemas. Se trata de una guía general y no un conjunto de reglas rígidas. Será más útil si la utiliza a lo largo de todo el proceso de resolución de problemas, empezando en la fase de exploración.

I. CONSIDERACIONES PREVIAS A LA EVALUACIÓN

Los puntos que aquí se exponen deben tenerse en cuenta durante las fases de exploración, análisis y respuesta.

A. ¿QUÉ DECISIÓN LE AYUDARÁ A TOMAR LA EVALUACIÓN?

- ¿Debería continuar este esfuerzo de resolución de problemas? Si esta es la única cuestión que la evaluación le ayudará a decidir, bastará con un diseño de evaluación simple (véase la pregunta III.A).
- ¿Debería utilizarse esta respuesta para problemas similares (por parte de su organismo o de otros organismos)? Si se marca, debería considerar la posibilidad de utilizar un grupo de control en el diseño de la evaluación de impacto (véase la pregunta III.A).
- Sin decisión. Si está marcada, una evaluación no será útil. Deténgase aquí.

B. ¿CONOCE EL PROBLEMA?

Estas preguntas ayudan a desarrollar una respuesta rentable y a evaluar la respuesta. Si no puede responder a estas preguntas con cierta precisión, necesita hacer más para analizar el problema.

- ¿A quién perjudica el problema? ¿A quién no perjudica?
- ¿Cómo puede medirse el daño?

- ¿Dónde ocurre el problema? ¿Dónde está ausente el problema?
- ¿Cuándo ocurre el problema? ¿Cuándo no ocurre?
- ¿Qué causa el problema? ¿Qué lo evita o suprime?

C. ¿SABE CÓMO FUNCIONA LA RESPUESTA?

Es necesario responder a estas preguntas para determinar si es probable que la respuesta sea eficaz y para garantizar la rendición de cuentas durante la implementación. Si no puede responderlas, los planes de respuesta son inadecuados y hay que prestar más atención a la fase de respuesta.

- ¿Cómo influye la respuesta en las causas del problema?
- ¿Quién es el responsable de llevar a cabo la respuesta?
- ¿Cuándo debe implementarse la respuesta?
- ¿Dónde debe implementarse la respuesta?
- ¿Cuánto tiempo pasará antes de que la respuesta tenga un impacto perceptible en el problema?
- ¿Quién tiene autoridad legal para llevar a cabo la respuesta?
- ¿Cuáles son los posibles obstáculos a la implementación de la respuesta?

II. EVALUACIÓN DE PROCESO

Las evaluaciones del proceso comienzan hacia el final de la fase de respuesta y continúan hasta bien entrada la fase de evaluación.

A. ¿SE IMPLEMENTÓ LA RESPUESTA?

Entre más se acerque la implementación real a la respuesta prevista, mayor confianza tendrá en que la respuesta fue la causa de los cambios en el problema detectados por la evaluación de impacto (sección III, más adelante). Cuanto mayor sea la variación entre lo previsto y lo ocurrido, mayor será la probabilidad de que los cambios en el problema se deban a causas distintas de la respuesta.

- ¿Se implementó cuando se suponía que debía implementarse?
- ¿Se implementó la respuesta donde se suponía que debía implementarse?
- ¿Se implementó la respuesta para las personas adecuadas?
- ¿La implementación se llevó a cabo según lo previsto?

B. ¿FUE SUFICIENTE LA IMPLEMENTACIÓN DE LA RESPUESTA?

Es posible que la respuesta se haya implementado según lo previsto, pero que haya carecido de los recursos, la duración o la intensidad necesarios para que fuera eficaz.

- ¿Se dispuso de recursos suficientes para implementar plenamente la respuesta?
- ¿Se llevó a cabo la respuesta durante el tiempo suficiente para tener un impacto?
- ¿Se llevó a cabo la respuesta con suficiente intensidad?

III. EVALUACIÓN DE IMPACTO

Muchas de las decisiones necesarias para llevar a cabo una evaluación de impacto deben considerarse en las fases de análisis y respuesta. Es el caso, en particular, de las decisiones de medición.

A. ¿NECESITA UN GRUPO DE CONTROL?

Responder a estas preguntas ayuda a decidir la complejidad del diseño de la evaluación.

- ¿Ha marcado la pregunta I.A.1? Si la respuesta es SÍ, no necesita un grupo de control.
- ¿Ha marcado la pregunta I.A.2? Si la respuesta es SÍ, debería utilizar un grupo de control.

B. ¿CON QUÉ FRECUENCIA SE PUEDE MEDIR EL PROBLEMA?

Responder a estas preguntas le ayudará a decidir si es posible un diseño de series temporales.

- ¿Se puede medir el problema de forma coherente durante muchos periodos de tiempo antes (al menos 10 periodos, aunque es mejor 30) y después (al menos 10, pero es preferible 30) de aplicar la respuesta? Si se comprueba, un diseño de series temporales es factible.
- ¿Se puede medir el problema sólo unas pocas veces antes de la respuesta y unas pocas veces después de la respuesta? En caso afirmativo, no es factible un diseño de series temporales y será necesario algún tipo de diseño pre-post.
- ¿Tiene algunas mediciones del problema que puedan examinarse durante muchos periodos de tiempo antes y después de la respuesta, y otras mediciones que sólo puedan examinarse durante unos pocos periodos de tiempo antes y después de la respuesta? Si la respuesta es afirmativa, puede utilizar un diseño de series temporales y una forma de diseño pre-post.

C. ¿QUÉ TIPO DE DISEÑO DE EVALUACIÓN DEBO ELEGIR?

Sus respuestas a las preguntas A y B, inmediatamente anteriores, proporcionan algunas orientaciones básicas para responder a esta pregunta. Esto se muestra en la siguiente tabla. Obviamente, las respuestas precisas dependen de las circunstancias particulares de cada problema.

TABLA C I: ¿QUÉ DISEÑO DE EVALUACIÓN TIENE MÁS SENTIDO?

III.B. PREGUNTA MARCADA	III.A. PREGUNTA MARCADA	
	1	2
1	Diseño de series temporales interrumpidas	Diseño de series temporales interrumpidas múltiples
2	Diseño pre-post	Diseño pre-post con grupo de control
3	Combinación de los diseños anteriores	Combinación de diseños anteriores

D. ¿QUÉ TIPO DE GRUPO DE CONTROL NECESITO?

(Sólo se aplica si se ha seleccionado uno de los diseños de la columna 2 de la Tabla D I. Si el diseño procede de la columna 1, sáltese esta sección y vaya directamente a la parte IV).

- ¿Se aplicará la respuesta en una zona geográfica identificable (lugar, barrio, etc.)? En caso afirmativo, el grupo de control debe ser una zona geográfica muy similar que tenga problemas parecidos a los que no recibirá la respuesta, y preferiblemente situada a cierta distancia de la zona problemática (para evitar que la respuesta se filtre a la zona de control y la contamine).
- ¿Se aplicará la respuesta a un conjunto de víctimas potenciales identificables (hombres jóvenes, mujeres mayores, personas que viajan diariamente al trabajo, etc.)? En caso afirmativo, el grupo de control debe ser un grupo muy similar de víctimas potenciales a las que no se aplicará la respuesta.
- ¿Se aplicará la respuesta a un conjunto de delincuentes potenciales identificables? En caso afirmativo, el grupo de control debe ser un grupo muy similar de delincuentes potenciales a los que no se aplicará la respuesta.
- ¿Se aplicará la respuesta a algún otro grupo identificable de personas o cosas? En caso afirmativo, el grupo de control debe ser un grupo muy similar de personas o cosas potenciales a las que no se dará la respuesta.

- Es imposible identificar un grupo de control para esta evaluación. Si se marca esta casilla, vuelva a la Tabla D I y elija el diseño adecuado de la columna I. A continuación, pase a la parte IV.

Si ha respondido Sí a alguna de las preguntas entre la 1 y la 4, compare sistemáticamente las características del grupo de respuesta con las del grupo de control y enumere las principales diferencias. En la última parte de esta lista de verificación (parte V) se le pide que emita un juicio sobre la posibilidad de que otros factores hayan podido causar el cambio en el problema relativo al control. Su lista de diferencias es una lista de otros factores potenciales que podrían explicar el cambio en el problema.

IV. SACAR CONCLUSIONES

Estos puntos corresponden a la fase de evaluación y son aplicables una vez documentados los resultados de la evaluación. Estas preguntas están diseñadas para ayudar a formular conclusiones que sean coherentes con los resultados de sus evaluaciones de proceso y de impacto y con su diseño de evaluación. Tendrá que formular más preguntas que las enumeradas aquí para interpretar plenamente los resultados de su evaluación particular.

A. ¿CUÁLES SON SUS CONCLUSIONES DE LA EVALUACIÓN DE PROCESO?

1. La respuesta no se implementó.
2. La respuesta se implementó de forma radicalmente distinta a la prevista.
3. La respuesta se implementó con recursos inadecuados, una duración demasiado limitada o sin la intensidad necesaria.
4. La respuesta se implementó casi según lo previsto y con los recursos adecuados, durante el tiempo necesario y con la intensidad requerida.

B. ¿CUÁLES SON SUS CONCLUSIONES DE LA EVALUACIÓN DE IMPACTO?

(Seleccione el diseño o diseños aplicados: pre-post, pre-post con control, series temporales o series temporales múltiples. Si se utilizaron varios diseños, interprete las Tablas 2 y 3 para cada diseño por separado).

Diseño pre-post (sin control) — utilice la Tabla 2 para interpretar su evaluación.

1. El problema empeoró después de la respuesta.
2. Problema sin cambios después de la respuesta.
3. Problema disminuido tras la respuesta.

Diseño pre-post (con control) — utilice la Tabla 3 para interpretar su evaluación.

1. El problema con la respuesta empeoró, en relación con el control.
2. Problema con la respuesta se mantuvo sin cambios, en relación con el control.
3. Problema con la respuesta disminuyó, en relación con el control.

Diseño de series temporales (sin control) — utilice la Tabla 3 para interpretar su evaluación.

1. El problema empeoró después de la respuesta.
2. Problema sin cambios después de la respuesta.
3. El problema se redujo después de la respuesta.

Diseño de series temporales múltiples (con control) — utilice la Tabla 3 para interpretar su evaluación.

1. El problema con la respuesta empeoró, en relación con el control.
2. Problema con la respuesta se mantuvo sin cambios, en relación con el control.
3. Problema con la respuesta disminuyó, en relación con el control.

V. CONCLUSIONES GENERALES DE LA EVALUACIÓN DE IMPACTO

1. ¿Las personas o lugares que recibieron la respuesta experimentaron menos el problema que las personas o lugares que no la recibieron?
2. ¿Disminuyó el problema a un ritmo más rápido después de la respuesta que antes de ella? Para responder a esta pregunta, se requiere un diseño de series temporales (con o sin grupo de control). No marque la casilla si sólo se utilizó un diseño pre-post con o sin grupo de control.
3. ¿Puede eliminar todas las demás explicaciones plausibles del cambio en el problema, aparte de que la respuesta hizo que el problema disminuyera? Utilice la lista de diferencias (véase la parte inferior de la parte III) entre los grupos de respuesta y de control para responder a estas preguntas. No marque si el único diseño utilizado fue un diseño pre-post sin grupo de control.

Se trata de juicios de valor; las respuestas deben considerarse como su grado de confianza en los hallazgos, más que como una evaluación totalmente objetiva de lo ocurrido. Otras personas que examinen las mismas pruebas podrían llegar a conclusiones diferentes. Por esta razón, estas preguntas (y la que sigue) se responden mejor después de que varias personas con perspectivas diferentes hayan examinado la información de la evaluación.

Basándose en sus respuestas a estas tres preguntas:

¿ESTÁ RAZONABLEMENTE SEGURO DE QUE LA RESPUESTA QUE HA EVALUADO HA SIDO LA PRINCIPAL RESPONSABLE DE LOS CAMBIOS EN EL PROBLEMA?

- **SÍ** — Si se ha realizado un examen exhaustivo de estas preguntas y se han marcado *todas* ellas, utilice la Tabla 3 para interpretar sus resultados.
- **NO** — Si no puede marcar todas las casillas de las preguntas 1, 2 y 3, esta es la respuesta adecuada. Debe interpretar la Tabla 3 con extrema precaución. Utilice la Tabla 2 para interpretar sus resultados.

TABLA C2: INTERPRETACIÓN DE LOS RESULTADOS DE LAS EVALUACIONES DEL PROCESO

RESULTADOS DE LA EVALUACIÓN DE PROCESO RESPUESTAS A LA PREGUNTA IV.A.			
4 MARCADA IMPLEMENTADO CASI SEGÚN LO PREVISTO		1, 2 O 3 MARCADA NO SE HA IMPLEMENTADO O SE HA IMPLEMENTADO DE FORMA RADICALMENTE DISTINTA A LA PREVISTA	
RESULTADOS DE LA EVALUACIÓN DE IMPACTO Respuestas a la pregunta IV.B (pre-post sin controles)	3 MARCADA El problema disminuyó	A. La respuesta puede o no haber causado la disminución del problema. No obstante, el problema ha disminuido	C. Sugiere que otros factores podrían haber causado la disminución del problema o que la respuesta fue accidentalmente eficaz. No obstante, el problema ha disminuido
	1 O 2 MARCADA El problema no disminuyó	B. La respuesta no parece haber funcionado; aunque es posible que el problema hubiera empeorado sin ella.	D. No está claro si debe implementarse la respuesta prevista o si debe volver a analizarse el problema y aplicarse una respuesta diferente.
Independientemente de la interpretación (A, B, C o D), no hay pruebas suficientes para relacionar la respuesta con el problema. La información disponible sobre la evaluación de impacto no puede utilizarse para promover el uso de la respuesta en otros problemas similares. Tampoco puede utilizarse para descartar el uso de la respuesta en circunstancias similares.			

TABLA C3:INTERPRETACIÓN DE LOS RESULTADOS DE LAS EVALUACIONES DEL PROCESO Y DE IMPACTO (OTROS DISEÑOS)

RESULTADOS DE LA EVALUACIÓN DE PROCESO RESPUESTAS A LA PREGUNTA IV.A.			
4 MARCADA IMPLEMENTADO CASI SEGÚN LO PREVISTO		1, 2 O 3 MARCADA NO SE HA IMPLEMENTADO O SE HA IMPLEMENTADO DE FORMA RADICALMENTE DISTINTA A LA PREVISTA	
<p>RESULTADOS DE LA EVALUACIÓN DE IMPACTO Respuestas a la pregunta IV.B (pre-post con controles, series temporales o series temporales múltiples)</p>	<p>3 MARCADA El problema disminuyó</p>	<p>A. Hay pruebas de que la respuesta ha causado una disminución del problema. Esta respuesta es una opción potencialmente útil en circunstancias similares.</p>	<p>C. Sugiere que otros factores podrían haber causado la disminución del problema o que la respuesta fue accidentalmente eficaz. No puede recomendar esta respuesta para abordar problemas similares, ya que no sabe si tendría un impacto en caso de utilizarse.</p>
	<p>1 O 2 MARCADA El problema no disminuyó</p>	<p>B. Hay pruebas de que la respuesta fue ineficaz o empeoró las cosas y debe intentarse una respuesta diferente. Debería volver a analizar el problema y rediseñar la respuesta. La respuesta implementada probablemente no debería utilizarse en circunstancias similares.</p>	<p>D. Se aprendió poco; tal vez se habrían observado mejores resultados si la respuesta se hubiera implementado según lo previsto, pero esto es especulativo. Ninguna recomendación, a favor o en contra de la respuesta, es válida.</p>

APÉNDICE D

RESUMEN DE LOS PUNTOS FUERTES Y DÉBILES DE LOS DISEÑOS DE EVALUACIÓN

DISEÑO	FORTALEZAS	DEBILIDADES
COMPARACIÓN ESTÁTICA	<ul style="list-style-type: none"> - Mejor que nada. - Puede ser útil para el examen preliminar. 	<ul style="list-style-type: none"> - Es probable que las diferencias preexistentes entre los grupos sean la causa de los distintos niveles del problema. - No se puede determinar si la respuesta se produjo antes o después de las diferencias preexistentes. - No se puede descartar la posibilidad de que los resultados se deban a tendencias preexistentes.
PRE-POST	<ul style="list-style-type: none"> - Aplicación sencilla y rápida. - Puede utilizarse fácilmente con encuestas. - Puede proporcionar una estimación razonable del cambio en un problema tras una respuesta. 	<ul style="list-style-type: none"> - Sólo puede mostrar cambios a corto plazo en los problemas. - No se pueden tener en cuenta las tendencias preexistentes. - Muy débil a la hora de eliminar explicaciones alternativas para el cambio en el problema. - No puede tener explicar la posibilidad de que algún otro factor se produjera al mismo tiempo que la respuesta y provocara el cambio del problema.

DISEÑO	FORTALEZAS	DEBILIDADES
<p>SERIES TEMPORALES INTERRUMPIDAS</p>	<ul style="list-style-type: none"> - Muy fácil de usar con datos recogidos de forma rutinaria durante muchos periodos de tiempo. - Puede eliminar tendencias preexistentes y muchas otras explicaciones alternativas. 	<ul style="list-style-type: none"> - Muy difícil de utilizar si se utilizan métodos especiales de recopilación de datos, como encuestas, para medir el problema. - No se puede tener en cuenta la posibilidad de que algún otro factor se haya producido al mismo tiempo que la respuesta y haya provocado el cambio del problema. - Los resultados tardan mucho tiempo en establecerse - Difícil de interpretar cuando hay pocos sucesos problemáticos por periodo de tiempo antes de la respuesta.
<p>PRE-POST CON CONTROL</p>	<ul style="list-style-type: none"> - Puede utilizarse fácilmente con encuestas. - Puede explicar la posibilidad de que algún otro factor se haya producido al mismo tiempo que la respuesta y haya provocado el cambio del problema. 	<ul style="list-style-type: none"> - Sólo pueden mostrar cambios a corto plazo en los problemas - Requiere un grupo de control con problemas que no reciba la respuesta y que sea similar al grupo de respuesta.
<p>SERIES TEMPORALES MÚLTIPLES</p>	<ul style="list-style-type: none"> - Fácil de usar con datos recogidos de forma rutinaria durante muchos periodos de tiempo. - Puede eliminar tendencias preexistentes y muchas otras explicaciones alternativas. - Puede explicar la posibilidad de que algún otro factor se haya producido al mismo tiempo que la respuesta y haya provocado el cambio del problema. 	<ul style="list-style-type: none"> - Muy difícil de utilizar si se utilizan métodos especiales de recopilación de datos, como encuestas, para medir el problema. - Requiere un grupo de control problemático que no reciba la respuesta y sea similar al grupo de respuesta. - Los resultados tardan mucho tiempo en establecerse. - Difícil de interpretar cuando hay pocos sucesos problemáticos por cada periodo de tiempo antes de la respuesta.

REFERENCIAS

Clarke, Ronald V. 1992. *Situational Crime Prevention: Successful Case Studies*. Albany, Nueva York: Harrow and Heston.

Clarke, Ronald V., y David Weisburd. 1994. "Diffusion of Crime Control Benefits: Observations on the Reverse of Displacement." En R. V. Clarke (ed.), *Crime Prevention Studies* (Vol. 2, pp. 165-184). Monsey, Nueva York: Criminal Justice Press.

Cornish, Derek, y Ronald V. Clarke. 1986. "Situational Prevention, Displacement of Crime and Rational Choice Theory." En K. Heal y G. Laycock (eds.), *Situational Crime Prevention: From Theory into Practice*. Londres, Reino Unido: Her Majesty's Stationery Office.

Corsaro, Nick, Dan W. Gerard, Robin S. Engel, y John E. Eck. 2012. "Not By Accident: An Analytical Approach to Traffic Crash Harm Reduction." *Journal of Criminal Justice* 40 (6), 502-514.

Eck, John E. 1993. "The Threat of Crime Displacement." *Criminal Justice Abstracts* 25, 527-546.

Eck, John E. 1997. "Preventing Crime at Places." In Lawrence W. Sherman, Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter y Shawn Bushway eds. *Preventing Crime: What Works, What Doesn't, What's Promising – A Report to the Attorney General of the United States*. Washington, DC: United States Department of Justice, Office of Justice Programs.

Eck, John E., y Edward R. Maguire. 2000. "Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence." En A. Blumstein y J. Wallman (eds.), *The Crime Drop in America* (pp. 207-265). Nueva York: Cambridge University Press.

Eck, John E., y William Spelman. 1987. *Problem Solving: Problem Oriented Policing in Newport News*. Washington, D.C.: Police Executive Research Forum.

Hesseling, René B. P. 1995. "Displacement: A Review of the Empirical Literature." En R. V. Clarke (ed.), *Crime Prevention Studies* 3, pp. 197-230. Monsey, Nueva York: Criminal Justice Press.

Lancashire Constabulary. 2006. *Operation SeaQuest*. Presentación al Premio Tilley. Disponible en [www.popcenter.org/library/awards/tilley/2006/06-52\(W\).pdf](http://www.popcenter.org/library/awards/tilley/2006/06-52(W).pdf).

Matthews, Roger. 1992. "Developing More Effective Strategies for Curbing Prostitution." En R. V. Clarke (ed.), *Situational Crime Prevention: Successful Case Studies* (1ª ed., pp. 89-98). Nueva York: Harrow and Heston.

Office of Community Oriented Policing Services. 1998. *Problem-Solving Tips: A Guide to Reducing Crime and Disorder through Problem-solving Partnerships*. Washington, D.C.: U.S. Department of Justice, Office of Community Oriented Policing Services.

Pawson, Ray, y Nick Tilley. 1997. *Realistic Evaluation*. Londres: Sage.

SOBRE LOS AUTORES

JOHN E. ECK

John E. Eck es profesor de la Escuela de Justicia Penal de la Universidad de Cincinnati, donde imparte cursos de posgrado sobre eficacia policial y prevención de la delincuencia. El Dr. Eck obtuvo su doctorado en la Universidad de Maryland en 1994 y su maestría en política pública en la Universidad de Michigan en 1977. De 1977 a 1994, dirigió investigaciones en el Foro de Investigación de Ejecutivos Policiales, en Washington D.C., donde realizó estudios sobre la gestión de las investigaciones criminales y los mercados de la droga, y ayudó a poner a prueba y aplicar la actuación policial orientada a la resolución de problemas en organismos de todo Estados Unidos. De 1995 a 1998, el Dr. Eck fue coordinador de evaluación de la Zona de Alta Intensidad de Tráfico de Drogas de Washington/Baltimore, donde desarrolló métodos para evaluar la eficacia de la represión del tráfico de drogas. El Dr. Eck ha escrito extensamente sobre la policía orientada a la resolución de problemas, la cartografía de la delincuencia, los mercados de la droga y la prevención de la delincuencia en los lugares. Los intereses de investigación del Dr. Eck se centran en los lugares de delincuencia: cómo surgen y qué se puede hacer para reducir la delincuencia en estos lugares. Entre sus numerosas publicaciones, el Dr. Eck es coautor, junto con Ronald V. Clarke, de *Crime Analysis for Problem Solvers: In 60 Small Steps* [Análisis delictivo para encargados de la resolución de problemas: En 60 pequeños pasos]. En los veranos, el Dr. Eck ayuda a su esposa a restaurar cementerios antiguos y reparar lápidas rotas, lo que ha alimentado su interés por evitar el vandalismo en los cementerios.

LISTA DE LECTURAS RECOMENDADAS

Existen numerosos libros y artículos que describen cómo realizar evaluaciones. A continuación, se enumeran algunos que me han resultado útiles para los estudiantes de posgrado, que son lecturas estándar esenciales o que se escribieron específicamente para la policía y están disponibles en internet.

Si desea adquirir conocimientos especializados en el ámbito de la evaluación, deberá seguir cursos de maestría y doctorado. Pero es mucho lo que se puede aprender y aplicar útilmente leyendo algo más que esta guía introductoria.

Bachman, Ronet, y Russell K. Schutt. 2013. *The Practice of Research in Criminology and Criminal Justice*. Thousand Oaks, California: Sage. Este texto de nivel universitario proporciona una descripción bien redactada de la teoría y la práctica de la recopilación de datos, la medición y el diseño de la investigación aplicados a la investigación y la evaluación en el ámbito de la justicia penal.

Campbell, Donald T., y Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Nueva York: Houghton Mifflin. Aunque tiene más de cincuenta años, sigue siendo la "biblia" de los diseños de evaluación. Prácticamente todos los textos sobre métodos adaptan material de esta fuente. Sigue siendo indispensable y, aunque breve y directo, este libro no es de lectura rápida.

Clarke, Ronald V., y John E. Eck. 2005. *Crime Analysis for Problem Solvers: In 60 Small Steps*. Washington, DC: Office of Community Oriented Policing. Disponible en: <http://www.popcenter.org/learning/60steps/>. Contiene varias secciones breves que describen la evaluación de los esfuerzos de resolución de problemas, así como un breve resumen del desplazamiento y la difusión de los beneficios.

Converse, Jean M., y Stanley Presser. 1986. *Survey Questions: Handcrafting the Standardized Questionnaire*. Thousand Oaks, California: Sage. Este libro es una referencia estándar en la investigación con encuestas.

Blair, Johnny, Ronald F. Czaja, y Edward Blair. 2013. *Designing Surveys: A Guide to Decisions and Procedures*. Los Ángeles: Sage. Una buena introducción al diseño de encuestas.

Eck, John E., y Nancy LaVigne. 1993. *Police Guide to Surveying Citizens and Their Environment*. Washington, D.C.: Bureau of Justice Assistance. Número NCJ: 143711. Disponible en: http://www.popcenter.org/library/reading/PDFs/Surveying_Citizens.pdf. Esta monografía describe los aspectos básicos de la realización de encuestas a los ciudadanos y del entorno físico. Contiene varios ejemplos e instrumentos de encuesta.

Eck, John E. 2005. "Evaluations for Lesson Learning." En N. Tilley, ed. *A Handbook for Crime Prevention: Theory, Policy and Practice*. Pp. 699-733. Cullompton, Devon: Willan. Un tratamiento académico del papel de la evaluación en la mejora de la prevención de la delincuencia.

Guerette, Rob T. 2009. *Analyzing Crime Displacement and Diffusion. Problem-Solving Tools Guide N.º 10*. Washington, D.C.: Office of Community Policing Services. Disponible en: <http://www.popcenter.org/tools/displacement/>. Una guía exhaustiva y de fácil lectura sobre el desplazamiento, escrita para profesionales de la policía.

Kosslyn, Stephen M. 1993. *Elements of Graph Design*. Nueva York: W.H. Freeman. Este libro de diseño gráfico, bien organizado, ofrece consejos prácticos y directos sobre cómo crear cuadros, gráficos y figuras eficaces con datos. Está repleto de buenos y malos ejemplos.

Shadish, William R., Thomas D. Cook, y Donald T. Campbell. 2001. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Belmont, California: Wadsworth. Se trata de la continuación, mucho más extensa, de la monografía de Campbell y Stanley. Es densa pero exhaustiva. Si pretende dominar el tema de las evaluaciones, debe consultar este texto estándar.

Trochim, William, y James P. Donnelly. 2007. *The Research Methods Knowledge Base*. Boston, Massachusetts: Cengage Learning. Este texto universitario se diseñó para su uso en línea, pero está disponible en versión de bolsillo. Es muy práctico y muestra cómo crear diseños de evaluación complejos a partir de diseños más sencillos para abordar situaciones peculiares. También contiene un excelente análisis de la medición y el muestreo.

Weisburd, David, y Chester Britt. 2014. *Statistics in Criminal Justice*. Nueva York: Springer. Se trata de un texto universitario de introducción a la estadística muy bien escrito, que lleva al lector desde lo más básico hasta un nivel intermedio.

Weisel, Deborah. 1999. Conducting Community Surveys: A Practical Guide for Law Enforcement Agencies. Washington, D.C.: Bureau of Justice Statistics and Office of Community Oriented Policing. (October) N.º NCJ:178246. Disponible en: http://www.popcenter.org/library/reading/PDFs/Conducting_Surveys.pdf. Esta guía práctica para los organismos encargados de hacer cumplir la ley acompaña al software de encuestas sobre victimización desarrollado por la Oficina de Servicios Policiales Orientados a la Comunidad y la Oficina de Estadísticas de Justicia. Describe cómo se han utilizado las encuestas para mejorar los servicios policiales, las formas de identificar los objetivos de las encuestas y los procedimientos para su administración y análisis.

NOTAS FINALES

¹ Pawson y Tilley (1997)

² Eck y Spelman (1987); Office of Community Oriented Policing Services (1998).

³ Lancashire Constabulary (2006).

⁴ Matthews (1992).

⁵ Eck (1997).

⁶ Clarke (1992).

⁷ Cornish y Clarke (1986); Eck (1993); Hesseling (1995).

⁸ Clarke y Weisburd (1994).

⁹ Corsaro et al. (2012).



**REVISANDO LAS
RESPUESTAS A
LOS PROBLEMAS:
¿FUNCIONÓ? UNA INTRODUCCIÓN
PARA SOLUCIONADORES
DE PROBLEMAS POLICIALES,
2DA. EDICIÓN**

AGOSTO 2023