



BANCO MUNDIAL

1818 H Street, N.W.
Washington, D.C. 20433, EE.UU.
Teléfono: 202-477-1234
Fax: 202-477-6391
Télex: MCI 64145 WORLDBANK
MCI 248423 WORLDBANK
Internet: www.worldbank.org

Grupo de Evaluación Independiente
Grupo de programas de gestión de conocimientos
y desarrollo de la capacidad de evaluación
Correo electrónico: eline@worldbank.org
Teléfono: 202-458-4497
Fax: 202-522-3125

GRUPO DE EVALUACIÓN INDEPENDIENTE DEL BANCO MUNDIAL Y GRUPO TEMÁTICO PARA EL ANÁLISIS,
SEGUIMIENTO Y EVALUACIÓN DEL IMPACTO EN LA POBREZA • DESARROLLO DE LA CAPACIDAD DE EVALUACIÓN

IEG
GRUPO DE EVALUACIÓN INDEPENDIENTE

REALIZACIÓN DE EVALUACIONES DE IMPACTO DE CALIDAD CON LIMITACIONES DE PRESUPUESTO, TIEMPO E INFORMACIÓN



REALIZACIÓN DE EVALUACIONES DE IMPACTO DE CALIDAD CON LIMITACIONES DE PRESUPUESTO, TIEMPO E INFORMACIÓN

IEG
ECD

Grupo de Evaluación Independiente
Banco Mundial

Grupo temático para el análisis, seguimiento
y evaluación del impacto en la pobreza
Red PREM
Banco Mundial



Nota de agradecimiento

El presente folleto ha sido preparado por Michael Bamberger y copatrocinado por el Grupo de Evaluación Independiente (IEG) y el Grupo temático para el análisis, seguimiento y evaluación del impacto de la pobreza (Red PREM) del Banco Mundial. Está basado en los trabajos del IEG (como se lo conoce por sus siglas en inglés) —antes llamado Departamento de Evaluación de Operaciones—, así como en un documento mucho más detallado elaborado posteriormente por Michael Bamberger, Jim Rugh y Linda Mabry (2006). Diversos miembros del personal del Banco, como Judy Baker, Tara Bedi, Ariel Fiszbein, Emanuela Galasso y Howard White, han contribuido con valiosas observaciones sobre las versiones preliminares de este folleto. El proyecto ha sido dirigido conjuntamente por Keith Mackay, Aline Coudouel y Markus Goldstein.

Agradecemos al Organismo Noruego de Cooperación para el Desarrollo del Ministerio de Relaciones Exteriores por el generoso apoyo financiero prestado a esta publicación.

Klaus Tilmes

Jefe

Grupo de programas de gestión de conocimientos y desarrollo
de la capacidad de evaluación

Grupo de Evaluación Independiente

© 2006

Banco Internacional de Reconstrucción
y Fomento/Banco Mundial
1818 H Street, N.W.
Washington, D.C. 20433, EE.UU.

Derechos reservados.

Impreso en los Estados Unidos de América.

Las opiniones expresadas en este informe no reflejan necesariamente las del Banco Mundial o de los gobiernos de sus países miembros. El Banco Mundial no garantiza la exactitud de los datos que

ÍNDICE

1	Panorama general.....	1
2	Simplificación del diseño de la evaluación.....	5
	Estudio de caso 1. Reducción de los costos de evaluación mediante un diseño transversal posterior al proyecto; impactos de género y empleo del tiempo en el sector de las flores cortadas de Ecuador	9
3	Trabajo con diseños de grupos de comparación.....	10
	Estudio de caso 2. Utilización de las puntuaciones de la propensión para seleccionar un grupo de comparación compatible: Proyecto de caminos rurales de Viet Nam.....	12
4	Utilización de datos secundarios.....	13
	Estudio de caso 3. Utilización de fuentes múltiples de datos secundarios: Proyecto de nutrición integrada de Bangladesh.....	15
5	Reconstrucción de datos de referencia	16
	Estudio de caso 4. Reconstrucción de datos de referencia: el Fondo de Inversión Social de Emergencia de Nicaragua	18
6	Reducción de los costos de la recopilación de datos	19
	Estudio de caso 5. Utilización de técnicas de evaluación participativa rápida (EPR) para reducir el costo de la recopilación de datos: Flores, Proyecto de suministro de agua y saneamiento a aldeas en Indonesia	21
7	Fortalecimiento de los diseños de evaluación cuando se trabaja con limitaciones de presupuesto, tiempo e información	22
	Cuadro 1. Diseño de un grupo de control anterior y posterior a la prueba	8
	Cuadro 2. Repercusiones de las limitaciones de presupuesto, tiempo e información en la calidad de una evaluación de impacto	23
	Notas	27
	Referencias	29
	Recursos adicionales sobre vigilancia y evaluación.....	31

1. Panorama general

Contexto y objetivo de esta publicación

En la comunidad del desarrollo se reconoce cada vez más claramente la utilidad de realizar evaluaciones. La evaluación es un poderoso instrumento para saber qué es lo que funciona, lo que no, y por qué. Es, asimismo, un importante vehículo de rendición de cuentas. De ahí el interés del Banco Mundial en la evaluación de todos sus proyectos. Un número cada vez mayor de países en desarrollo reconoce, además, los beneficios que se derivan de la evaluación y muchos de ellos se esfuerzan por institucionalizar los sistemas de seguimiento y evaluación como parte de una gestión de gobierno sólida¹.

En este contexto, se está realizando un número cada vez mayor de rigurosas evaluaciones de impacto, y es importante que los métodos, conclusiones y recomendaciones de éstas sean lo más fiables posible. Por otro lado, estas evaluaciones pueden ser costosas. Los jefes de proyectos y de programas que desean llevar a cabo una evaluación suelen encontrarse con importantes limitaciones de presupuesto, tiempo o información, lo cual puede desincentivar la realización de evaluaciones rigurosas. El objetivo del presente folleto es brindar asesoramiento a las personas que se ocupan de la planificación de las evaluaciones de impacto para que estén en condiciones de elegir los métodos más rigurosos disponibles en función de las limitaciones existentes. Asimismo, se pretende aclarar las necesarias soluciones de compromiso entre el rigor, por un lado, y el presupuesto, el tiempo y la información disponibles para una evaluación, por el otro². Se espera que este folleto anime al personal directivo a realizar evaluaciones de impacto en situaciones en que habrían considerado demasiado costoso o lento emprender una evaluación de alta calidad. Así pues, el resultado deseado es que aumenten la calidad y la cantidad de las evaluaciones de impacto rigurosas realizadas.

Existen numerosas publicaciones sobre metodologías adecuadas para la evaluación de los impactos de los proyectos y programas de desarrollo. En este folleto se aplican esas metodologías a situaciones y restricciones reales con que se encuentran algunos jefes de proyecto e investigadores. Nuestra intención es completar otras publicaciones recientes del Banco Mundial, como las de Baker (2000), el Departamento de Evaluación de Operaciones (2004), Ravallion (2001, 2005) y White (2006), así como las directivas metodológicas y los estudios de casos prácticos de evaluaciones de impacto divulgados a través del sitio web del Banco *Poverty Impact Analysis, Monitoring and Evaluation*³.

Circunstancias reales que obstaculizan las evaluaciones

Existen dos grupos de limitaciones que determinan la elección de los métodos de evaluación. La primera abarca las limitaciones de presupuesto, tiempo e información. Las **limitaciones presupuestarias** influyen en el número de entrevistas que pueden realizarse, la capacidad de combinar la recopilación y el análisis de datos cualitativos y cuantitativos, el tipo de experiencia profesional del equipo de investigación y el análisis que puede llevarse a cabo. Las **limitaciones temporales** tienen que ver con el momento en que se inicia y termina la evaluación, el tiempo que pueden pasar los investigadores sobre el terreno y el tiempo disponible para recibir información de las partes interesadas. Cuando se realizan nuevas encuestas, las **limitaciones de**

carácter informativo influyen en la capacidad de obtener datos de un grupo de comparación adecuado e información de referencia sobre la población de beneficiarios, o de reunir información de carácter confidencial y entrevistar a grupos de difícil acceso. Cuando la evaluación se basa en datos secundarios o cuando los datos proceden de estudios realizados por otros organismos (estudios concatenados o sincronizados), las limitaciones de información pueden mermar la compatibilidad de la cobertura o cronología de la muestra o la posibilidad de que los datos cubran las variables necesarias o las definan adecuadamente.

En cambio, *los mecanismos de diseño y ejecución de los programas* determinan el modo de seleccionar a los participantes en el proyecto (aleatoriamente, según criterios administrativos o mediante la autoselección) y la calidad y uniformidad de la ejecución del proyecto. Estas características del diseño, que se derivan de la naturaleza misma de un proyecto, suscitan toda una serie de cuestiones relacionadas específicamente con el proyecto que determinan el alcance de los métodos de evaluación que se pueden aplicar. Estas cuestiones no se abordarán aquí; puede encontrarse más información sobre cómo tratar estas cuestiones de los métodos de evaluación en Ravallion (2001, 2005).

Marco hipotético de evaluaciones reales

Examinaremos dos marcos hipotéticos comunes. En el primero, el evaluador participa en el proyecto desde el principio, pero los problemas de presupuesto, tiempo e información o los mecanismos de diseño y ejecución del programa restringen el abanico de diseños de evaluación que pueden utilizarse. Por ejemplo, tal vez sea imposible incluir un grupo de control o realizar un estudio de referencia exhaustivo sobre la población beneficiaria, o tal vez sólo pueda realizarse un número limitado de entrevistas.

En el segundo marco hipotético, la evaluación no se encarga hasta que el proyecto está a punto de finalizar o ha terminado. La información puede obtenerse por uno de los cuatro medios siguientes (véase White, 2006): mediante una encuesta específica para el proyecto; agregando un módulo especial a una encuesta en curso; mediante una encuesta sincronizada en la que se entrevista a la población beneficiaria, pero tomando el grupo de control de otra encuesta (encuesta nacional de hogares, etcétera), o basando la evaluación en datos secundarios, como la información sobre el proyecto y los grupos potenciales de comparación. Una importante limitación de todos los diseños de evaluación posteriores a la ejecución del proyecto es la falta de estudios de referencia. Las opciones para abordar esta cuestión se examinan en la Sección 5.

Condiciones para una evaluación de impacto de calidad con limitaciones reales

El reto para el evaluador y el cliente consiste en decidir si es posible realizar una evaluación de impacto de calidad con las limitaciones existentes en la práctica, y elegir el diseño más sólido posible en función de una serie concreta de restricciones de presupuesto, tiempo e información. Por ejemplo, ¿a partir de qué punto se considera que la muestra es demasiado pequeña, o su cobertura demasiado limitada, para que el análisis estadístico sea válido? ¿Cuáles son los criterios para evaluar si los datos secundarios son adecuados para estimar las condiciones iniciales? ¿Y cuándo es posible formular una hipótesis válida a falta de un estudio de referencia?

Una evaluación de impacto de calidad debe:

- Elaborar una serie de indicadores que puedan definir y cuantificar de forma válida y fiable los insumos, procesos de ejecución, productos, resultados esperados e impactos del proyecto;
- Establecer una *hipótesis* lógicamente válida que demuestre en forma convincente que los cambios observados en los indicadores de resultados a raíz del proyecto se deben realmente al proyecto y no a otros factores ajenos al mismo, como las mejoras en la economía local o programas organizados por otros organismos;
- Determinar, de conformidad con procedimientos estadísticos aceptados, si un proyecto ha contribuido a los impactos esperados y ayudado a una proporción significativa de la población beneficiaria.

Además, muchas evaluaciones deben:

- Evaluar la distribución de los beneficios entre los distintos sectores de la población beneficiaria;
- Identificar los factores que influyen en la magnitud y la distribución de los impactos;
- Evaluar la sostenibilidad de los impactos a lo largo del tiempo.

Cuando no hay problema de falta de recursos, en las evaluaciones convencionales se suele incluir un grupo de beneficiarios y un grupo de comparación antes y después de la intervención, en la forma que puede verse a continuación:

Tiempo	T ₁ Inicio del proyecto (referencia)	T ₂ Intervención del proyecto (puede durar unas pocas semanas o muchos años)	T ₃ Fin del proyecto
Grupo de beneficiarios	P ₁	X	P ₂
Grupo de control aleatorio o no aleatorio	C ₁		C ₂

La comparación de un grupo de control antes y después de la prueba representa la **hipótesis**, lo que habría sucedido a la población de beneficiarios si el proyecto no se hubiera llevado a cabo.

Los diseños de evaluaciones en que los sujetos se asignan aleatoriamente al proyecto y a los grupos de control presentan varias ventajas metodológicas, ya que se evitan diferencias sistemáticas entre los dos grupos antes del proyecto. Sin embargo, en muchas situaciones reales la asignación aleatoria no es posible, de modo que se buscará la correspondencia más próxima posible entre los dos grupos utilizando procedimientos como el “emparejamiento” en función de *características observables o puntuaciones de la propensión* (véase la Sección 3). Se encuesta a los dos grupos tanto al inicio del proyecto (T₁) como después de su ejecución (T₃). Si la correspondencia entre los grupos es buena, cualquier diferencia estadísticamente significativa entre ambos respecto de las variables de impacto en T₃ indicará un efecto potencial del proyecto. Sin embargo, las diferencias podrían explicarse también, en parte, por una experiencia distinta entre los dos grupos durante el período de ejecución del proyecto (por ejemplo, un proyecto no relacionado que afecta solamente al grupo de control).

Dos elementos importantes del diseño para la estimación de los impactos de los proyectos son la existencia de un grupo de control (o comparación) cuidadosamente seleccionado y la comparación de ambos grupos antes y después de la intervención. El primer elemento es esencial para la formulación de una hipótesis lógicamente sólida. El segundo es aconsejable, pero sólo es posible utilizar datos transversales relativos al grupo de beneficiarios y al grupo de control después del proyecto. Si el conjunto de datos de corte transversal tiene las características observables utilizadas para asignar el programa a los beneficiarios, y creemos que las características no observables no desempeñan un papel importante en la participación en el programa, el emparejamiento por puntuación de la propensión puede ser una opción aceptable y menos costosa. Aunque una de las limitaciones de la puntuación de propensión es que no es posible controlar todas las diferencias preexistentes entre los dos grupos, suele ser la mejor opción cuando no se dispone de datos de referencia, y también es mucho más barata. En términos prácticos, cuanto mejor se entienda el contexto del proyecto y el proceso de selección de participantes, mayor será la confianza en la validez de la correspondencia por puntuación de propensión.

Muchos de los marcos hipotéticos reales examinados en las siguientes secciones se refieren a situaciones en que se ha eliminado uno o varios de estos cuatro puntos de observación (P_1 , P_2 , C_1 , y C_2), sea como estrategia deliberada para reducir los costos y el tiempo invertido, o porque las circunstancias no posibilitaron reunir datos en todos los puntos (por ejemplo, cuando la evaluación no comenzó hasta que el proyecto estaba demasiado avanzado o cuando, por falta de presupuesto, no fue posible recopilar datos de referencia sobre un grupo de control). Una de las cuestiones capitales que abordaremos es la siguiente: ¿Cómo influyen los enfoques de evaluación limitados por los costos y el tiempo en la validez del diseño de la evaluación y en las conclusiones extraídas de su análisis? ¿Cuáles son los compromisos aceptables que permiten llegar a conclusiones válidas y cuáles las condiciones metodológicas mínimas sin las cuales no puede seguir considerándose el estudio como una evaluación de impacto de calidad?

En la Sección 2 se analiza una variedad de alternativas para simplificar el diseño de una evaluación y se examina su influencia en la rigurosidad y el costo de las evaluaciones. En la Sección 3 se describen opciones para seleccionar los grupos de comparación y en la Sección 4 se considera la utilización de datos secundarios. Las estrategias para reconstruir los datos de referencia y mejorar la rigurosidad de una evaluación se presentan en la Sección 5, mientras que las alternativas para reducir los costos de la recopilación de datos se indican en la Sección 6. Por último, en la Sección 7 se señalan las formas en que las limitaciones de presupuesto, tiempo e información reducen la rigurosidad de una evaluación. También se presenta una serie de opciones para abordar cada una de estas limitaciones.

2. Simplificación del diseño de la evaluación

Diseños aleatorios en contextos reales

En los diseños de evaluación aleatorios, los sujetos (personas, comunidades, centros docentes o sanitarios, etcétera) se asignan al azar al grupo de beneficiarios y al grupo de control, asegurándose de que los dos grupos tienen la misma distribución de características observadas y no observadas al comienzo del proyecto. Ello garantiza que las diferencias de puntuación del impacto después de la intervención no se deban a diferencias iniciales (sesgo de selección) en las características de los dos grupos. A pesar de los beneficios potenciales, la asignación aleatoria sólo se ha usado en un reducido porcentaje de evaluaciones de desarrollo por una de las siguientes razones: las comunidades, organizaciones o personas beneficiarias se seleccionan en función de determinados criterios administrativos o políticos (los más pobres o necesitados, o los lugares donde es más probable que el proyecto tenga éxito); los individuos sujetos se autoseleccionan (personas o grupos que toman ellos mismos la decisión de participar), o situaciones en que la selección aleatoria no es posible por consideraciones de índole política. En estos dos marcos hipotéticos de selección —aleatorio y no aleatorio— es probable que existan diferencias sistemáticas entre el grupo de beneficiarios y el de no beneficiarios con respecto a los factores que determinan el impacto. Por ejemplo, las personas que solicitan pequeños préstamos empresariales a bancos comunitarios pueden tener ya confianza o experiencia para poner en marcha un negocio rentable.

Tal vez exista un número significativo de casos en que la asignación al azar no sólo es posible, sino que es la mejor manera de proceder: por ejemplo, los proyectos piloto, los proyectos de recursos limitados en relación con la demanda, o cuando se desconoce la eficacia del tratamiento. Sin embargo, por todas las razones mencionadas más arriba, la asignación aleatoria no es factible en la mayoría de las intervenciones de desarrollo y, dadas estas limitaciones, en la mayor parte las observaciones siguientes se hace referencia a casos en que un programa no se administra aleatoriamente y en que no es posible realizar diseños de evaluación aleatorios.

Puntuación de la ganancia

Diferencia en la puntuación del indicador de impacto (*diferencia simple*) antes y después del proyecto. Cuando se usa un grupo de control, la *puntuación de la ganancia* es la diferencia en el cambio antes y después del proyecto para los grupos de beneficiarios y de comparación (*diferencia doble*).

Factores que explican las diferencias entre el grupo de beneficiarios y el grupo de control en diseños de evaluación aleatorios y no aleatorios

Los diseños de evaluación en los que debe emplearse un procedimiento de muestreo diferente para el grupo que no participa en la intervención se denominan “diseños de *grupo de control no equivalente*” o “*grupo de comparación*”. Técnicamente, el término “*grupo de control*” sólo debería utilizarse cuando la asignación es aleatoria. Los factores que pueden contribuir a las diferencias *ex-post* en los indicadores de impacto de los proyectos entre grupos de beneficiarios y no beneficiarios (*puntuaciones de la ganancia*) en diseños verdaderamente experimentales, diseños aleatorios en entornos sobre el terreno y diseños de grupo de control no equivalente pueden resumirse como sigue:

- En diseños verdaderamente experimentales, las condiciones de los dos grupos se controlan cuidadosamente durante el tratamiento, y las puntuaciones de la ganancia deberían ser atribuibles exclusivamente a los efectos del proyecto.

- Cuando se trata de diseños aleatorios en situaciones reales, muy pocas veces se puede controlar el entorno del proyecto. Por lo tanto, las puntuaciones de la ganancia pueden guardar relación con *efectos diferenciales de variación con el tiempo* tales como los efectos de contaminación o secundarios que repercuten en el grupo de control, los modelos de atribución distintos y la influencia de la selección o la no selección en el comportamiento de los individuos y las partes interesadas⁴. Por ejemplo, en una evaluación del impacto de los rotafolios en el rendimiento académico en las escuelas de Kenya se comprobó que las asociaciones de padres y personal docente eran más activas en las escuelas que recibían ese material y que los padres animaban más a los hijos a estudiar, de modo que algunos de los cambios en el rendimiento académico tal vez no estuviesen directamente relacionados con la utilidad práctica de los rotafolios. En otros casos, la existencia de un proyecto sufragado por donantes puede incitar a los organismos públicos a prestar más apoyo (no previsto en el diseño del proyecto), ya que quieren que el proyecto sea fructífero, mientras que en otros casos puede ocurrir que las colectividades o las organizaciones no seleccionadas se desmoralicen o que la ayuda estatal sea menor.
- En lo que concierne a los diseños de grupos de control no equivalentes, las puntuaciones de la ganancia pueden verse influidas por un sesgo en el muestreo con respecto a las características recogidas en la encuesta (*observables*) — pueden controlarse en el análisis — y a características no recogidas en la encuesta (variables omitidas o *no observables*) que no pueden controlarse en el análisis. Mientras que algunas de las variables omitidas podrían incluirse con facilidad en encuestas futuras, sería muy difícil recoger otras, como la motivación o la inteligencia.

Reducción de costos y tiempo mediante la simplificación del diseño de la evaluación

Dado que la recopilación de datos suele representar más de la mitad del presupuesto de una evaluación (Baker, 2000), en esta sección abordaremos el modo de reducir los costos asociados al acopio de información simplificando el diseño de la evaluación. En el Cuadro 1 se estima que, en las evaluaciones que requieren la obtención de datos mediante encuestas y otros tipos de información primaria (en lugar del análisis de datos de encuestas existentes y otras fuentes de datos secundarios), algunos de los diseños simplificados de evaluación permiten reducir el costo de la recopilación de información en más del 50%. Sin embargo, hay que hacer importantes compromisos entre las estrategias para economizar fondos, la calidad de la evaluación y la validez de las conclusiones. De hecho, algunos de los diseños menos costosos (los Diseños 5 y 6 expuestos más adelante) han dejado de considerarse evaluaciones de impacto de calidad, aunque proporcionan conclusiones útiles para las operaciones.

Diseños robustos y menos robustos en situaciones reales

Utilizaremos como punto de referencia el *diseño robusto de evaluación de impacto de uso general* descrito en el Cuadro 1 (Diseño 1). Todos los diseños simplificados descritos en esta sección eliminan uno o más de los siguientes puntos de observación: el grupo de comparación de referencia (antes de la intervención), el grupo de beneficiarios de referencia (tratamiento) o el grupo de comparación después de la intervención. No es posible eliminar el grupo de beneficiarios posterior a la intervención, ya que es imprescindible para cuantificar los efectos conseguidos. Las cinco opciones de diseño simplificado son (véase el Cuadro 1):

- Diseño 2: *Diseño aplazado de grupo de comparación antes y después de la intervención*. Es similar al Diseño 1, pero la evaluación no empieza hasta que el proyecto lleva cierto tiempo

en marcha, normalmente como parte del examen a mitad del período. Si el proyecto se retrasa, este diseño no será mucho más débil que el Diseño 1, pero lo será si empieza puntualmente. Se obtendrá una reducción de costos moderada porque se contratará al consultor por menos tiempo.

- **Diseño 3: Grupo de beneficiarios antes y después de la intervención y grupo de comparación después de la intervención.** En este diseño no existe un grupo de comparación antes de la intervención. Aunque es menos robusto metodológicamente que el Diseño 1, puede asegurar un poder analítico razonable. Se consigue una disminución potencial del 25% en los costos asociados a la recopilación de datos debido a la eliminación del grupo de comparación de referencia.
- **Diseño 4: Grupo de beneficiarios y grupo de comparación posteriores a la intervención, sin datos de referencia.** Este diseño, ampliamente utilizado, define el grupo de comparación posterior a la intervención como hipótesis, suponiendo que cualquier diferencia observada entre los dos grupos (tras controlar las características observables de los hogares) se debe a los efectos del proyecto y no a diferencias no observables entre ambos grupos antes de la intervención (véase el Estudio de caso 1, más adelante). La eliminación de todas las encuestas de referencia puede propiciar un ahorro de hasta el 50% en el acopio de datos.

Dos diseños habituales de evaluación que no pueden considerarse como diseños de evaluación de *impacto sólidos*

Los dos diseños siguientes no son diseños de evaluación de impacto sólidos (véase la Sección 1), pero los hemos incluido debido a su gran aceptación y porque, si se emplean con las debidas precauciones, pueden aportar información sobre algunos de los efectos del proyecto.

- **Diseño 5: Comparación del grupo de beneficiarios antes y después de la intervención (comparación reflexiva).** Este diseño elimina el grupo de comparación y, por consiguiente, no proporciona una hipótesis fundada lógicamente. El diseño determina los impactos del proyecto sólo con la premisa muy potente (y por lo general improbable) de que no se producen cambios asociados al paso del tiempo. Este diseño se utiliza con frecuencia porque la eliminación del grupo de comparación permite recortar los gastos de acopio de datos a la mitad y porque existen muchas situaciones en que se dispone de datos para el grupo de beneficiarios (por lo general, procedentes de encuestas para el proyecto y de registros administrativos), pero no para el grupo de comparación.
- **Diseño 6: Grupo de beneficiarios después de la intervención sin datos de referencia ni grupo de comparación.** Este es el diseño más débil de todos y, aunque su uso está muy extendido para calcular los efectos de un proyecto⁵, no ofrece estimaciones cuantitativas rigurosas del impacto del mismo. Muchas veces, este diseño se usa cuando el evaluador tiene un presupuesto muy limitado (en ocasiones, apenas US\$10.000) y sólo puede pasar unas semanas sobre el terreno. Las estimaciones del cambio (el impacto) se basan en una combinación de datos cualitativos, como los cuestionarios recordatorios, los informantes clave, los grupos focales y las técnicas de grupos participativos como la evaluación participativa rápida (EPR), los registros de proyectos y los datos secundarios de organismos públicos (por ejemplo, el Ministerio de Salud o de Educación), censos y otros datos del gobierno. Los datos secundarios se utilizan únicamente para realizar comparaciones globales de comunidades participantes en el proyecto con áreas semejantes, y no para análisis de hogares. Según la naturaleza del diseño, los costos pueden reducirse un 75% o más.

Cuadro 1. Opciones de diseño para reducir los costos de la obtención de datos ^(a)						
Diseño de la evaluación	Referencia	Tratamiento	Mitad de período	Después de la intervención	Porcentaje de ahorro con respecto al Diseño 1	Observaciones
Diseño de referencia: diseño robusto de evaluación de impacto de uso general						
1. Diseño con grupo de beneficiarios y de control antes y después de la intervención con asignación aleatoria y no aleatoria.	P ₁ C ₁	X		P ₂ C ₂	0	Diseño más sólido de todos en la mayoría de los casos reales en que la evaluación comienza al mismo tiempo que el proyecto.
Diseños progresivamente menos robustos						
2. Diseño aplazado de grupo de comparación antes y después de la intervención. La evaluación no comienza hasta mediados del período de ejecución.		X	P ₁ C ₁	P ₂ C ₂	0	Diseño útil cuando la evaluación comienza a mediados del período. Si la ejecución se retrasa, no será mucho más débil que el Diseño 1. Eventual ahorro moderado porque el período del control de evaluación es más breve.
3. Grupo de beneficiarios antes y después de la intervención y grupo de comparación sólo después de ésta.	P ₁	X		P ₂ C ₂	25	Aunque es más débil que los diseños anteriores, es relativamente sólido, ya que permite hacer comparaciones a lo largo del tiempo y análisis transversales con posterioridad al proyecto.
4. Grupo de beneficiarios y de comparación posteriores a la intervención sin datos de referencia.				P ₁ C ₁	50	Diseño frecuente cuando la evaluación comienza hacia el final del ciclo del proyecto o cuando éste ha terminado. La falta de datos de referencia dificulta el control de las diferencias iniciales entre los dos grupos, pero eso se puede resolver en parte con puntuaciones de propensión. Sin embargo, el uso de puntuaciones de propensión puede requerir muestras más grandes.
Dos diseños muy utilizados que, por falta de una hipótesis aceptable, casi nunca son considerados como diseños de impacto de calidad						
5. Análisis del grupo de beneficiarios antes y después de la intervención sin grupo de comparación.	P ₁			P ₂	50	Diseño frecuente cuando sólo se compilan datos sobre el grupo de beneficiarios. Metodológicamente débil, ya que el uso de P1 como hipótesis requiere suposiciones audaces sobre la variación de los efectos con el paso del tiempo y las variables individuales no observables.
6. Grupo de beneficiarios después de la intervención sin datos de referencia ni grupo de comparación.				P ₁	75-90	Es el diseño más débil de todos, pero se utiliza con frecuencia cuando hay que realizar evaluaciones hacia el final del proyecto y con poco tiempo y recursos. Se emplean métodos cualitativos, los registros del proyecto y datos secundarios agregados para estimar la hipótesis.
<p>Símbolos: P = grupo de beneficiarios de selección aleatoria o no aleatoria. C = grupo de control o comparación de selección aleatoria o no aleatoria (no equivalente). P₁, P₂, C₁, C₂ indican la primera y la segunda utilización del instrumento de encuesta con los grupos de beneficiarios y de comparación, respectivamente.</p> <p>^(a) La reducción de costos se basa en la hipótesis de que se necesitarán encuestas u otros métodos de obtención de datos (antes y después de la intervención, y grupo de beneficiarios y de recopilación de datos, en comparación con el Diseño 1, dependen del número de puntos de obtención de datos) eliminados. En el caso de encuestas concatenadas o sincronizadas, o cuando la evaluación se basa en datos secundarios, el ahorro puede ser mayor.</p>						

Estudio de caso 1. Reducción de los costos de evaluación mediante un diseño transversal posterior al proyecto; impactos de género y empleo del tiempo en el sector de las flores cortadas de Ecuador

La evaluación ilustra cómo pueden reducirse considerablemente los costos asociados a la obtención de datos mediante un diseño transversal posterior al proyecto (Diseño 4) en que los datos se recopilan sólo una vez, cuando la intervención ya está en marcha. El punto débil del diseño es que no determina en qué medida los aparentes impactos del proyecto se deben, de hecho, a diferencias preexistentes entre los grupos que no se han reflejado en la encuesta (no observables).

El objetivo del estudio era evaluar el impacto del empleo femenino en la asignación del trabajo remunerado y no remunerado en el hogar. El estudio compara la asignación del trabajo familiar en las zonas geográficas con acceso al sector de las flores cortadas, que ofrece a las mujeres mejores salarios, con la situación en zonas similares sin acceso a esa fuente de empleo. El estudio utiliza un diseño transversal posterior a la intervención en el que se realizaron entrevistas socioeconómicas en mayo y junio de 1999 a una muestra de 562 hogares y se obtuvieron observaciones concernientes a 2.567 personas de más de 10 años. La muestra incluía hogares de “tratamiento” que vivían en el valle donde funcionaba el sector de las flores y hogares de “control” que vivían en un valle parecido a 200 km de distancia, sin acceso a esas actividades. Se obtuvo información pormenorizada sobre el empleo del tiempo en un período de 24 horas y en la semana anterior. Se utilizó el análisis de regresión múltiple, con una variable ficticia para el empleo en el sector de las flores, como control de las características de las familias que podían influir en la variable dependiente (cantidad de tiempo dedicado por hombres y mujeres a las tareas domésticas).

En el estudio se comprobó que, cuando la mujer trabaja, el marido dedica más tiempo a actividades domésticas y que la mayor diferencia se producía cuando la mujer trabajaba en el sector de las flores, donde sus ingresos eran más altos en comparación con los del marido.

Aunque el análisis de regresión mejora la correspondencia entre los hogares beneficiarios y de control después de la intervención, el punto débil de este diseño es que no determina si las diferencias observadas en el empleo del tiempo existían antes de que se introdujera el sector de las flores cortadas. Ello es muy importante para el diseño de la evaluación, ya que quizás uno de los factores determinantes de la ubicación del sector floral fuera el hecho de que las mujeres de esta zona constituirían una parte importante de la mano de obra. El problema podría haberse solventado mediante alguna de las técnicas presentadas en este folleto para la reconstitución de datos de referencia (véase la Sección 5).

Fuente: Newman, 2001.

3. Trabajo con diseños de grupos de comparación

Hay tres maneras de seleccionar a los participantes en un proyecto:

- al azar entre todos los miembros de la población beneficiaria (personas, comunidades, escuelas, etcétera que reúnen las condiciones exigidas);
- mediante el uso de criterios administrativos o políticos (familias más pobres, viabilidad técnica para proporcionar acceso a la infraestructura existente, grupos con más probabilidad de éxito, etcétera), o
- personas que participan por decisión propia (selección libre).

A los efectos de la evaluación de impacto, la selección aleatoria de participantes tiene la ventaja de que se puede suponer que la distribución de las características observadas y no observadas para el grupo de beneficiarios y el grupo de control es análoga, de modo que las puntuaciones de la ganancia tras la intervención no estarán determinadas por diferencias iniciales entre los dos grupos y, por ende, es probable que se deban a la intervención del proyecto.

Sin embargo, por las razones expuestas en la Sección 2, la asignación aleatoria sólo se realiza en un reducido porcentaje de proyectos de desarrollo, por lo que en la mayoría de las evaluaciones de impacto debe utilizarse un **diseño cuasiexperimental**, en el que se emplean distintos procedimientos de muestreo para elegir al grupo de beneficiarios y al de comparación. Esto tiene importantes repercusiones para el análisis de los impactos del proyecto, ya que las puntuaciones de la ganancia posteriores al proyecto pueden deberse a un sesgo de muestreo (diferencias en las características de ambos grupos), y no a los efectos del proyecto. En esta sección se examinan los problemas y enfoques posibles en el uso de diseños cuasiexperimentales para la evaluación de los impactos de los proyectos.

Opciones para seleccionar grupos de comparación cuando existen limitaciones de presupuesto, tiempo e información

Correspondencia de zonas en función de características observables. El investigador, en consulta con los clientes y otras personas bien informadas, establece las características que deben tener las zonas del proyecto y de comparación (por ejemplo, acceso a servicios, tipo o calidad de las viviendas, nivel económico, calidad de las vías de comunicación, tipos de producción agrícola, etcétera). Después, combina la información de mapas (y, a veces, del Sistema de Información Geográfica (GIS) y fotografías aéreas), la observación, información secundaria (censos, encuestas de hogares, registros escolares, etcétera) y los datos aportados por informantes clave para seleccionar las zonas de comparación con las características de correspondencia más próximas. Muchas veces, por imperativos prácticos, es necesario buscar características fácilmente observables o identificables, como el tipo de vivienda o la infraestructura. Aunque de este modo se puede acelerar la labor, es importante recordar la posibilidad de que haya diferencias no observables, a fin de abordarlas en la medida de lo posible mediante investigaciones cualitativas, y tomar los resultados con la debida cautela.

Correspondencia de individuos o familias en función de características observables. Se utilizan procedimientos similares para la correspondencia tanto de individuos como de familias. La selección de la muestra puede basarse, a veces, en encuestas de hogares anteriores o en curso, pero muchas veces los investigadores establecen su propio método de muestreo. En algunos casos, la selección se basa en características físicas observables (tipo de vivienda, distancia de las fuentes de abastecimiento de agua y de otros servicios, tipo de cultivo o extensión del área cultivada, etcétera), mientras que en otros la selección se basa en características que requieren entrevistas

de comprobación previa, como el nivel económico, la actividad laboral o la asistencia a la escuela. En estos últimos casos, el entrevistador debe localizar y entrevistar a un número determinado de individuos con diversas características (muestreo por cuotas).

Muestreo de sujetos en cartera: El grupo de comparación está integrado por personas, hogares o comunidades seleccionados para participar en el proyecto pero que todavía no han participado (véanse Ravallion, 2005 y White, 2006). A menudo hay proyectos de gran envergadura, por ejemplo los relacionados con la construcción de viviendas o de infraestructura comunitaria, que se llevan a cabo por fases a lo largo de varios años y en los que algunos de los beneficiarios no comenzarán a disfrutar de los servicios hasta varios años después de la primera fase. Cuando no hay grandes diferencias entre las características de las familias o colectividades previstas para cada fase, las últimas fases pueden ser un buen grupo de comparación para las primeras. El uso de estos procedimientos también resulta económico. Sin embargo, es necesario examinar detenidamente el diseño y los criterios de selección del proyecto porque, a menudo, habrá diferencias sistemáticas entre las fases. Por ejemplo, puede ocurrir que en la primera fase se incluyan las familias más pobres o las que viven en las zonas mejor comunicadas o en mejor situación, en cuyo caso las características de las colectividades que participen en las últimas fases serán, probablemente, diferentes⁶.

Diseños de discontinuidad de la regresión: En los casos en que se asigna un programa utilizando un umbral claro formado por uno o varios criterios, la pauta de asignación al programa puede emplearse también para la evaluación. La idea básica es comparar las personas, colectividades o unidades que se sitúen justo por encima del umbral y que, por ende, no puedan incluirse en el proyecto (grupo de comparación) con las que se encuentren justo por debajo del umbral y sí puedan incluirse (grupo de tratamiento). Este procedimiento presupone que la norma de tratamiento se aplica realmente y que los criterios de selección no son manipulados por los posibles beneficiarios.

Correspondencia por puntuación de la propensión. Es posible emplear procedimientos de emparejamiento estadístico cuando se dispone de datos secundarios para seleccionar a los individuos, las comunidades o los lugares (escuelas, centros de salud, etcétera) con características análogas a las del grupo beneficiario. El método más común es la **puntuación de la propensión**, donde cada individuo del proyecto se empareja estadísticamente con un grupo de no participantes (vecinos más próximos) atendiendo a una serie de características pertinentes. Se calcula el valor medio de la variable de resultados del vecino más próximo y se compara con la puntuación obtenida para el participante en el proyecto a fin de calcular la *puntuación de la ganancia* (véase en Ravallion, 2006, y Baker, 2000 un resumen de la correspondencia por puntuación de la propensión⁷). Un problema que surge al trabajar con limitaciones presupuestarias es que el uso de este método suele requerir una muestra más grande para garantizar la obtención de variables con la máxima correspondencia.

Diseños de grupos de comparación múltiples. Cuando los proyectos se ejecutan de distintas maneras o los participantes reciben combinaciones de servicios diferentes, se pueden utilizar grupos de comparación distintos para los diferentes tratamientos.

Sesgos y problemas potenciales cuando se utilizan grupos de comparación

Los grupos de comparación están sujetos a diversos sesgos o problemas potenciales. Muchos de ellos pueden agravarse cuando las evaluaciones se realizan con las limitaciones que existen en las situaciones reales. Algunos de los problemas habituales son:

- Los proyectos pretenden abarcar a *todos* los individuos o comunidades con características concretas (familias más pobres, barrios de tugurio más grandes, colectividades, personas con más probabilidades de éxito), lo que dificulta la localización de equivalentes para el grupo de comparación. Cuando existen reglas claras para la selección, es posible utilizar la discontinuidad de la regresión para comparar las zonas participantes en el proyecto con otros lugares. Sin embargo, a veces se producen situaciones en que, por ejemplo, casi todas las familias urbanas más pobres se concentran en un reducido número de barrios de tugurio grandes y la mayoría de las familias con un nivel de vida algo mejor viven en comunidades muy diferentes, por lo que la correspondencia no es muy próxima.
- Cuando las personas o las colectividades deciden participar libremente en el programa, es difícil encontrar los factores que determinan la decisión de participar y, por consiguiente, encontrar una buena comparación. Ello representan un grave problema analítico, ya que las personas o comunidades que deciden participar pueden ser las que tienen más probabilidades de prosperar (por ejemplo, las mujeres que desean participar en programas de desarrollo de pequeños negocios podrían tener experiencia empresarial previa) y, por lo tanto, es posible que los resultados positivos se deban más a las características de los participantes que al proyecto mismo.

El grupo de beneficiarios del proyecto y el grupo de comparación pueden diferir en relación con factores no incluidos en la encuesta (*variables o características no observables omitidas*). En algunos casos, esos factores podrían incluirse fácilmente en estudios futuros, mientras que en otros puede suceder que el evaluador no sea consciente de algunos factores importantes, o que éstos no se puedan cuantificar fácilmente (por ejemplo, los motivos para decidir participar en un proyecto).

En la Sección 7 se describen varias estrategias para abordar problemas frecuentes que afectan a los diseños de grupos de control no equivalentes y la manera de fortalecer los diseños. Se consideran los siguientes temas: los sesgos en la selección muestral, los problemas relacionados con el tamaño de la muestra, los sesgos posteriores a la selección, las incoherencias en la ejecución de los proyectos, la falta de fiabilidad de las medidas de resultados, las influencias del contexto en la ejecución y los resultados del proyecto, los estudios de evaluación rápida y la triangulación.

Estudio de caso 2. Utilización de las puntuaciones de la propensión para seleccionar un grupo de comparación compatible: Proyecto de caminos rurales en Viet Nam

La muestra de la encuesta estaba integrada por 100 comunas participantes y otras 100 no participantes, todas ellas situadas en los mismos distritos. Esta correspondencia geográfica simplificó la logística de la encuesta y redujo los costos; al mismo tiempo, las comunas se encontraban lo bastante alejadas como para que no hubiera “contaminación” (influencia del proyecto en las zonas de control). Se empleó un modelo *logit* de la probabilidad de participar en el proyecto para calcular la puntuación de la propensión para cada comuna participante y no participante en el proyecto. A continuación, se seleccionaron comunas con *puntuaciones de la propensión* análogas a las comunas beneficiarias. Para la evaluación se utilizaron también datos sobre las comunas, recopilados con fines administrativos, en relación con la infraestructura, el empleo, la educación, la asistencia sanitaria, la agricultura y la organización comunitaria. Esta información permitirá realizar un análisis contextual y establecer los indicadores comunitarios de bienestar, así como hacer un seguimiento de los impactos del programa a largo plazo. Los datos administrativos se usarán, asimismo, como modelo para el proceso de selección de participantes y para determinar si existen o no sesgos de selección.

Fuente: Van De Walle y Cratty, 2005.

4. Utilización de datos secundarios

En la Sección 1 indicábamos que la recopilación de datos se puede llevar a cabo de cuatro maneras (véase White, 2006):

- realización de una encuesta específica para el proyecto;
- incorporación de un módulo especial en una encuesta diferente en curso;
- una encuesta sincronizada en que se entrevista a la población incluida en el proyecto pero el grupo de control se obtiene de otra encuesta (encuesta nacional de hogares, etcétera);
- evaluación basada en datos secundarios obtenidos con otros fines, pero que contienen información sobre el grupo beneficiario y los posibles grupos de control.

Casi todas las evaluaciones se basan en datos secundarios, incluso cuando se realizan encuestas, y en muchos casos los datos secundarios son la fuente principal o única de información. Por consiguiente, en la mayoría de las evaluaciones la cuestión no es si se debe utilizar la información secundaria, sino cómo garantizar la calidad y la idoneidad de una evaluación determinada.

Las ventajas de los datos secundarios

Los datos secundarios pueden ser una forma útil de reducir costos y ahorrar tiempo, y cuando se emplean diseños de grupos de beneficiarios y de comparación después de la intervención, suelen ser la única manera de *reconstruir* condiciones de referencia del grupo de beneficiarios y del grupo de comparación antes del inicio del proyecto. En este diseño, y en casi todos los diseños, pueden utilizarse para fortalecer la *estimación hipotética* de cuál habría sido la situación de la población del proyecto si éste no se hubiera llevado a cabo.

Algunos de los tipos más habituales de datos secundarios son los siguientes:

- Los censos nacionales;
- Las encuestas generales de hogares como el Estudio de Medición de los Niveles de Vida (EMNV);
- Las encuestas especializadas, como las encuestas demográficas y de salud;
- Los datos administrativos recopilados por departamentos ministeriales y otros organismos públicos (matrícula escolar, uso de centros de salud, precios de mercado de la producción agrícola, etcétera);
- Los estudios realizados por organismos donantes, organizaciones no gubernamentales y universidades;
- Los datos administrativos del organismo o ministerio encargado del proyecto;
- Los medios de comunicación (prensa, televisión, documentales, etcétera). Estos pueden ser útiles, entre otras cosas, para comprender el contexto económico y político local de cada lugar donde se ejecute el proyecto.

Otra importante aplicación de los datos secundarios es el metaanálisis, en que los impactos de proyectos o intervenciones comparables en este o en otros países proporcionan estimaciones de la magnitud de los efectos que cabe esperar de un proyecto bien diseñado. El metaanálisis puede resultar especialmente útil para estimar el tamaño que deben tener las muestras del grupo de beneficiarios y del grupo de comparación, ya que (si no se modifican los demás factores) cuanto menor sea el efecto previsto, menor será la potencia de la prueba y mayor la muestra necesaria para detectar los impactos del proyecto, si los hubiere⁶.

Los datos secundarios son muy valiosos para la evaluación, pero es probable que la información se haya recopilado para otros fines, por lo que las fuentes de información deben evaluarse cuidadosamente antes de su utilización. Entre los posibles problemas que deben examinarse antes de

hacer uso de datos secundarios se encuentran los siguientes (véase Bamberger, Rugh y Mabry, 2006, Capítulo 5):

- Hay un desfase cronológico entre el inicio el proyecto (que es cuando se necesitan los datos de referencia) y la recopilación o transmisión de los datos secundarios.
- ¿Qué semejanza hay entre la muestra y la población destinataria? Por ejemplo, ¿se incluyen en la encuesta tanto las escuelas públicas como las privadas? ¿Se tiene en cuenta el empleo del sector formal y del sector informal? ¿Se consideran hombres y mujeres, así como otros grupos de interés, como los ancianos?
- ¿Se obtuvo información sobre todas las variables importantes y los indicadores de resultados del proyecto? ¿Son los datos adecuados para la evaluación? A menudo, han de utilizarse uno o dos indicadores indirectos simples para medir indicadores complejos de resultados (por ejemplo, el uso de indicadores de servicios sanitarios prestados como indicador indirecto de los impactos en la salud, o el uso de los volúmenes y tipos de vehículos, y el número de empresas nuevas, como indicador del impacto de los caminos rurales).

Es asimismo importante evaluar la calidad y la exhaustividad de la información. A veces, cuando la información se recopila con fines administrativos, no hay controles de calidad y la información es incompleta, inexacta o sesgada (por ejemplo, cuando las escuelas tienen un incentivo para falsificar a su favor las tasas de matrícula o las puntuaciones de los exámenes, o la policía para no comunicar todos los delitos). Esto reviste especial importancia en el caso de las evaluaciones de impacto, ya que el incentivo de comunicar datos ficticios es mayor si la unidad que presta el servicio sabe que esos datos se utilizarán para una evaluación.

Utilización de datos secundarios para superar las limitaciones de presupuesto, tiempo e información

Los datos secundarios permiten reducir el tiempo o los recursos necesarios para reconstruir condiciones de referencia o grupos de comparación. Entre los procedimientos para ello cabe citar los siguientes:

- Los datos administrativos del proyecto se utilizan como referencia aproximada. Por ejemplo, los interesados en obtener una vivienda económica suelen tener que proporcionar información sobre su vivienda actual; los candidatos a microcréditos facilitan información sobre sus actividades económicas e ingresos actuales, y los estudios de fomento de infraestructura reúnen información sobre la accesibilidad y la calidad de la infraestructura existente.
- Los datos de encuestas de hogares se utilizan para establecer correspondencias entre los beneficiarios de los proyectos y otras comunidades u hogares no incluidos en ellos y que servirán como grupo de comparación.
- Es posible emplear los datos de censos y encuestas de hogares para crear un grupo de comparación.
- Los registros de los centros docentes o de salud locales en zonas de comparación que no están bajo la influencia del proyecto se emplean para estimar el marco hipotético de los programas que repercuten en la salud y la educación. Dichos registros pueden servir también para comparar la utilización de instalaciones nuevas respaldadas por el proyecto con escuelas, dispensarios y otros centros existentes en las zonas de los proyectos. Esto puede ser importante para tener en cuenta el hecho de que los usuarios pueden comenzar a utilizar las instalaciones más modernas y tal vez no sean comparables con el usuario medio de las instalaciones no mejoradas.
- Pueden utilizarse dos o tres fuentes secundarias diferentes para constituir grupos de comparación distintos a fin de lograr una mayor coherencia o realizar distintos tipos de análisis.

Estudio de caso 3. Utilización de fuentes múltiples de datos secundarios: Proyecto de nutrición integrada de Bangladesh

El objetivo del Departamento de Evaluación de Operaciones (DEO) era evaluar la contribución del impacto del componente de nutrición comunitaria del Proyecto de nutrición integrada de Bangladesh (PNIB) a la reducción de la desnutrición severa en los niños y en las mujeres embarazadas. La evaluación utilizó tres fuentes diferentes de datos secundarios: la evaluación del PNIB, que recopiló datos de referencia, a mitad del período y al final de la ejecución, la encuesta realizada por Save the Children al final del proyecto y el Proyecto de vigilancia nutricional (PVN) realizado por Helen Keller International. En estas encuestas se incluyeron no sólo resultados nutricionales, sino también una amplia gama de indicadores de proceso, lo que permitió la aplicación de un enfoque teórico. Los resultados de estas evaluaciones fueron contradictorios y se plantearon dudas acerca de la pertinencia de los grupos de comparación. Además, el estudio del DEO realizó un metaanálisis de estudios en países similares para averiguar el porcentaje de reducción de la mortalidad por 1.000 nacidos vivos que podía esperarse razonablemente de la contratación de parteras tradicionales. La repetición del análisis de los datos del PNIB con información del PVN para elaborar un grupo de comparación más apropiado mediante la correspondencia por puntuación de la propensión llevó a concluir que tanto la *estatura para el peso* como el *peso para la edad* habían mejorado en las zonas del proyecto.

La combinación de estas fuentes diferentes de datos secundarios permitió construir una cadena causal y localizar varios eslabones perdidos y que permitían explicar por qué el impacto del proyecto, que en términos estadísticos era significativo, resultaba decepcionante en términos operacionales (apenas un 5%, aproximadamente).

Se pudieron extraer varias enseñanzas del uso de datos secundarios para la evaluación de proyectos nutricionales. A veces se dispone de varias fuentes con abundantes datos secundarios que, si se combinan, pueden mejorar las estimaciones del impacto del proyecto fortaleciendo los grupos de comparación. Estas fuentes de datos pueden contribuir también a describir la cadena causal mediante la que deben conseguirse los impactos del proyecto. Por otro lado, el estudio reveló también que a menudo hay lagunas en la información por lo que respecta a la administración y la ejecución del proyecto y, además, se disponía de muchos menos datos acerca de factores relativos a la demanda.

Fuente: White, 2006.

5. Reconstrucción de datos de referencia

Los diseños posteriores acusan los efectos de la falta de datos de referencia

Las evaluaciones que no se encargan hasta que el proyecto está casi finalizado no ofrecen la opción de recopilar datos de referencia. Si bien algunas técnicas estadísticas, como la **puntuación de la propensión** (véase la Sección 3), pueden mejorar la correspondencia entre el grupo de beneficiarios y el de comparación, las evaluaciones realizadas después de la intervención suelen verse obligadas a definir el grupo de comparación posterior a la ejecución como *marco hipotético*, y esos diseños no pueden controlar las diferencias *no observadas* entre los dos grupos al inicio del proyecto (véase el Estudio de caso 1). El problema se agrava cuando los participantes se obtienen mediante *autoselección*, ya que las personas o comunidades que deciden participar en un proyecto suelen ser las que tienen más posibilidades de éxito (por ejemplo, las mujeres que reciben microcréditos y ya tienen experiencia empresarial, confianza en sí mismas y control de los recursos de la familia). Por consiguiente, la falta de datos de referencia dificulta la distinción entre los efectos del proyecto y las diferencias preexistentes.

A veces, hay encuestas o datos de censos anteriores que permiten estimar la situación reinante antes de realizar la intervención (véase la Sección 3), o existen otros tipos de datos secundarios como los registros del proyecto, los archivos de las escuelas y los centros de salud locales u otros organismos públicos que pueden proporcionar estimaciones generales de la situación de referencia (véase la Sección 4), pero a menudo sin comparaciones a nivel de hogares. Sin embargo, cuando no se dispone de datos secundarios inmediatos, el investigador debe plantearse algunas de las técnicas cualitativas que se describen a continuación.

Evaluación de estrategias para reconstruir datos de referencia

Los *datos secundarios* de encuestas nacionales de hogares, censos y estudios similares realizados en la misma época en que comenzó el proyecto pueden ayudar a identificar a la población beneficiaria en el proyecto y aportar datos pertinentes sobre esa población o sobre grupos de comparación potenciales. En ocasiones, las encuestas facilitan datos útiles sobre el grupo de comparación, pero el número de participantes en el proyecto es demasiado reducido para realizar análisis detallados.

- Es importante evaluar en qué medida el conjunto de datos satisface las necesidades de la evaluación. Por ejemplo, es posible que los datos no se hayan recopilado en el momento adecuado, que no abarquen a toda la población beneficiaria o de comparación, que no incluyan toda la información fundamental necesaria o que haya dudas sobre su fiabilidad. Sin embargo, la experiencia de las evaluaciones del Banco Mundial muestra que a menudo es posible encontrar datos secundarios de gran calidad que permiten atender la mayoría de las necesidades de información de referencia.

Los *registros administrativos* recopilados por el proyecto (por ejemplo, las características de las familias o comunidades que solicitan una vivienda, una plaza en una escuela, un crédito empresarial pequeño o servicios básicos) a menudo proporcionan datos de referencia sobre los participantes en el proyecto, pero casi nunca sobre los grupos de comparación.

- Hay que comprobar la fiabilidad, la integridad y la coherencia de los datos comunicados. Los organismos pueden introducir sesgos en la información que transmiten si saben que se utilizará para evaluar los resultados del proyecto. Otro problema es vincular los registros

administrativos con los datos recabados con posterioridad al proyecto, ya que muchas veces los números de identificación de los individuos están incompletos o no son fiables. Si no hay correspondencia entre los individuos, la evaluación no puede utilizar estudios de panel, que son mejores para manejar las variables *no observables* (invariante en el tiempo) de los individuos o las comunidades.

Los *registros de las escuelas, los centros de salud, las cooperativas de crédito y ahorro, las ventas en los mercados agrícolas locales*, pueden proporcionar, a veces, datos de referencia sobre el grupo de comparación.

- A veces no se sabe si los registros están completos o si existen sesgos sistemáticos (por ejemplo, puede que se incentive a los maestros para que exageren el número de niños que asisten a la escuela o a la policía para que comunique menos delitos de los que se producen en la realidad).

Recuerdo: A veces se pide a los beneficiarios que recuerden su situación antes de la intervención por lo que respecta a la asistencia a la escuela, el uso de centros de salud o el tiempo y el costo de un viaje determinado⁹.

- El recuerdo de información está sujeto a sesgos potenciales debido a problemas de memoria, a exageración o minimización de la realidad y a la distorsión de comportamientos socialmente deseables o indeseables. Los resultados son también muy sensibles al período de tiempo cubierto y a la forma de formular las preguntas. Salvo en unos pocos ámbitos, como el gasto y el comportamiento reproductivo, donde se han realizado investigaciones sistemáticas sobre la base del recuerdo, lo normal es que existan muy pocos estudios empíricos que estimen la magnitud o la dirección del sesgo (véase Bamberger, Rugh y Mabry, 2006, págs. 97-99). Esto puede favorecer la comparación de medidas múltiples de la variable de impacto.

El término *evaluación participativa rápida* (EPR) se usa actualmente como nombre genérico de toda una serie de técnicas con participación de los beneficiarios en que son más bien las comunidades, y no los individuos, quienes proporcionan información sobre su situación, los problemas de la colectividad y los cambios que se producen con el tiempo (Kumar, 2002). Los grupos comunitarios pueden proporcionar estimaciones, por ejemplo, sobre el volumen y la calidad del agua, la producción agrícola y su venta, el tiempo y el costo de un desplazamiento determinado y el empleo del tiempo. Deberían triangularse siempre varias fuentes independientes de información (véase el Estudio de caso 5); la triangulación consiste en realizar una comparación sistemática de las estimaciones obtenidas de dos o más fuentes independientes.

- Además de los problemas de sesgo del recuerdo, las EPR experimentan problemas de representatividad (¿quiénes participan en las consultas colectivas?) y de dinámica de grupo (cuando algún grupo domina las conversaciones). La recopilación de datos colectivos modifica asimismo la unidad de observación utilizada para medir el impacto, reduce el tamaño de la muestra y dificulta la integración de los hallazgos a nivel comunitario y los datos de encuestas individuales y de hogares¹⁰.

Es posible entrevistar a *informantes clave* sobre la situación que existía antes de realizar el proyecto. La opinión de varios informantes diferentes independientes debe triangularse, ya que puede incluir sesgos (voluntarios o involuntarios). Si bien los informantes principales sólo proporcionan información general sobre la comunidad o el grupo, ésta puede ser útil cuando no se dispone de datos de encuestas. Este enfoque puede emplearse también para evaluar algunas características *no observables* importantes como, por ejemplo, las características peculiares de los individuos o comunidades participantes que los hacen tener más posibilidades de éxito en el proyecto.

Los *datos administrativos* recopilados por el proyecto se pueden adaptar de modo que incluyan preguntas que serán de utilidad para una evaluación de impacto posterior¹¹.

Fortalecimiento de la calidad y análisis de los datos de referencia reconstruidos

- La lista de la Sección 4 puede utilizarse para evaluar los puntos débiles de todas las fuentes potenciales de datos secundarios. En la Sección 7 se proponen modos de abordar algunas amenazas habituales a la validez de los distintos enfoques descritos en esta sección.
- Debería emplearse la triangulación para comprobar las fuentes secundarias. Puede usarse para comparar estimaciones de dos o más preguntas incluidas en una encuesta o para comparar estimaciones cualitativas de, por ejemplo, datos de encuestas sobre el nivel económico, la participación en la fuerza laboral y la matrícula escolar. Algunas fuentes potenciales de triangulación son la observación directa, otras fuentes secundarias, los informantes clave, las encuestas de partes interesadas, las EPR, las fotografías y los artículos de prensa. Estas técnicas ayudarán al investigador a comprender las características *no observables* potenciales u otras fuentes de sesgo en los datos.

Estudio de caso 4. Reconstrucción de datos de referencia: el Fondo de Inversión Social de Emergencia de Nicaragua

Este caso ilustra el uso de cuatro fuentes de datos secundarios para obtener estimaciones independientes de accesibilidad e impacto del Fondo de Inversión Social de Emergencia (FISE) de Nicaragua y reconstituir condiciones de referencia. También ilustra la necesidad de estimaciones de impacto independientes cuando las comunidades y los beneficiarios participantes en el proyecto no se seleccionan aleatoriamente.

El FISE proporciona letrinas, escuelas, dispensarios y suministro de agua a comunidades de bajo ingreso seleccionadas entre las que solicitan participar en el programa. Los criterios de selección dieron prioridad a las comunidades pobres, pero tuvieron también en cuenta la capacidad de la colectividad para llevar a cabo el proyecto. Las comunidades no se seleccionan al azar, ni pueden definirse simplemente como las más pobres. Se utilizaron varias fuentes de datos secundarios para crear grupos de comparación independientes:

- El estudio EMNV de 1998 permitió elaborar mapas de pobreza para orientar la selección de zonas de intervención del FISE e identificar comunidades participantes y no participantes para cada componente del proyecto (suministro de agua y saneamiento, salud, educación, etcétera). A continuación, se combinaron las muestras con datos administrativos sobre los criterios de selección del proyecto para calcular **puntuaciones de propensión** que estimasen la probabilidad de vivir en una zona de influencia del respectivo componente del FISE. Las puntuaciones de la ganancia del proyecto (matrícula escolar, edad para cada curso, repetición de cursos, etcétera) se estimaron como la diferencia media entre el proyecto y las correspondientes comparaciones para cada variable de impacto.
- El censo se combinó con datos administrativos para seleccionar *muestras basadas en la selección* de beneficiarios directos y beneficiarios potenciales porque la muestra del FISE no incluía suficientes hogares para realizar comparaciones estadísticas.

Los estudios de hogares del EMNV y del FISE se llevaron a cabo cuando el Fondo había estado funcionando ya cinco años en algunas comunidades y no se habían realizado estudios de referencia. La evaluación se basó principalmente en una comparación posterior a la intervención de las zonas de tratamiento y no tratamiento utilizando las puntuaciones de propensión para buscar correspondencias entre los dos grupos. Las fuentes descritas más arriba se emplearon para fortalecer el análisis mediante la reconstrucción de datos de referencia. La técnica del recuerdo se utilizó también para obtener datos sobre medidas de las variables de resultados anteriores al inicio del proyecto.

Fuente: Pradhan y Rawlings, 2000.

5. Reducción de los costos de la recopilación de datos

Reducción de la duración o la complejidad del instrumento de encuesta

Las consultas al cliente para determinar qué información es fundamental y cuál es sólo “interesante” pueden propiciar, a menudo, reducciones considerables en la duración o la complejidad del instrumento de encuesta y, por ende, en los costos y el tiempo necesario para la recopilación de datos.

Reducción del número de entrevistas

Dado que la recopilación de datos puede representar muchas veces más de la mitad del presupuesto de la evaluación (Baker, 2000), la reducción del tamaño de la muestra puede redundar en un ahorro considerable. Sin embargo, hay que llegar a ciertas soluciones de compromiso, ya que cuando la muestra es más pequeña, disminuye la precisión estadística de las estimaciones y el nivel de desagregación del análisis. Entre los principales factores determinantes del tamaño de la muestra necesario para una evaluación concreta se encuentran los siguientes: el tamaño del efecto de tratamiento promedio estimado, la potencia deseada de la prueba estadística, la media y la varianza de las variables subyacentes, el nivel requerido de precisión estadística, si se usa o no un grupo de comparación, los tipos de análisis desagregado y si se dispone de las siguientes opciones para reducir el número de entrevistas:

- Aceptar un nivel más bajo de precisión estadística (por ejemplo, un intervalo de confianza del 90% en lugar del 95%) o de **potencia estadística de la prueba** (por ejemplo, un riesgo del 20% en lugar del 10% de rechazar un impacto real del proyecto)¹³. Por supuesto, esto aumenta la posibilidad de juzgar erróneamente si un proyecto ha tenido un impacto o no.
- Reducir el nivel de desagregación estadística del análisis (por ejemplo, obtener resultados sólo para la población beneficiaria total en lugar de comparar impactos en grupos diferentes o la efectividad de componentes diferentes del proyecto).
- Cuanto mayor sea la magnitud del efecto esperado, menor será la muestra necesaria para encontrar un impacto estadísticamente significativo.

Sustituir entrevistas individuales por la recopilación de datos de alcance comunitario

Si bien las entrevistas individuales proporcionan la información más detallada y precisa estadísticamente, hay otras formas más económicas de reunir información comparable, aunque no tan exacta en términos cuantitativos, en la comunidad o el grupo:

- Puede utilizarse una lista de comprobación en la comunidad para estimar, por ejemplo, los patrones de viaje y transporte, el tiempo dedicado a la recogida de agua y combustible o la calidad y la utilización de los servicios. Las observaciones se recopilan en distintos lugares de la comunidad (carreteras y caminos, rutas hacia las fuentes de agua o a los centros de salud locales). Aunque su administración es fácil y económica, esas listas suelen reducir el total de observaciones, pues se pasa de muchos individuos a un reducido número de comunidades. Después, habrá que considerar si será necesario incorporar más comunidades para obtener resultados estadísticamente significativos y, en caso afirmativo, si se mantendrá la reducción general de costos. Idealmente, para fortalecer el análisis debería recopilarse información adicional sobre los atributos relacionados (covariables) a nivel de hogares o comunidades, lo cual podría reducir aún más los costos.

- Asimismo, se puede recurrir a los grupos focales, las técnicas de evaluación participativa rápida o las entrevistas comunitarias para obtener estimaciones comunitarias de la utilización y la calidad de los servicios, la producción agrícola, el empleo del tiempo o la división del trabajo entre el hombre y la mujer. Una vez más, se suscita la cuestión, abordada más arriba, de las consecuencias de disminuir el número de observaciones.

Reducción del costo de las entrevistas

- Por lo que atañe a las evaluaciones de programas de salud o educación, puede contratarse a personal sanitario o docente que conozca la zona y tenga aceptación social para que realice entrevistas y, a veces, ayude a codificar los datos. En encuestas socioeconómicas más generales, puede contratarse también a estudiantes universitarios o a graduados de secundaria. Aunque la remuneración por cada entrevista será más baja que la contratación de una empresa especializada en encuestas, hay que prever más tiempo y dinero para la formación y, quizás, para una mayor supervisión.
- Se pueden utilizar cuestionarios administrados por los propios encuestados en lugar de contratar a un equipo de entrevistadores para recabar la información. Naturalmente, esta opción sólo es viable cuando se encuesta a poblaciones con cierto nivel de formación, como en el caso de las evaluaciones de los programas de enseñanza secundaria. Este enfoque puede introducir también sesgos en la tasa de respuesta, incluida la diferencia sistemática de sesgo entre el grupo de tratamiento y el grupo de comparación. Los sesgos son más difíciles de manejar cuando el agente censal no está presente en el momento de cumplimentar la encuesta.

Utilización de diarios para reducir el costo de la recopilación de datos sobre los ingresos y los gastos para una evaluación de impacto

En una evaluación de los impactos de inversiones en viviendas sobre la capacidad de las familias pobres para cubrir sus gastos básicos, 100 familias aceptaron mantener un registro cotidiano de sus ingresos y gastos a lo largo de un año. La única compensación para las familias fue la posibilidad de elegir un pequeño regalo para el hogar cada mes. Esto resultó mucho más económico que utilizar un equipo de agentes para visitar a todas las familias al menos una vez a la semana. Sin embargo, el éxito de este planteamiento dependió del grado excepcionalmente alto de cooperación de las familias que, en su totalidad, participaban en un proyecto de autoayuda para la construcción de viviendas.

Fuente: Valadez y Bamberger, 1994.

Recopilación electrónica de datos

Aunque muchos países en desarrollo donde se realizan evaluaciones no disponen de muchas de las actuales tecnologías electrónicas, su acceso a esos recursos es cada vez mayor, al igual que las posibilidades de reducir considerablemente los fondos y el tiempo dedicados a la recopilación de datos:

- Se pueden sustituir las entrevistas directas por entrevistas telefónicas. Aunque no todos los encuestados tengan teléfono, muchos pueden recibir llamadas en el centro de la comunidad o en casa de algún amigo. A medida que los teléfonos móviles se abaratan, es posible también dar o prestar un teléfono móvil a los encuestados.

Las entrevistas telefónicas pueden ahorrar tiempo y dinero

En la evaluación de un programa de cupones escolares en Colombia, la mayoría de las entrevistas se realizaron por teléfono. Dado que el acceso a un teléfono era un criterio de participación, las entrevistas telefónicas no introdujeron un sesgo significativo de la selección de la muestra y, ciertamente, supusieron un ahorro de tiempo y dinero.

Fuente: Angrist y cols., 2002.

- Las encuestas por correo electrónico se usan cada vez más a nivel organizativo. Muchos encuestados, incluso los que son muy pobres, pueden acudir a centros comunitarios de telecomunicaciones o a cibercafés, de modo que este medio de comunicación ya se está utilizando más ampliamente en algunos tipos de encuesta.
- Es posible utilizar contadores automáticos para registrar, por ejemplo, el número de personas que entran en un edificio, o el tráfico de peatones o vehículos.
- Los sistemas de información geográfica y las fotografías aéreas pueden ser, en ocasiones, medios económicos de obtener información física, por ejemplo, sobre el número y el tamaño de las casas y otras construcciones, las pautas de transporte y las tierras cultivadas.

Participación en los costos

A veces es posible compartir los gastos de la recopilación y el análisis de datos con otro organismo o incorporar el estudio (en especial cuando el instrumento es breve) a otra encuesta. En ocasiones, puede administrarse un módulo dentro de una submuestra de hogares incluidos en una encuesta nacional de hogares en curso.

Estudio de caso 5. Utilización de técnicas de evaluación participativa rápida (EPR) para reducir el costo de la recopilación de datos: Flores, Proyecto de suministro de agua y saneamiento a aldeas en Indonesia

Este caso ilustra el uso de técnicas de EPR como medio económico para reconstruir las condiciones de referencia y evaluar los cambios en el acceso y la calidad de los servicios. La muestra de comunidades es suficientemente amplia para realizar un análisis estadístico.

La evaluación examinó los efectos de un enfoque de gestión comunitaria del suministro de agua y saneamiento en el acceso, el uso efectivo y la sostenibilidad de los servicios. Se empleó la metodología de evaluación participativa, que es una técnica de EPR, para conseguir la perspectiva de grupos comunitarios diferentes, así como de organismos públicos locales y nacionales. De un universo integrado por 260 puntos donde varios organismos públicos y ONG estaban llevando a cabo proyectos de suministro de agua y saneamiento, además del proyecto del Banco Mundial, se extrajo una muestra aleatoria estratificada de 63 puntos. Se utilizaron las siguientes técnicas participativas: a) clasificación de bienestar de hogares, dividiendo los hogares durante una reunión comunitaria abierta en tres grupos (situación más favorecida, intermedia y menos favorecida); b) mapa social y secciones transversales en que representantes masculinos y femeninos de los tres estratos evaluaron el acceso, la calidad y la sostenibilidad de los servicios; c) entrevistas de comités en que se emplearon diversas técnicas de extracción de tarjetas para establecer una evaluación ordinal de la gestión y la calidad de los servicios. Las conclusiones se presentaron y debatieron en reuniones comunitarias abiertas. Las puntuaciones ordinales de los distintos aspectos del programa se compararon con las puntuaciones del grado de participación comunitaria en la planificación y la gestión y la medida en que se hizo partícipes a las mujeres. Aunque el diseño de la evaluación permitió comparar proyectos hídricos de distintos organismos, no incluyó un grupo de comparación de aldeas sin proyectos y, por lo tanto, la evaluación sólo compara la eficacia de los distintos sistemas de prestación de servicios y no permite una comparación entre participantes y no participantes. Además, no se debatieron los sesgos de selección debido a las características especiales de las comunidades seleccionadas para los proyectos. La evaluación costó US\$150.000 (incluidos US\$45.000 para asesores internacionales) y duró 12 meses.

Fuente: Hopkins y Mukherjee, 2005.

7. Fortalecimiento de los diseños de evaluación cuando se trabaja con limitaciones de presupuesto, tiempo e información

Efectos de las limitaciones de presupuesto, tiempo e información en la calidad de las evaluaciones de impacto

Cuando se realizan evaluaciones de impacto con limitaciones de presupuesto, tiempo o información, aumenta la dificultad de manejar las cuatro series de amenazas a la calidad del diseño y la validez de las conclusiones¹⁴. Aunque esas cuatro series de *amenazas a la validez de las conclusiones de la evaluación* pueden afectar a todas las evaluaciones, son más difíciles de gestionar en las situaciones reales. En el Cuadro 2 se describen los problemas comunes de estas cuatro categorías y se especifican las limitaciones que suelen contribuir a cada problema:

- *Amenazas a la calidad general del diseño y la ejecución de la evaluación*: Las limitaciones de recursos pueden restringir la atención prestada al diseño de la evaluación, al desarrollo y a la verificación de instrumentos, así como a las consultas a clientes; asimismo, puede haber presiones para recortar los costos de la recopilación de datos mediante la contratación de entrevistadores más económicos, o reduciendo la formación y la supervisión de los entrevistadores. Por otra parte, puede ser más difícil utilizar enfoques de métodos mixtos y triangulación para controlar la calidad y comprobar plenamente la idoneidad de las fuentes de datos secundarios.
- *Amenazas al análisis estadístico*: Las limitaciones dificultan la toma de medidas para fortalecer el marco muestral o abordar los sesgos de muestreo. Puede haber, asimismo, presiones para reducir el número de puntos de recopilación de datos (por ejemplo, eliminar datos de referencia o grupos de comparación). También puede haber presiones para reducir el tamaño de la muestra, lo que disminuye la potencia de la prueba estadística y limita la posibilidad de realizar un análisis desagregado.
- *Coherencia teórica e idoneidad del marco hipotético*: Las limitaciones dificultan la realización de estudios exploratorios, la consulta a clientes y los seminarios necesarios para elaborar una teoría sobre el programa que explique cómo se espera que éste alcance sus objetivos y cómo afectan las variables contextuales y el proceso de ejecución del proyecto a la magnitud y la distribución de los impactos. Las limitaciones debilitan también el marco hipotético al eliminar puntos de recopilación de información o reducir el acceso a datos secundarios que pueden fortalecer el grupo de comparación.
- *Posibilidad de generalizar los resultados*: Cuando la evaluación no puede manejar los sesgos muestrales o analizar factores contextuales que influyan en los resultados en lugares concretos, aumenta el riesgo de sacar conclusiones erróneas sobre la posibilidad de repetir el proyecto.

Modos de abordar los efectos de las limitaciones de presupuesto, tiempo e información en la validez de las conclusiones de la evaluación

En el párrafo anterior se ha descrito cómo las limitaciones de presupuesto, tiempo e información agravan muchas de las amenazas más habituales a la validez de los diseños de evaluación. Cuando existe restricción de recursos, habrá que llegar casi siempre a soluciones de compromiso en el modo de utilizar los recursos, por ejemplo, recortando gastos mediante la contratación de entrevistadores que cobren menos o reduciendo el número de estudios minuciosos de casos prácticos, mientras que otros compromisos se refieren a decisiones sobre si deben invertirse recursos escasos para aumentar el tamaño de la muestra, mejorar la cobertura y la calidad del marco muestral o reducir las tasas de falta de respuesta pidiendo que se repitan con mayor frecuencia las visitas a las familias. En el Cuadro 2 se ofrece al lector una lista indicativa básica sobre los aspectos potenciales en los que debe encontrarse un compromiso (véase Bamberger, Rugh y Mabry, 2006). Aunque tal vez esos compromisos sean inevitables, el hecho de determinar cómo pueden afectar a los resultados es una forma útil de orientar las decisiones. No obstante, puede haber maneras de disminuir las limitaciones y elementos que no deben ponerse en peligro nunca. En la siguiente sección se brinda orientación a este respecto.

Cuadro 2. Repercusiones de las limitaciones de presupuesto, tiempo e información en la calidad de una evaluación de impacto			
Problemas (amenazas a la validez) causadas por las distintas limitaciones	Limitaciones que contribuyen a cada problema		
	Presupuesto	Tiempo	Información
A. Problemas que afectan a la calidad general del diseño de la evaluación (amenazas a la validez interna)			
Atención insuficiente a la planificación, la consulta a los clientes y el desarrollo de la relación con consultores locales	✓	✓	
Atención insuficiente al desarrollo y a la verificación de instrumentos	✓	✓	
Falta de tiempo para el seguimiento de los resultados de la evaluación	✓	✓	
Exclusión de grupos difíciles de alcanzar y sobre los cuales es difícil obtener información	✓	✓	✓
Menor aplicación de enfoques de métodos mixtos, por lo que no pueden utilizarse comprobaciones de coherencia por triangulación	✓	✓	
Presión para encontrar entrevistadores más baratos y menos recursos para la formación y la supervisión	✓	✓	
Más dependencia de métodos cualitativos rápidos	✓	✓	
Mayor dificultad de comprobar la idoneidad de los datos secundarios	✓	✓	✓
B. Problemas que afectan al diseño de la muestra y al análisis estadístico (amenazas a la validez estadística)			
Menor posibilidad de aplicar enfoques de métodos mixtos	✓	✓	✓
Menos recursos para mejorar la calidad del marco muestral	✓	✓	
Mayor dificultad para hacer frente a los sesgos muestrales y mejorar las correspondencias	✓	✓	✓
Menor calidad de la aplicación de la muestra	✓	✓	✓
Menor tamaño de la muestra — riesgo de falsos negativos	✓		
Presiones para eliminar la recopilación de datos sobre el grupo de tratamiento o el grupo de control, o el grupo de comparación posterior a la intervención.	✓		✓
Menor análisis desagregado	✓	✓	✓
C. Problemas que afectan a la coherencia de la teoría y la validez del marco hipotético (amenazas a la validez de constructo)			
Menos tiempo y recursos para elaborar un modelo teórico para el proyecto, de modo que los conceptos e indicadores principales pueden estar menos definidos y tal vez no se puedan establecer las principales hipótesis o se establezcan erróneamente	✓	✓	✓
Menor uso de enfoques basados en métodos múltiples y de la triangulación	✓	✓	✓
Grupo de control/comparación más débil (más pequeño o con una correspondencia peor)	✓		✓
Datos de referencia poco fiables o inexistentes	✓	✓	✓
D. Problemas que afectan a la posibilidad de generalizar los resultados y las recomendaciones sobre la posibilidad de repetir el proyecto en otros entornos o con otros grupos (amenazas a la validez externa)			
Falta de atención a los sesgos muestrales	✓		✓
Análisis débil de factores contextuales que contribuyen al éxito o al fracaso en lugares concretos	✓	✓	✓

Fortalecimiento de la calidad general del diseño de la evaluación

- Incluso cuando se trabaja con limitaciones de presupuesto y tiempo, es importante disponer siempre de tiempo suficiente para reunirse con los clientes y con los principales interesados para comprender sus necesidades de información, plazos y limitaciones. Todas las decisiones de diseño para ahorrar tiempo o dinero deben tomarse en consulta con los clientes para que éstos entiendan y acepten plenamente los compromisos que deben realizarse.
- Cuando el tiempo es la principal limitación, se puede comenzar rápidamente organizando videoconferencias con los organismos locales y los investigadores, así como encargando estudios preparatorios que deben terminarse *antes* de que lleguen los asesores extranjeros.
- Deben estudiarse las repercusiones financieras del modo de tratar los sesgos de selección y las variables instrumentales. En general, resulta más económico trabajar con personal del proyecto para tratar los sesgos de selección durante el diseño que realizar costosas encuestas después de la intervención para buscar correspondencias entre el grupo beneficiario y el grupo de comparación respecto de un conjunto de criterios de selección de difícil identificación. Los jefes de proyecto pueden ponerse de acuerdo para introducir criterios de selección de participantes más explícitos a fin de evitar algunos sesgos, o puede fortalecerse la recopilación de datos administrativos durante la selección de los participantes para que los criterios de selección efectivos estén mejor documentados. La aplicación y la documentación de criterios claros de selección y rechazo fortalecen los posibles tipos de análisis.
- La elaboración de un modelo teórico (por ejemplo, lógico) para el programa y la articulación de una *cadena de efectos* a través de la cual se espera conseguir los impactos pueden contribuir a determinar las hipótesis fundamentales y los aspectos en que deben centrarse los recursos escasos para la evaluación.
- La revisión por homólogos, que en principio es un componente habitual de cualquier evaluación, resulta muy útil cuando se examinan las amenazas a la validez de las medidas para abordar las limitaciones reales, al aportar otra perspectiva para apoyar los compromisos que se realizan.
- Cuando el presupuesto no permite realizar una encuesta de hogares por muestreo, la EPR y otras técnicas cualitativas pueden proporcionar, a menudo, estimaciones de alcance comunitario, por ejemplo, del consumo de agua o del uso de nuevas instalaciones sanitarias (véase el Estudio de caso 5). Sin embargo, si deben utilizarse estas técnicas para la evaluación del impacto, se precisará una muestra de comunidades lo suficientemente amplia para poder llevar a cabo análisis estadísticos y habrá que evaluar si se mantiene la reducción de costos.
- Cuando el tamaño de la muestra no es todo lo grande que cabría desear, a veces es posible formular enfoques rentables de métodos mixtos que fortalezcan la validez proporcionando dos o más estimaciones independientes de productos, resultados o indicadores de impacto fundamentales. Por ejemplo, si las estimaciones de los informantes clave sobre los cambios en la cantidad y la fiabilidad del suministro de agua son compatibles con las encuestas de hogares, aumenta la credibilidad de las estimaciones. Sin embargo, el empleo de métodos mixtos incrementa los costos de la compilación de datos, de modo que es importante determinar si esta estrategia favorece una reducción general de los gastos. Asimismo, es importante recordar que si la muestra es demasiado pequeña, seguirá siendo imposible aplicar pruebas estadísticas para el examen de las hipótesis, aun en el caso de que los métodos mixtos aumenten la credibilidad de una relación causal.

Fortalecimiento del diseño de la muestra y del análisis estadístico

- *Cuestiones relacionadas con el tamaño de la muestra:* La reducción del tamaño de la muestra es una forma muy tentadora de ahorrar fondos, pero las muestras más pequeñas agravan el riesgo de *negativos falsos* (suposición errónea de que el proyecto no tuvo un impacto determinado). El análisis del poder estadístico (véase la Sección 4) es un modo útil de asegurar que la muestra propuesta sea lo suficientemente grande para los fines del análisis.

Fortalecimiento del marco teórico y de la validez de la hipótesis

- Los estudios de evaluación rápida son un medio rentable de elaborar los modelos teóricos de programas de los que hemos hablado más arriba.
- Cuando se trabaja con muestras pequeñas, los enfoques de métodos mixtos pueden constituir un medio rentable de comprender los conceptos fundamentales o mejorar su medición.
- El marco hipotético se puede fortalecer reconstituyendo las condiciones de referencia (aun cuando se usan diseños de correspondencia de puntuaciones de propensión basados en comparaciones posteriores a la prueba sin datos de referencia) y, cuando sea necesario, reforzando los grupos de comparación mediante las técnicas abordadas en la Sección 5.
- Deben asignarse tiempo y recursos para formular una teoría del programa. Los modelos más sencillos pueden elaborarse en un período de tiempo relativamente breve. A menudo, el evaluador comprobará que, aunque el personal del programa tenga sus propias ideas sobre los objetivos de éste y el modo de alcanzarlos, el modelo del programa no se ha formulado oficialmente y debe perfilarse durante las entrevistas, los seminarios y el estudio de los documentos relativos al programa. Sin embargo, es posible hacerlo en forma rápida y económica.

Fortalecimiento y generalización de las conclusiones

- Deben utilizarse métodos de evaluación rápida (informantes clave, grupos focales, observación, etcétera) para establecer las similitudes y las diferencias entre el grupo de beneficiarios y el de comparación, y comprender el modo en que esto podría afectar a la posibilidad de generalizar los resultados. Asimismo, pueden utilizarse métodos mixtos para interpretar los resultados: un buen análisis cualitativo puede esclarecer el proceso que produjo el resultado cuantificado, aunque inicialmente no se haya procesado el análisis.
- Conviene recurrir al análisis contextual (descripción y análisis cualitativo o cuantitativo del contexto económico, político, institucional y sociocultural local en cada lugar donde se realice el proyecto) para comprender la influencia de los factores locales en los resultados y evaluar en qué medida esos factores son específicos de ese contexto o se pueden generalizar¹⁵.
- Siempre que sea posible, debería emplearse el análisis de variables múltiples para fortalecer la correspondencia entre el grupo de beneficiarios y el de comparación para, de ese modo, aumentar la validez de las proyecciones de las condiciones en las que podría repetirse el proyecto.

Evaluación de la viabilidad de una evaluación de impacto

Antes de llevar a cabo una evaluación, debería efectuarse un *análisis de la "evaluabilidad"* para determinar si se pueden cumplir los requisitos para llevar a cabo una evaluación de impacto de calidad (véase la Sección 1) con los recursos, plazos y datos disponibles. Esto se hace después de

26

haber estudiado todas las posibilidades para fortalecer el diseño de la evaluación, empleando las técnicas mencionadas más arriba, y de proponer el mejor diseño disponible. Si no es posible llevar a cabo una evaluación de impacto aceptable con estas limitaciones, hay que volver a negociar los recursos y plazos, revisar el alcance y los objetivos de la evaluación, o anular la evaluación. Cabe recordar que muchas veces se puede efectuar una evaluación operacionalmente útil de los efectos potenciales del proyecto incluso cuando las condiciones no permiten realizar una evaluación de impacto rigurosa.

NOTAS

¹ El Banco Mundial, con el apoyo del Grupo de Evaluación Independiente, está ayudando a un número creciente de gobiernos a desarrollar y reforzar sus sistemas de seguimiento y evaluación.

² Se han elegido varios estudios de casos prácticos para ilustrar el uso de determinados métodos, pero ello no significa necesariamente que sean evaluaciones de impacto en el sentido del término utilizado en el presente documento.

³ <http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/0,,contentMDK:20381417~menuPK:773951~pagePK:64165401~piPK:64165026~theSitePK:469372,00.html>.

⁴ Si esos factores contextuales tienen un efecto sistemáticamente diferente en el grupo de beneficiarios y en el de comparación, se puede incrementar o reducir artificialmente el efecto medio del proyecto.

⁵ CARE Internacional ha estimado recientemente que en torno al 50% de las evaluaciones de proyectos que ha realizado en todo el mundo sufren restricciones de presupuesto, tiempo, información y logística para usar algunas variantes de este diseño (véase Bamberger, Rugh y Mabry, 2006, Capítulo 10).

⁶ El Departamento de Evaluación de Operaciones del Banco Mundial intentó usar un grupo transitorio de comparación para evaluar un proyecto de riego en Andhra Pradesh (India), pero comprobó que los agricultores incluidos en las últimas fases se encontraban, en general, en zonas más remotas y presentaban características diferentes de los agricultores de la fase uno (véase White, 2006, p. 14).

⁷ Véase en Baker, 2000, Cuadro 3.1, una breve explicación de las puntuaciones de propensión y en Ravallion, 2005, un análisis más detallado.

⁸ Por ejemplo, en el Proyecto de nutrición integrada de Bangladesh se realizó un metaanálisis internacional para estimar el número de muertes por 1.000 nacidos vivos que sería razonable esperar salvar mediante la contratación de parteras tradicionales. Se comprobó que el intervalo era de cinco a siete muertes evitadas por 1.000 nacidos vivos. Este dato se empleó para confirmar que la meta de evitar siete fallecimientos por 1.000 nacidos vivos era realista (véase White, 2006, Caso 3). Podría haberse utilizado ese mismo análisis para calcular la magnitud del efecto esperado al estimar el tamaño de la muestra necesario para una evaluación de impacto.

⁹ Por ejemplo, la evaluación del Fondo de Desarrollo Comunitario de Eritrea utilizó un diseño de evaluación posterior a la intervención. A fin de reconstituir las condiciones de referencia existentes antes de la construcción de escuelas comunitarias, se pidió a las familias que recordaran el período anterior a la construcción de las mismas, cuáles de sus hijos asistían a la escuela (en especial las diferencias de género en la asistencia), cuánto tiempo necesitaban los niños para desplazarse desde la escuela hasta la escuela y el costo del desplazamiento. Se hicieron preguntas similares acerca de los centros de salud de las aldeas. Las respuestas de las familias se cotejaron con información proporcionada por los principales informantes (ancianos de la aldea, maestros, funcionarios locales, etcétera). Los evaluadores consideraron que el recuerdo era relativamente fiable en este caso porque la construcción de una escuela o un dispensario en una aldea remota era un acontecimiento que todos podían recordar fácilmente y no había ninguna razón obvia por la que los encuestados pudieran querer alterar la información. Una técnica análoga de recuerdo podría resultar mucho más difícil de aplicar en una zona urbana, donde las familias tienen acceso a menudo a varias escuelas o centros de salud y se pueden haber puesto en marcha otros programas en el mismo período, lo que haría más difícil para los encuestados centrarse en el proyecto concreto de interés para la evaluación.

¹⁰ El estudio EMNV comprende un módulo para análisis a nivel de la comunidad y los servicios (Frankenberg, 2000, Volumen 1, págs. 315-338) y existe experiencia en el tratamiento de estas cuestiones analíticas a varios niveles.

¹¹ Por ejemplo, solía preguntarse a los interesados en participar en un programa de viviendas de bajo costo dónde estaban viviendo en ese momento y cuáles eran sus ingresos. Sería relativamente fácil y económico pedir información más detallada sobre sus condiciones de vida (tamaño y calidad de la vivienda, acceso a servicios, medios de transporte disponibles), la situación económica y las actividades del mercado laboral. Sin embargo, dependiendo de la presión laboral del personal de recepción y el número de solicitantes, tal vez fuera necesario contratar a personal adicional para recabar esa información, lo que significa que esa actividad quizá podría llevar aparejado un costo.

¹² Sobre el análisis del efecto de cada uno de estos factores en el tamaño de la muestra, véase Bamberger, Rugh y Mabry, 2006, Capítulo 14.

¹³ La potencia estadística de una prueba es la probabilidad de rechazar erróneamente una asociación significativa, desde el punto estadístico, entre el proyecto y la variable dependiente (indicador de impacto). El riesgo de rechazar equivocadamente un impacto significativo puede reducirse aumentando la potencia de la prueba (pasando, por ejemplo del típico 0,8, que acepta un riesgo del 20% de rechazar erróneamente el impacto, a una potencia del 0,9, por ejemplo, que reduce ese riesgo al 10%). El aumento de la potencia incrementa el tamaño de la muestra, mientras que la aceptación de una potencia más baja puede reducir significativamente el tamaño de la muestra (véase Bamberger, Rugh y Mabry, 2006, Capítulo 14).

¹⁴ Estas categorías se basan en cuatro series de amenazas a la validez de las conclusiones de la evaluación examinadas en las publicaciones sobre diseños cuasiexperimentales. Las amenazas son las siguientes: validez interna de las conclusiones (calidad general del diseño de la evaluación); validez de la conclusión estadística; validez de constructo (coherencia de la teoría e idoneidad de la hipótesis), y validez externa (posibilidad de generalizar las constataciones). Puede verse un análisis pormenorizado de las amenazas a la validez, así como listas y hojas de trabajo para aplicar los conceptos sobre el terreno, en Bamberger, Rugh y Mabry, 2006, Capítulo 7 y Anexos 1, 2 y 3.

¹⁵ Puede verse un examen del análisis contextual en Bamberger, Rugh y Mabry, 2006, Capítulo 9.

REFERENCIAS

- Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King y Michael Kremer. 2002. 'Vouchers for private schooling in Colombia: evidence from a randomized natural experiment'. *American Economic Review* 92 (5): 1535-1558.
- Baker, Judy. 2000. *Evaluating the Impacts of Development Projects on Poverty: A Handbook for Practitioners*. Washington, D.C.: Banco Mundial.
- Bamberger, Michael, Jim Rugh y Linda Mabry. 2006. *Real World Evaluation: Working under Budget, Time, Data and Political Constraints*. Thousand Oaks, CA: Sage.
- Frankenberg, Elizabeth. 2000. 'Community and price data', en Margaret Grosh y Paul Glewwe (comps.), *Designing Household Survey Questionnaires for Developing Countries. Lessons from 15 Years of the Living Standards Measurement Study*, Capítulo 9. Washington, D.C.: Banco Mundial.
- Hopkins, Richard y Nilanjana Mukherjee. 2005. 'Assessing the effectiveness of water and sanitation interventions in villages in Flores, Indonesia', Operations Evaluation Department, 2005, *Influential Evaluations*, 22-30. Washington, D.C.: Banco Mundial.
- Kumar, Somesh. 2002. *Methods for Community Participation*. Londres: ITDG Publications.
- Newman, Constance. 2001. *Gender, Time Use, and Change: Impacts of Agricultural Export Employment in Ecuador*. Policy Research Report on Gender and Development. Working Paper Series No 18. Washington, D.C.: Banco Mundial.
- Operations Evaluation Department (OED). 2004. *Monitoring and Evaluations: Some Tools, Methods and Approaches*. Washington, D.C.: Banco Mundial.
- _____. 2004. *Influential Evaluations: Evaluations that Improved Performance and Impacts of Development Programs*. Washington, D.C.: Banco Mundial.
- _____. 2005. *Influential Evaluations: Detailed Case Studies*. Washington, D.C.: Banco Mundial.
- Pradhan, Menno y Laura Rawlings. 2000. 'The impact and targeting of social infrastructure investments: lessons from the Nicaraguan Social Fund'. *World Bank Economic Review* 16 (2): 275-295.
- Ravallion, Martin. 2001. 'The Mystery of the Vanishing Benefits: An introduction to Impact Evaluation,' *World Bank Economic Review* 15 (1): 115-140.
- _____. 2005. *Evaluating Anti-Poverty Programs*. Policy Research Working Paper No. 3625. Washington, D.C.: Banco Mundial.
- _____. 2006. *Evaluating anti-poverty programs. Handbook for Agricultural Economics* (editado por Robert Evenson y Paul Schulz), Volumen 4. North-Holland.
- Valadez, Joseph y Michael Bamberger. 1994. *Monitoring and Evaluating Social Programs in Developing Countries*. Washington, D.C.: Banco Mundial.

Van De Walle, Dominique y Dorothyjean Cratty. 2005. *Do Donors Get What They Paid For? Micro Evidence on the Fungibility of Development Project Aid*. World Bank Policy Research Working Paper No. 3542. Washington, D.C.: Banco Mundial.

White, Howard. 2006. *Impact Evaluation: The Experience of the Independent Evaluation Group of the World Bank*. Washington, D.C.: Banco Mundial (de próxima aparición).

Recursos adicionales sobre seguimiento y evaluación

Sitios web

- Grupo de Evaluación Independiente del Banco Mundial:
<http://www.worldbank.org/ieg/>
- Grupo de Evaluación Independiente del Banco Mundial —
evaluación del impacto: <http://www.worldbank.org/ieg/ie/>
- Banco Mundial — evaluación del impacto:
<http://www.worldbank.org/impactevaluation/>
- Fortalecimiento de los sistemas públicos de seguimiento
y evaluación: <http://www.worldbank.org/ieg/ecd/>
- Noticias sobre seguimiento y evaluación:
<http://www.mande.co.uk/>