

N° 00000 1022

27.34

P. 831

V. 12

"PROGRAMA UNIVERSITARIO DE
ESTUDIOS DE GENERO" - U.N.A.M.

"PROGRAMA UNIVERSITARIO DE ESTUDIOS DE GÉNERO" - U.N.A.M. 006056

Investigación operativa en planificación familiar y salud reproductiva

Federico R. León¹

1. Population Council.

EL POPULATION COUNCIL busca ayudar a mejorar el bienestar y la salud reproductiva de las generaciones presentes y futuras del mundo, y ayudar a alcanzar un balance humano, equitativo y sustentable entre la población y los recursos. El Council analiza temas y tendencias poblacionales; realiza investigación biomédica para desarrollar anticonceptivos nuevos; trabaja con organizaciones públicas y privadas para mejorar la calidad y alcance de los servicios de planificación familiar y salud reproductiva; ayuda a los gobiernos a influir sobre las conductas demográficas; divulga resultados de investigación en el área de la población a públicos apropiados; y ayuda a fortalecer la capacidad en investigación en países menos desarrollados. El Council es una organización no lucrativa que fue fundada en 1952. Se dedica a la investigación y está dirigida por una junta directiva multinacional. Su oficina matriz en Nueva York apoya una red global de oficinas en regiones y países, entre ellas la Oficina Regional para América Latina y el Caribe. Las direcciones de estas oficinas son:

Population Council
One Dag Hammarskjold Plaza
Nueva York, Nueva York 10017, E.U.A.
Tel.: (212) 339-0500
Fax: (212) 755-6052
Correo electrónico: pubinfo@popcouncil.org
<http://www.popcouncil.org>

Population Council
Oficina Regional para América Latina y el Caribe
Escondida 110 Villa Coyoacán, 04000 México, D.F.
Tel. (525)554-8610; Fax: (525)554-1226;
Correo electrónico: inopalmx@mpsnet.com.mx

Esta publicación fue financiada por el Proyecto INOPAL III (Investigación Operativa y Cooperación Técnica en Planificación Familiar y Salud Reproductiva para América Latina y el Caribe) bajo el Contrato No. CCP-95-00-00007-00 con la Oficina de Población de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID).

Responsable de la publicación: Silvia Llaguno
Impresión: Solar, Servicios Editoriales, S.A. de C.V.
©1998 Population Council/INOPAL III

Cualquier parte de este documento podrá reproducirse sin permiso de los autores, siempre y cuando se reconozca la fuente y la información no se utilice con fines de lucro. Agradeceremos cualquier comentario o sugerencia de los lectores. Para solicitar más copias de este documento, materiales de difusión o requerir más información, por favor solicitarla a la Oficina Regional para América Latina y el Caribe.

**Investigación operativa
en planificación familiar
y salud reproductiva**

Resumen

El concepto "operations research" en planificación familiar y salud reproductiva está disponible en libros y artículos de amplia circulación, como los de Fisher y cols. (1991) y Ross y cols. (1987). Las definiciones del concepto, sin embargo, presentan algunas contradicciones y no dejan enteramente claro en qué se diferencia este tipo de investigación de otras aplicaciones de las ciencias sociales en el área de la salud. Este artículo presenta un modelo conceptual que resuelve las contradicciones al contextualizar y extender la definición y describe los principales diseños de investigación operativa.

Definiciones

Naturaleza de los programas

Antes de entrar en materia conviene delimitar los campos temáticos, comenzando por el de planificación familiar. Aunque en sus orígenes buena parte de los esfuerzos para implementar los servicios de planificación familiar en países en desarrollo respondió a una política de reducción de la natalidad con propósitos demográficos, y aun cuando muchos programas mantienen en el presente metas de esta naturaleza, este artículo trata de los servicios de anticoncepción diseñados para ayudar a los usuarios a realizar sus intenciones reproductivas, es decir, servicios que se ofrecen a las parejas o personas que desean limitar permanentemente su fecundidad o aplazarla. Estos servicios incluyen la entrega de los anticonceptivos, que va desde la venta de condones hasta las intervenciones quirúrgicas de esterilización, pero también abarcan la información, la consejería y el seguimiento a los usuarios. La investigación de operaciones tiene como propósito incrementar la eficiencia, la efectividad y la calidad de los servicios de planificación familiar que los proveedores brindan, así como la disponibilidad, la accesibilidad y la aceptabilidad de los servicios según los deseos de los usuarios actuales o potenciales (Fisher y cols., 1991).

El siguiente concepto por abordar es el de salud reproductiva, del que existen varias definiciones. Las tradicionales están formuladas en términos puramente médicos. Así, los trabajos presentados en el Seminario sobre Enfoques Innovativos para la Evaluación de la Salud Reproductiva que se realizó en la Universidad de Filipinas en 1996 se centraron en la *morbilidad reproductiva* que los programas de salud debían evitar o reducir.

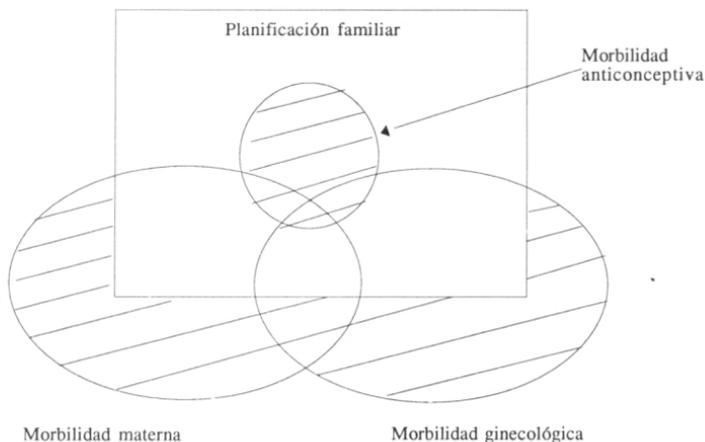
En la morbilidad reproductiva se distinguen tres categorías (Fortney, 1996). Una es la morbilidad materna —que en grado extremo se convierte en mortalidad— y contiene a su vez tres subcategorías: 1) la morbilidad obstétrica directa, que se origina en el embarazo o el parto (abortos, hemorragias, infecciones, etc.), aunque puede presentarse en el puerperio; 2) la morbilidad obstétrica indirecta, que se refiere a condiciones que no son causadas por el embarazo pero que la gestación empeora, como la hepatitis, hipertensión crónica o cáncer de mama; 3) la morbilidad materna no obstétrica que abarca eventos como depresiones o suicidios de madres solteras cuyo origen es el embarazo.

La morbilidad ginecológica abarca las enfermedades de transmisión sexual (incluyendo el SIDA), las infecciones del tracto reproductivo, los cánceres del sistema reproductivo (mamas, cuello uterino, etc.), y otras

morbilidades, como la infertilidad, las disfunciones sexuales, etc. Por último, la morbilidad anticonceptiva es aquella causada por el uso de métodos anticonceptivos y puede ser local (por ejemplo, infecciones) o sistémica (por ejemplo, impacto cardiovascular).

La figura 1 ofrece un modelo gráfico de las relaciones entre planificación familiar y morbilidad reproductiva. La planificación familiar reduce la morbilidad materna al prevenir embarazos, especialmente en mujeres con riesgo reproductivo. La planificación familiar, sin embargo, no es aplicable cuando la mujer quiere el embarazo. En este caso, la morbilidad materna se reduce mediante otros esfuerzos de prevención (*v.g.*, reforzamiento nutricional), el buen manejo de las emergencias obstétricas y el parto institucional. Otro efecto positivo de la planificación familiar ocurre en la intersección con la morbilidad ginecológica. Aquí se incluyen los problemas ginecológicos que pueden prevenirse directamente con métodos anticonceptivos, como el uso del condón y las enfermedades de transmisión sexual, así como los problemas ginecológicos originados en la morbilidad materna y que la anticoncepción puede prevenir. La morbilidad anticonceptiva, por definición, es causada enteramente por la planificación familiar y puede tener repercusiones en la salud materna y la ginecológica. La anticoncepción, por supuesto, permite a las parejas gozar de su sexualidad manteniendo el tamaño de familia deseado. En este sentido, va más allá de la prevención de las morbilidades materna y ginecológica.

Figura 1
Relaciones entre planificación familiar
y componentes de la morbilidad reproductiva



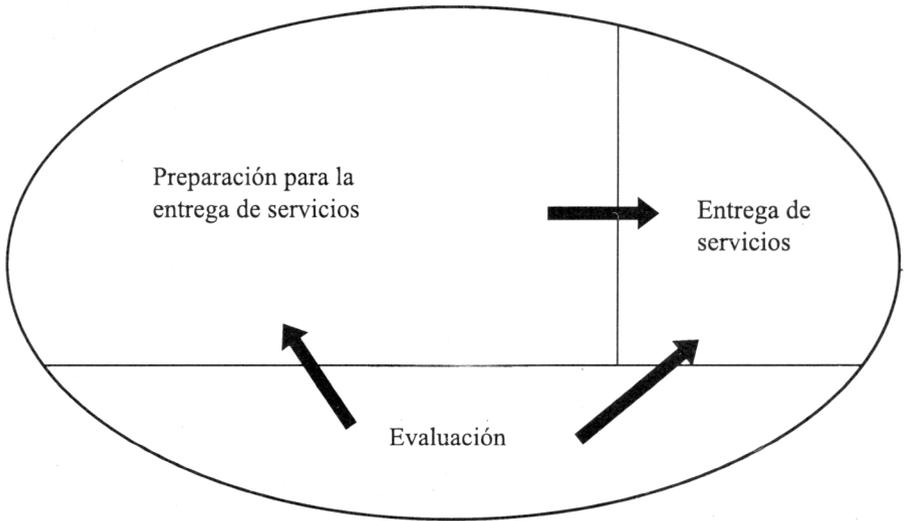
Este modelo, basado en el concepto de morbilidad reproductiva, ofrece una definición mínima de salud reproductiva. La Décima Conferencia Internacional sobre Población y Desarrollo (ICPD, 1994) ofreció una definición más amplia, entendiendo la salud reproductiva como "un estado de completo bienestar físico, mental y social, y no sólo la ausencia de enfermedad, en todos los aspectos relacionados con el sistema reproductivo". Esta definición pone especial énfasis en la educación sexual y el empoderamiento de la mujer como medios para la realización de las intenciones reproductivas, la prevención de la morbilidad reproductiva, el logro de una sexualidad plena y el desarrollo de la mujer como persona.

Los programas de salud establecen sus metas, formulan sus estrategias y organizan sus operaciones siguiendo definiciones estrechas o amplias de salud reproductiva. Asimismo, pueden ser tan pequeños como el programa de salud reproductiva de una organización no gubernamental (ONG) boliviana cuya clientela no pasa de algunas docenas de familias o tan grandes como el de planificación familiar del gobierno de China, que atiende a millones de personas a través de miles de establecimientos.

Operaciones de los programas

El *Handbook of Indicators for Family Planning Program Evaluation* (Bertrand y cols., s/f) considera cinco clases de operaciones de entrega de servicios: gerencia, capacitación, suministros y logística, información-educación-comunicaciones (IEC), e investigación y evaluación. En realidad, la mayor parte de estas funciones organizacionales son de preparación del programa para la entrega de servicios. Las operaciones programáticas, en efecto, pueden ser clasificadas en tres clases: de preparación para la entrega de servicios, de entrega de servicios y de evaluación de ambas (figura 2). Estas categorías corresponden casi totalmente a la infraestructura, los procesos y los resultados en el esquema de calidad de atención de Donabedian (1988). Al definirse la entrega concreta de servicios como una operación, se pone el foco en lo esencial y se pueden manejar con mayor coherencia conceptos como los de IEC y calidad de atención. Por ejemplo, se puede distinguir entre la preparación del programa para dar un servicio de calidad (los equipos y suministros están en su sitio, el personal tiene los conocimientos técnicos, etc.) y la calidad de atención que se plasma en la interacción individual entre un proveedor y una usuaria (la cantidad y calidad de la información que fluye, la ausencia de sesgos en la consejería, etc.). Esta conceptualización también permite tratar como servicios los esfuerzos de IEC dirigidos a la comunidad.

Figura 2
Tipos de operaciones de los programas



Lo que busca la investigación operativa es modificar las operaciones preparatorias de los programas con el fin principal de mejorar la entrega de servicios y sus resultados. Por ejemplo, el Instituto Peruano de Seguridad Social quiso mejorar el acceso de las mujeres a la anticoncepción en el post-parto inmediato. Esto requirió un conjunto de acciones institucionales, como la adquisición de equipo y suministros para la inserción del DIU en el post-parto, la reorganización de los servicios por pisos del hospital, la capacitación del personal en la técnica de inserción, su capacitación en técnicas de consejería de planificación familiar, etc. Los cambios introducidos en estos niveles preparatorios mejoraron la calidad de atención y produjeron una mayor aceptabilidad de la anticoncepción por parte de las mujeres en el post-parto (Foreit y cols., 1993).

Diversos tipos de operaciones de planificación familiar y salud reproductiva y sus productos han sido objeto de estudio en el Perú. Por ejemplo, Foreit y cols. (1990) estudiaron el costo y efectividad de diversas frecuencias de visita por parte de equipos itinerantes de prestación de servicios anticonceptivos en puestos comunitarios de INPPARES. Del Valle y cols. (1993) evaluaron la infraestructura de servicios de planificación familiar en el ámbito nacional, incluyendo aspectos de capacitación del personal, disponibilidad de suministros en los establecimientos, etc.

León y cols. (1994) ayudaron al Ministerio de Salud en la introducción del inyectable Depo-Provera en ámbitos rurales, poniendo a prueba la efectividad de prestadores no profesionales de servicios bajo nuevos sistemas de capacitación y logística. Benson y cols. (1998) modificaron los procedimientos de atención del post-aborto en un hospital, logrando reducir el tiempo y costo de la atención y aumentando la satisfacción de las usuarias. Los informes finales de INOPAL (1990, 1995) y Fisher y cols. (1991) ofrecen otros ejemplos de Perú y el mundo.

Resultados inmediatos y mediatos de las operaciones

La figura 3 ofrece una visión más global de las aplicaciones y metas de la investigación operativa. Como se dijo, ésta tiene que ver con cambios introducidos en las operaciones preparatorias con el propósito de mejorar los programas.

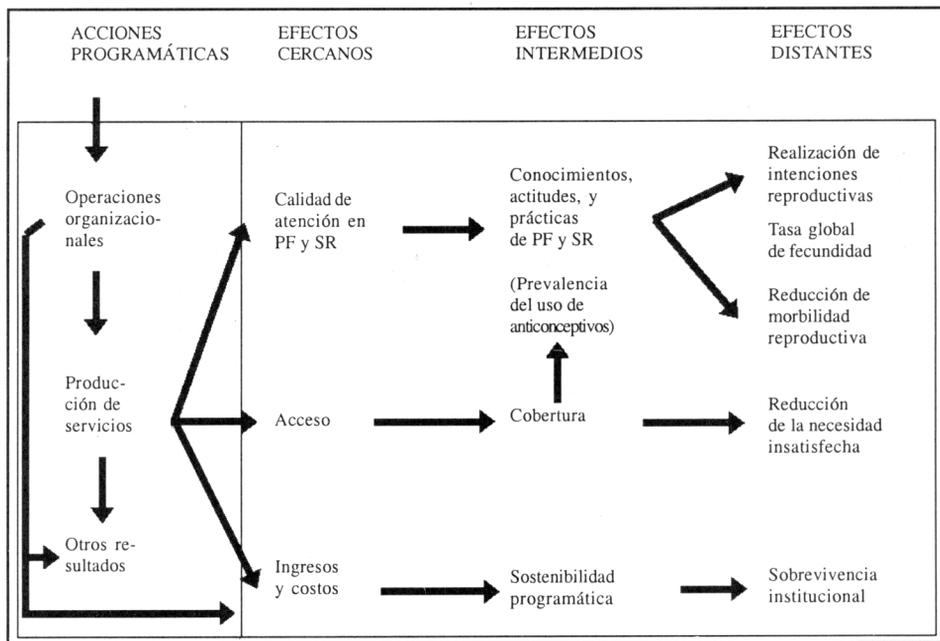
La primera columna de la figura se refiere a las acciones programáticas. La investigación operativa puede limitarse a establecer si una intervención programática que modifica operaciones organizacionales tiene un efecto en la preparación para dar servicios (por ejemplo, si una nueva organización logística hace llegar oportunamente equipos y suministros a los establecimientos). O si un nuevo sistema de capacitación causa mejoras en la cantidad y calidad de los conocimientos obtenidos por los proveedores de servicios. También los investigadores pueden interesarse en efectos independientes de las operaciones de entrega de servicios, como, por ejemplo, establecer si un nuevo esquema de captación de afiliados institucionales está generando un mayor número de afiliados en la Organización No Gubernamental.

La siguiente columna de la figura 3 se refiere a los efectos cercanos, que corresponden a los “productos” en la clasificación de indicadores de Bertrand y cols. (s/f). Lo que más interesa mejorar es, generalmente, la calidad de atención, el acceso y la relación de ingresos a costos. Por ejemplo, en Paraguay, CEPEP introdujo un instrumento algorítmico de consejería que reducía el costo de capacitación sin desmejorar la calidad de atención (Carrón y cols., 1994). En la tercera columna están los efectos intermedios, que a veces se estudian como efectos directos de las intervenciones programáticas, omitiendo constatar los efectos cercanos. Por ejemplo, se capacita a los proveedores y luego se observa si las usuarias de los servicios han mejorado sus prácticas anticonceptivas, sin verificar si el proveedor realmente ha dado un servicio mejor a la usuaria. La mayoría de los “indicadores de prácticas anticonceptivas” de Bertrand y cols. son efectos intermedios.

Finalmente, en la cuarta columna están los efectos más distantes, que incluyen los “impactos” del sistema de indicadores de Bertrand y cols. Todavía no se ha realizado un estudio que abarque toda la cadena de causas. El proyecto ideal introduciría mejoras y determinaría si ello afecta la calidad de atención, si las usuarias modifican sus prácticas anti-conceptivas y, si, finalmente, realizan en mayor medida sus intenciones reproductivas.

La conceptualización del acceso (v.g., Bongaarts y Bruce, 1995; Ross, 1995) no está tan diferenciada como la de la calidad de atención (Bruce, 1990). En cuanto a los problemas de sostenimiento económico y sobrevivencia institucional, éstos son más agudos en ONG's que en los programas del sector público. Sin embargo, los costos tienen que ver con la eficiencia en ambos.

Figura 3
Aplicaciones y metas de la investigación operativa



La figura 3 muestra las tendencias temáticas contemporáneas mayores de la investigación operativa en planificación familiar y salud reproductiva. Es probable que en los próximos años el campo se enriquezca al abordarse otras temáticas, congruentes con la reciente am-

pliación del concepto de salud reproductiva (la definición de la ICPD). Por ejemplo, en un Seminario sobre las Relaciones entre el Microcrédito y la Fecundidad, que organizó el Population Council (Amin, 1997), varios estudios evaluaron esta vinculación asumiendo que el microcrédito empoderaba a la mujer y traía como consecuencia cambios en su comportamiento reproductivo. Los estudios fueron esencialmente evaluativos, pero se puede esperar que en el futuro desarrollen enfoques proactivos que resulten en propuestas de mejoramiento de las operaciones de microcrédito.

Tipos de estudios sobre operaciones

El *Handbook for Family Planning Operations Research Design*, publicado en segunda edición por el Population Council (Fisher y cols. 1991) y traducido al español y al francés, se refiere a tres tipos de estudio concernientes a las operaciones de los programas. En primer lugar están los estudios diagnósticos exploratorios que: “buscan establecer los parámetros de una situación problemática... examinan los factores subyacentes... [y] son de un diseño retrospectivo o transversal. Se necesita este tipo de estudio en el momento que se percibe un problema programático, pero no se conoce bien la naturaleza del problema. Un aspecto clave de estos estudios es la búsqueda de variables manipulables por el programa” (Fisher y cols., 1991, pp. 1-2).

Un ejemplo es el de Cobián y Reyes (1998), incluido en esta serie. Antes de realizarlo, las autoridades de salud de la provincia de Santa, Perú, sabían que los hombres de Chimbote y pueblos aledaños no se acercaban a los establecimientos de salud para solicitar servicios de planificación familiar ni acompañaban a sus esposas a la consulta; sin embargo, no tenían ideas precisas sobre las barreras limitantes. Cobián y Reyes descubrieron, mediante entrevistas con hombres en grupos focales, que estos concebían los servicios de salud del sector público en términos contrarios a los de su propia autoimagen. Los veían como dirigidos a niños, mujeres y ancianos, es decir, a los débiles y vulnerables y no a personas eficaces como ellos mismos. Los comentarios de los participantes sugirieron que acudirían en busca de información y, eventualmente, servicios, si variara el contexto de la oferta (si disminuyeran, por ejemplo, los elementos femeninos dominantes en los servicios).

El segundo tipo de investigación de operaciones que presenta el *Handbook* es el estudio de evaluación: “Muy frecuentemente se implementan actividades de salud y planificación familiar durante años, pero nunca se les evalúa. En tales casos, los estudios de evaluación pueden representar un excelente enfoque de investigación de opera-

ciones para examinar transversal o retrospectivamente los efectos de las actividades de los programas” (Fisher y cols., 1991, p. 2).

Un ejemplo de investigación de evaluación sobre operaciones es el estudio de continuidad en el uso de anticonceptivos de Ramos y cols. (1998) en el Perú. Ellos hicieron un seguimiento de usuarias que habían iniciado el uso del DIU, pastilla o inyectable en 1995 en establecimientos del Ministerio de Salud en la provincia de Santa. Las tasas de continuidad calculadas a los 6, 12 y 24 meses ofrecen retroalimentación a los proveedores de servicios de la provincia sobre lo que ocurre con las usuarias después de la consulta. (Para una descripción más detallada, véase el Documento de Trabajo núm. 19, *Continuidad de la planificación familiar: un seguimiento de usuarias de establecimientos públicos de la provincia de Santa*, de esta serie.)

No siempre son claras las diferencias entre el estudio exploratorio y el estudio de evaluación. Por lo general, el primero tiene como propósito profundizar en el entendimiento de un problema programático antes de actuar sobre él, mientras que el estudio de evaluación describe una situación de manera que permite comparar lo observado con lo esperado.

Finalmente, están los estudios de intervenciones de campo: “Estos estudios prueban, sobre una base experimental, nuevos enfoques para resolver un problema del programa. En muchas situaciones se conocen los factores responsables del problema, pero no las formas más eficientes y costo-efectivas de aliviarlo. Estos estudios son siempre prospectivos y longitudinales, y usualmente emplean un diseño de investigación experimental o cuasi experimental” (Fisher y cols., 1991, p. 2). También emplean series de tiempos no prospectivas y diseños pre-experimentales, como el antes-después sin grupo de control.

El estudio de Velásquez y cols. (1998), incluido en esta serie, ejemplifica este tipo de estudio. En él se identificó la necesidad de mejorar el funcionamiento del programa (dar mayor uso a la infraestructura de servicios de salud reproductiva de una clínica limeña, más conocida por sus servicios de planificación familiar). Se formuló una solución (informar mediante un folleto interactivo a los visitantes de la clínica sobre la existencia de servicios de salud reproductiva), se puso a prueba (mediante un experimento verdadero), y se verificó el logro de mejoras (incremento del uso de los servicios).

Definición y ejemplos de investigación operativa

Según el *Handbook* (Fisher y cols., 1991), se puede abordar las operaciones de los programas mediante distintos tipos de estudios (diagnósticos,

evaluaciones, intervenciones), pero hay una sola manera de hacer investigación operativa: "La investigación operativa es un proceso, una manera de identificar y resolver problemas de los programas. Tal como se le aplica actualmente en programas de salud, planificación familiar y otros programas de desarrollo, la investigación operativa puede definirse como un proceso continuo de cinco pasos básicos...", que son los que muestra la figura 4. La investigación operativa es propuesta como una forma de investigación-acción que dota a los programas de instrumentos de cambio e innovación.

Figura 4
Los cinco pasos de la investigación operativa
según Fisher y cols., 1991, p.1

- Identificación y diagnóstico del problema
- Selección de la estrategia de solución
- Experimentación y evaluación de la estrategia
- Divulgación de la información
- Utilización de la información

Fisher y cols. (1991) ofrecen criterios para cada paso. El siguiente ejemplo ilustra los significados de la figura 4 en la práctica concreta de la investigación operativa. Se trata del estudio sobre frecuencia de puestos comunitarios realizado en el Perú por Foreit y cols. (1990).

■ *Paso 1 de Foreit y cols.: Identificación y diagnóstico del problema.* A mediados de la década de 1980, la ONG INPPARES tenía un programa de distribución comunitaria de anticonceptivos (DCA) en áreas urbano marginales de Lima. Las promotoras vendían pastillas, condones y espermicidas y guardaban una modesta porción del pago. El apoyo médico a las promotoras se ofrecía en puestos comunitarios, y las camionetas de INPPARES se encargaban de traer a los médicos. Los servicios se ofrecían en un espacio donado por la comunidad. Los médicos hacían inserción de DIU y su seguimiento y comprobaban en las usuarias los efectos secundarios y contraindicaciones de los anticonceptivos orales. Durante su primer año (1984), el componente médico produjo resultados modestos que fueron atribuidos a la baja visibilidad de las postas (funcionaban irregularmente y, a menudo, su ubicación variaba de una sesión a otra) y a una reticencia de las distribuidoras a hacer referencias para inserción de DIU.

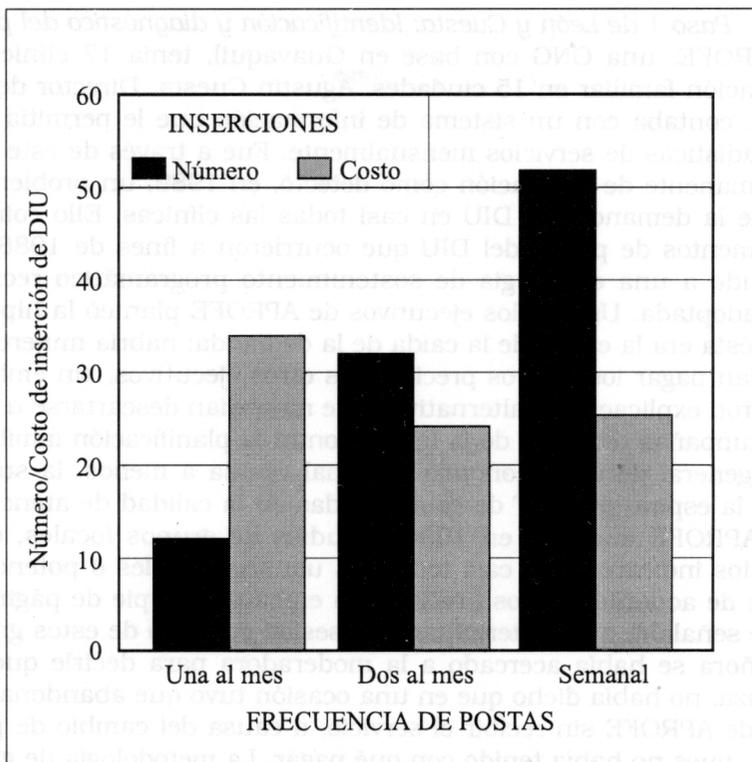
■ *Paso 2 de Foreit y cols.: Selección de la estrategia de solución.* Para superar estas dificultades, INPPARES capacitó promotoras en consejería de DIU y les dio un pequeño pago por cada referencia a la posta. Las promotoras traían a las mujeres para la inserción del DIU y recibían el pago por referencia al final de la sesión. Las sesiones de posta se realizaban ahora en locales fijos a un horario determinado. Sin embargo, no se sabía cuál era la frecuencia más efectiva, eficiente y efectiva en costos de las sesiones de posta por mes. INPPARES decidió probar tres frecuencias: una vez al mes, dos veces al mes y una vez por semana.

■ *Paso 3 de Foreit y cols.: Experimentación y evaluación de la estrategia.* Las tres frecuencias fueron comparadas en un estudio realizado en 42 postas. Se utilizó un diseño aleatorio por bloques. Las 42 postas se agruparon en 14 bloques de 3 postas cada uno, igualadas por desempeño previo, número de promotoras y tamaño de la población distrital. Se fijó un número igual de charlas sobre planificación familiar para cada posta y los médicos fueron asignados aleatoriamente a las tres frecuencias. El experimento duró 12 meses, desde agosto de 1985 hasta julio de 1986. Los indicadores de producción fueron tres. La *efectividad* se definió como el número total de servicios realizados durante el año del experimento. La *eficiencia* se definió como el número medio de servicios llevados a cabo por sesión. La *efectividad de costos* se definió como la proporción de costos netos asociados a la operación de una posta (los costos totales menos el ingreso de la posta) respecto a su propia producción. Los resultados mostraron que la frecuencia tenía un gran impacto en los tres indicadores. Dependiendo del indicador, las sesiones realizadas dos veces al mes producían entre 1.5 y 2.1 veces más que el resultado de aquellas realizadas una vez al mes. Las sesiones semanales producían entre 1.3 y 1.6 veces más que aquellas realizadas dos veces al mes. A un nivel de producción de casi 11,200 visitas por año, se estimó que las sesiones realizadas dos veces por mes eran de 7 a 38% más efectivas en costos, dependiendo del indicador, que las sesiones realizadas una vez al mes, y de 6 a 28% más efectivas en costos que las sesiones semanales. La gráfica 1 presenta datos relevantes respecto de las inserciones de DIU.

■ *Paso 4 de Foreit y cols.: Divulgación de la información.* Los resultados de los estudios pueden presentarse a los administradores en diversos formatos. En el presente caso, los investigadores tuvieron varias sesiones informales de análisis de los hallazgos con los administradores y luego realizaron una presentación formal en un seminario al que

Gráfica 1

Número promedio de inserciones de DIU por posta y costo neto de inserción (en dólares), según frecuencia de apertura de la posta en la comunidad, INPPARES, 1985-1986



fueron invitadas personas interesadas de diversas instituciones, además del personal de INPPARES. La gerencia de INPPARES decidió instalar en lo sucesivo la frecuencia quincenal de puestos.

■ *Paso 5 de Foreit y cols.: Utilización de la información.* La utilización se define como el nivel de cambio programático que resulta de los esfuerzos de investigación. Cerca de seis años después del seminario final del proyecto, Mostajo y León (1992) visitaron las postas de INPPARES y encontraron que la frecuencia quincenal seguía en pie en la mayoría; es decir, INPPARES continuaba utilizando la solución que había emergido del estudio como la más recomendable.

Evidentemente, los proyectos de investigación operativa no se desarrollan de manera idéntica. El siguiente ejemplo, tomando materiales de un estudio de León y Cuesta (1993) y agregando información antes inédita, hace notar las diferencias que pueden presentarse en cada paso.

■ *Paso 1 de León y Cuesta: Identificación y diagnóstico del problema.* APROFE, una ONG con base en Guayaquil, tenía 17 clínicas de planificación familiar en 15 ciudades. Agustín Cuesta, Director de Evaluación, contaba con un sistema de información que le permitía obtener estadísticas de servicios mensualmente. Fue a través de este sistema permanente de evaluación como detectó, en 1989, un problema: la caída de la demanda del DIU en casi todas las clínicas. Ello coincidía con aumentos de precio del DIU que ocurrieron a fines de 1988, respondiendo a una estrategia de sostenimiento programático recientemente adoptada. Uno de los ejecutivos de APROFE planteó la hipótesis de que ésta era la causa de la caída de la demanda: habría mujeres que no podían pagar los nuevos precios. Los otros ejecutivos, sin embargo, plantearon explicaciones alternativas que no podían descartarse *a priori*, como campañas recientes de la Iglesia contra la planificación familiar, el estado general de una economía nacional venida a menos, la sustitución de la espiral por la T de cobre, caídas en la calidad de atención, y otras. APROFE encargó, en 1990, estudios de grupos focales, cuyos resultados indicaron que casi todas las usuarias, reales o potenciales, estaban de acuerdo con los precios. Sin embargo, un pie de página del informe señalaba que, al terminar una sesión con uno de estos grupos, una señora se había acercado a la moderadora para decirle que, por vergüenza, no había dicho que en una ocasión tuvo que abandonar una clínica de APROFE sin recibir el servicio, a causa del cambio de precio del DIU, pues no había tenido con qué pagar. La metodología de grupos focales no había permitido que fluyera esta información y, por ende, no se había establecido cuántas mujeres estuvieron en una situación similar. Nótese cómo APROFE utilizó un estudio de evaluación (la estadística de servicios de su sistema de información gerencial) y un estudio exploratorio (los grupos focales) en este proceso. Hay situaciones, sin embargo, donde la identificación y diagnóstico del problema se efectúan sin contar con estudios tan formalizados como éstos.

■ *Paso 2 de León y Cuesta: Selección de la estrategia de solución.* El grupo ejecutivo de APROFE analizó detenidamente una estrategia de solución que consistió en fijar diferentes tasas de aumento de precio en las clínicas, para identificar el límite de tolerancia de la población de usuarias. Así, se podría identificar el aumento de precio que maximizara

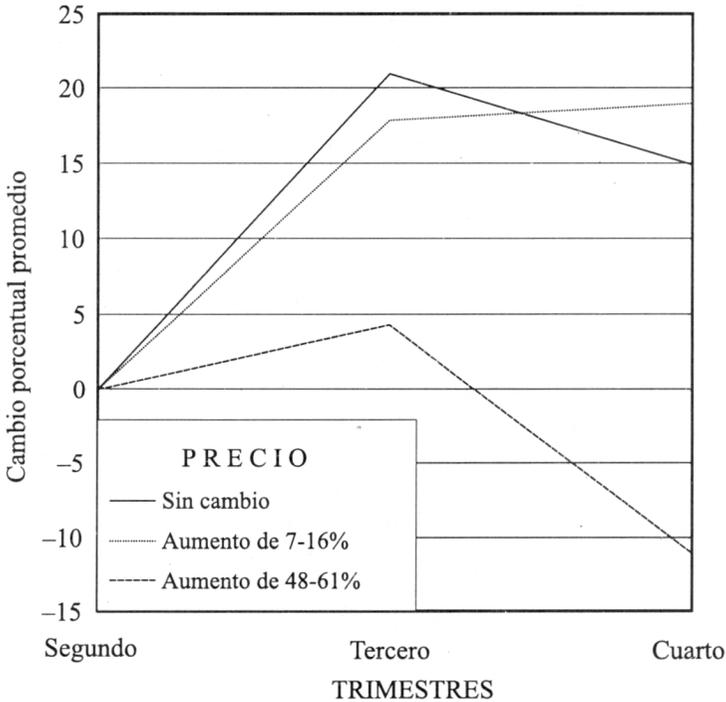
los ingresos institucionales sin disminuir la demanda. La estrategia, sin embargo, fue descartada momentáneamente por considerar que no había evidencias suficientes en favor de la hipótesis del precio como causa de las caídas observadas en la demanda. Entre tanto, APROFE encargó más estudios diagnósticos con el propósito de profundizar en la verificación de las otras hipótesis (estudios económicos y administrativos, más grupos focales, etc.). Nótese la diferencia de actitud frente a la estrategia considerada. Mientras que INPPARES sólo esperaba los resultados del experimento para optar por una de las tres frecuencias de posta, APROFE mantenía dudas sobre la naturaleza misma del problema.

■ *Paso 3 de León y Cuesta: Experimentación y evaluación de la estrategia.* La estrategia propuesta, y ante la cual APROFE había tenido inicialmente dudas, se implementó de manera natural. A mediados del año 1991, el Director Ejecutivo dejó en los directores de clínicas la responsabilidad de fijar los precios durante un periodo que les serviría para observar resultados. Seis clínicas mantuvieron el precio del DIU cobrado desde 1990, siete lo elevaron entre 7 y 16%, y cuatro lo aumentaron entre 48 y 61%. Aquí hay otra diferencia respecto al procedimiento de INPPARES, que estableció las opciones en forma vertical. León y Cuesta observaron el impacto del precio en la demanda al comparar los cambios producidos entre los tres grupos de clínicas del trimestre previo a la intervención (segundo del año) con los dos siguientes a ella. Los resultados se confundieron con un efecto estacional, pero dejaron ver claramente los efectos del precio (véase la gráfica 2). El grupo que introdujo aumentos de 48 a 61% presentó un incremento muy pequeño de la demanda en el tercer trimestre del año, en comparación con los importantes aumentos registrados para los otros grupos, y la caída de la demanda en el cuarto trimestre fue mucho más marcada para este grupo de clínicas. Al comparársele con los otros dos grupos, se pudo concluir que, de no haber efectuado un incremento tan fuerte del precio, podrían haber tenido un flujo de usuarias comparable al de los otros grupos. Un estudio de serie de tiempos en la principal clínica de APROFE, que cubría 24 trimestres, permitió observar que la demanda del DIU fluctuaba sistemáticamente en función del precio.

■ *Paso 4 de León y Cuesta: Divulgación de la información.* A principios de 1993, APROFE desarrolló un seminario en Guayaquil en el que fueron expuestos los resultados descritos arriba y otros, incluyendo tanto evidencias empíricas adicionales sobre el efecto del precio en la demanda, como hallazgos de los estudios diagnósticos (grupos focales, en-

Gráfica 2

Cambio porcentual promedio en la demanda de DIUs en tres grupos de clínicas de APROFE, de acuerdo con el cambio de precio del DIU, 1991



cuentas domiciliarias, estudios administrativos y económicos) que se habían realizado para evaluar las hipótesis alternativas (calidad de atención, etc.). Ninguna de éstas fue apoyada por los datos como lo fue la de precios, y los cuatro miembros del grupo ejecutivo de APROFE salieron de ese seminario convencidos de que los precios estaban directamente relacionados con la caída de usuarias nuevas de DIU observada desde 1989.

■ *Paso 5 de León y Cuesta: Utilización de la información.* La modificación de las creencias de los ejecutivos dio como resultado una reformulación del problema institucional de sostenimiento programático. Se hizo evidente para el equipo ejecutivo que su estrategia a base del aumento de precios debía tener en cuenta el impacto en la demanda y su significado respecto a su misión de dar servicios a una población pobre. Posteriormente, el Proyecto INOPAL III del Population Council ha realizado con ellos investigaciones de fijación sistemática de precios,

con un criterio de incremento de ingresos que minimice la pérdida de clientela. Nótese que en este caso la utilización de los hallazgos del experimento es menos operativa, algo ambigua en comparación con la de Foreit y cols. (1990). INPPARES instaló una frecuencia específica de postas que se ha mantenido por varios años. APROFE sólo decidió tener en cuenta el impacto del precio en la demanda en sus reformulaciones de la estrategia institucional de sostenimiento financiero.

Experimentación y utilización

La experimentación cumple un papel esencial en la investigación operativa por su capacidad para demostrar a los administradores de programas los efectos de las operaciones programáticas. Los estudios descriptivos, como los diagnósticos exploratorios y las evaluaciones no experimentales, carecen de este poder de convencimiento. Así, por ejemplo, si APROFE, en lugar de poner a prueba la estrategia de precios a través de un experimento natural, se hubiese limitado a profundizar el examen de las diversas hipótesis mediante estudios diagnósticos, tal vez hasta hoy continuaría dudando sobre la naturaleza de las relaciones entre precios y demanda del DIU. La experimentación, al ofrecer al administrador la oportunidad de introducir él mismo los cambios en las operaciones programáticas y observar las consecuencias, otorga un nivel de veracidad y confianza en los resultados, sin par en la metodología científica.

Por supuesto, esto no significa que los diagnósticos y evaluaciones carezcan de lugar en la investigación operativa. Como se vio en el caso de APROFE, ambos jugaron un papel importante en el proceso de identificación y diagnóstico del problema. La investigación operativa, sin embargo, puede dejar de lado los estudios descriptivos formales cuando los administradores poseen información e intuición suficientes para formular y seleccionar una o varias estrategias de acción. No puede prescindir, en cambio, de la experimentación y evaluación de la estrategia, al menos según la definición de Fisher y cols. (1991).

La otra dimensión clave de la definición de investigación operativa es la de utilización de resultados por el programa. En la dicotomía de investigación básica versus investigación aplicada, la investigación operativa se sitúa claramente en la segunda categoría, pues se diseña para generar hallazgos importantes para la toma de decisiones administrativas del programa en cuyo contexto se desarrolla. Si la investigación está bien planteada, sus resultados serán necesariamente particulares, relevantes y prácticos, en contraste con los de la investigación básica, que apunta a la generación de conocimientos más

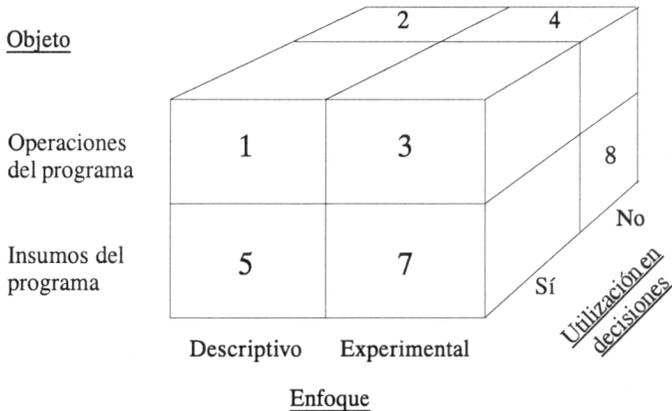
generales y ofrece resultados relevantes o irrelevantes para una situación programática específica, aunque haya estado perfectamente bien planteada respecto a sus propios fines. Por definición, la particularidad, relevancia, y practicidad de los hallazgos originados en procesos de investigación, como los que resume la figura 4, debieran favorecer la utilización de los resultados por parte del programa.

Modelo de investigaciones en la perspectiva del administrador

Ross y cols. (1987) objetaron la definición de investigación operativa presentada aquí. Consideraron que los estudios no deben limitarse a resolver problemas sino también a aprovechar oportunidades. Defendieron la relevancia de los estudios descriptivos por sí mismos en la toma de decisiones de los administradores y rechazaron, por ende, el paso 3 como obligatorio en la investigación operativa. Por último, arguyeron que, en ocasiones, los resultados de investigaciones quedan sin utilización práctica por razones fuera del control de los investigadores o administradores y no por eso deben perder la clasificación de "operations research".

El modelo tridimensional que se propone en la figura 5 para definir tipos de investigación sobre planificación familiar y salud reproductiva permite contextualizar el debate y resolverlo. El modelo se formula desde la perspectiva del usuario de las investigaciones, es decir, del administrador del programa.

Figura 5
Modelo de investigaciones sobre planificación familiar y salud reproductiva



Si nos quedamos por ahora en el examen del piso superior del modelo, bastará tomar en cuenta las dimensiones de enfoque y utilización. Para Fisher y cols. (1991), sólo los estudios que involucran al cubo 3 pertenecen a la categoría de investigación operativa, pues, además de referirse a operaciones de un programa, incluyen la experimentación con una estrategia y generan resultados que causan cambios programáticos. Según Ross y cols. (1987), en cambio, los cuatro cubos del piso superior del modelo calificarían.

Los estudios descriptivos (diagnósticos exploratorios, evaluaciones) relativos a las operaciones preparatorias o de entrega de servicios de los programas son, por definición, formas de investigación sobre operaciones, sea que constituyan o no el primer paso de los cinco requeridos por la definición de investigación operativa de Fisher y cols. (1991). Y al igual que los experimentales, los estudios descriptivos pueden tener una utilización práctica (cubo 1) o no tenerla (cubo 2). Ejemplo del primer caso es el análisis situacional que condujeron Mensch y cols. (1994) en 46 clínicas del programa de planificación familiar de la Comisión de la Ciudad de Nairobi. Los investigadores evaluaron varios aspectos de las operaciones preparatorias del programa, detectando falta de esterilizadores, escasez de estetoscopios y otras limitaciones para la entrega técnicamente apropiada de servicios (por ejemplo, se observó una incongruencia entre el volumen de usuarios y el de personal). Al conocer los hallazgos preliminares del estudio, las autoridades de salud de la ciudad dieron pasos inmediatos para mejorar los servicios: se movilizó hacia las clínicas equipos que estaban almacenados y enfermeras que trabajaban en clínicas sobredotadas de recurso humano, etc. Aquí se tiene un estudio descriptivo (paso 1), sobre la preparación de las clínicas para dar servicios, que condujo directamente a la divulgación (paso 4) y utilización de los resultados (paso 5). Evidentemente, no es indispensable dar todos los pasos de la investigación operativa para mejorar los programas.

No es difícil encontrar ejemplos del cubo 2, es decir, casos de falta de utilización de resultados de estudios descriptivos concernientes a operaciones programáticas. Snowden y cols. (1988) trabajaron con una muestra representativa de médicos de Lima, examinando sus conocimientos sobre métodos naturales de planificación familiar (calendario, Billings, temperatura), es decir, evaluaron su preparación para dar información de buena calidad al respecto, lo que era relevante para el Programa Nacional de Planificación Familiar de un país como Perú, donde el ritmo era el método más utilizado de todos. Los autores hallaron que aun aquellos médicos que dispensaban estos métodos en su práctica clínica tenían importantes vacíos de conocimientos, y recomendaron su capacitación sistemática, operación que ha quedado

pendiente hasta la fecha debido a las dificultades prácticas de esta solución y a su baja prioridad para el Programa. En contraste, pues, con el estudio de Mensch y cols. (1994), que encaja en el cubo 1 del modelo, al de Snowden y cols. (1988) le corresponde el cubo 2.

El estudio de Foreit y cols. (1990) sobre frecuencia de visitas de los equipos médicos itinerantes a las puestos comunitarios de INPPARES es uno de los mejores ejemplos de investigación operativa (cubo 3). Finalmente, en el cubo 4 caben los experimentos sobre estrategias programáticas que no llegaron a ser divulgados y/o plenamente utilizados. Por ejemplo, Profamilia introdujo en Lima el uso de hombres promotores de planificación familiar como parte de un proyecto de INOPAL (Foreit y cols., 1992). Al poco tiempo de terminado el proyecto, esta modalidad de acercamiento a la comunidad se fue diluyendo hasta extinguirse. Es decir, no hubo utilización real de los resultados después del experimento.

Demostrada la validez empírica de los cubos del 1 al 4, ya es posible retornar a la controversia entre Fisher y cols. (1991) y Ross y cols. (1987). Lo que propone el autor del presente artículo para resolverla es una aproximación semántica. Es indiscutible que los cuatro tipos de estudios representan subtipos de investigación de operaciones, pues todos abarcan operaciones programáticas o conciernen directa o indirectamente a la preparación del programa o su entrega de servicios. El cubo 3, sin embargo, se diferencia de los otros tres porque se refiere a estudios que, además de involucrar operaciones, operan cambios programáticos en dos instancias: durante el estudio, con el fin de conocer los resultados que traerían los cambios, y después, en la fase de institucionalización de los cambios (utilización). Para reconocer este doble énfasis operativo, se puede limitar a los estudios del cubo 3 la clasificación de investigación operativa.

Insumos de los programas

Refirámonos ahora al primer piso del modelo. En la perspectiva del administrador del programa, hay investigaciones que no estudian directamente las operaciones del suyo y, sin embargo, conciernen a la planificación familiar y/o salud reproductiva. Los hallazgos que brindan pueden ser de mucha relevancia para el mejoramiento de las operaciones programáticas. Conviene, por ende, reconocer su valor y diferenciarlas con claridad de la investigación relativa a las operaciones del programa. La mejor manera de hacerlo es concibiendo su objeto de estudio como insumos analizados que trascienden el contexto específico del programa. Los insumos de los programas son básicamente de dos tipos. Por un

lado están los físicos, como los métodos anticonceptivos, los equipos de laboratorio, los medicamentos para el tratamiento de las morbilidades reproductivas, y el resto de variables físicas manejadas en los servicios. El segundo tipo de insumos principales de los programas es la gente, que el programa convierte de personas no atendidas en usuarias de los servicios.¹

Es evidente que las investigaciones sobre insumos, más allá del contexto de un programa particular, pueden ser de gran relevancia para las decisiones de los administradores. Considérense los estudios epidemiológicos como la última Encuesta de Demografía y Salud realizada en Perú (Reyes y cols., 1997), que mostró cambios dramáticos en el porcentaje de adolescentes de 15 a 19 años que son madres en Lima metropolitana, el cual subió en más de 50% entre 1991 y 1996. O el estudio de Sánchez y cols. (1996), de mujeres que solicitaban certificados en dos establecimientos de salud de Lima norte, que halló que más de 33% mostraba anticuerpos respecto a tres enfermedades de transmisión sexual. Estos tipos de información tienen muchas implicaciones para los administradores de programas de planificación familiar y salud reproductiva y pueden generar decisiones administrativas inmediatas o sugerir investigaciones operativas.

Los estudios clínicos que se realizan con el propósito de mejorar el conocimiento sobre la planificación familiar y la morbilidad reproductiva también son relevantes. Considérese la nueva revisión de la literatura de Koblinsky (1995) sobre la relación entre el estado nutricional de la madre y los resultados del embarazo en países en desarrollo. Niveles moderados de anemia, como los detectados recientemente por Reyes y cols. (1997) en el Perú, se agravan con el embarazo y reducen la capacidad de la madre para resistir infecciones y sobrevivir a hemorragias, aumentando cuatro veces la probabilidad de mortalidad materna.

Finalmente están los estudios sociales, como el antropológico de Maynard-Tucker (1989) en una aldea de Cusco, donde encontró que los hombres tenían un conocimiento más exacto de los órganos reproductivos femeninos que las mujeres, pero éstas conocían mejor los efectos de los anticonceptivos modernos en el cuerpo humano.

Los estudios epidemiológicos, clínicos y sociales, al referirse a insumos del programa y no a sus operaciones, están excluidos del concepto de investigación de operaciones. Sin embargo, pueden ser muy útiles para identificar problemas operacionales que requieren soluciones y pueden generar procesos de investigación operativa.

Los estudios epidemiológicos, clínicos y sociales son investigaciones descriptivas, es decir, transversales o retrospectivas y, por ende, pertenecen a los cubos 5 o 6 del modelo presentado en la figura 5,

según sean o no utilizados sus hallazgos. Pero también se abordan los insumos de los programas con un enfoque experimental (cubos 7 y 8). Por ejemplo, en la búsqueda de un anticonceptivo hormonal masculino, Aravindan y cols. (1996) han conducido experimentos con un modelo animal de mecanismos genéticos de acción de los andrógenos, hormonas que regulan la producción de esperma. Luego están los ensayos clínicos con humanos. Por ejemplo, Rosenberg y cols. (1996) insertaron aleatoriamente la T de cobre 380-A en algunas mujeres, y en otras, un nuevo DIU llamado Cu-Fix e hicieron un seguimiento durante 14 meses, encontrando un mejor desempeño de la T de cobre. Además de los estudios experimentales sobre anticonceptivos, se han realizado experimentos sobre el comportamiento de las personas, y los resultados obtenidos son relevantes para los programas de salud interesados en temas reproductivos. Los psicólogos sociales norteamericanos, por ejemplo, frecuentemente estudian los estereotipos de género con un enfoque experimental (v.g., Conway y cols., 1996).

Las dos grandes categorías de estudios, los de operaciones del programa y los de insumos del mismo, difieren en el grado de generalización. La investigación de operaciones se centra en la realidad concreta e inmediata de un programa singular y, regularmente, no pretende hacer generalizaciones más allá del contexto singular estudiado. En este sentido, debe considerarse una forma de investigación estrictamente aplicada. En cambio, los estudios sobre insumos, trátense de anticonceptivos o de personas, suelen aspirar a mayores generalizaciones. Considérese, por ejemplo, una de las publicaciones recientes del *Collaborative Group on Hormonal Factors in Breast Cancer* (1996). Se trata de una comparación de 53,297 mujeres que desarrollaron cáncer de mama y 100,239 que no lo desarrollaron en 54 estudios realizados en 25 países. Sus conclusiones son relevantes para cualquier programa de planificación familiar.

Requerimientos de la investigación de operaciones

La investigación de operaciones (cubos del 1 al 4) se distingue de otras formas de aplicación de las ciencias sociales al campo de la salud, por su énfasis en dos dimensiones de significado: el objeto de estudio y la utilización de los resultados. Quien conduce los estudios es generalmente un investigador que trabaja en estrecha colaboración con los administradores del programa y debe tener en cuenta tales significados al conducir este tipo de estudios.

Comencemos por el criterio de *utilización*. Para que el estudio sea

práctico, tiene que ser entendido cabalmente por los administradores del programa, pues son ellos quienes tomarán las decisiones de utilización. Es improbable que un administrador implemente una estrategia que sólo entiende vagamente. A un científico social, el investigador puede decirle que va a usar “un enfoque constructivista para aproximarse a las representaciones sociales de los jóvenes”, pero a un administrador hay que hablarle en su lenguaje. Afortunadamente, muchos administradores de programas de planificación familiar y salud reproductiva son médicos u obstetras que han sido expuestos a cursos avanzados de administración y tienen familiaridad con el lenguaje técnico del área. Se les puede hablar de las creencias, actitudes y prácticas de las usuarias; la calidad de atención, la necesidad insatisfecha de planificación familiar, las tasas de continuidad de los métodos, el costo-efectividad de las estrategias, etc., sin necesidad de dar muchas explicaciones, pero también se les pueden presentar conceptos sociales complejos que, tal vez, sean nuevos para ellos. Si los conceptos son claros, se podrán presentar como esquemas similares a los de la teoría de conjuntos, los flujos de decisiones o alguna otra representación espacial o gráfica (Anderson, 1998). Este punto de vista sobre la necesidad de modelos conceptuales suficientemente claros como para ser representables gráficamente, trasciende las particularidades del abordaje metodológico del tema y se aplica por igual a estudios que utilizan técnicas cuantitativas o cualitativas.

El segundo requisito es ser prácticos: es definir el tema de la investigación en torno a alguna variable que esté realmente bajo el control del administrador. Hay situaciones que pueden ser difíciles para el administrador y que a él le gustaría cambiar, pero que están fuera de su control. Él tiene que operar sobre la realidad con todas sus limitaciones, y entre éstas sobresalen las del costo de las operaciones. En muchas ocasiones, el costo de la investigación, incluyendo la implementación de la solución nueva que va a ser probada, es asumido por financiamiento externo y, sin embargo, no se tiene en cuenta las implicaciones de este hecho. Si se va a efectuar una intervención que dará resultados maravillosos aplicando recursos cuantiosos, hay que preguntarse cuán realista es esperar que la solución propuesta sea luego realmente implementada por el programa con sus propios recursos. Será bueno distinguir aquí entre el poder del administrador para aceptar la aplicación de los cuantiosos recursos del proyecto con financiamiento externo y su poder para implementar luego la solución con sus propios recursos, que son generalmente limitados.

El tercer requisito es obtener resultados claros. Los programas traba-

jan con personas y éstas tienen intereses. Dentro del mismo programa se encontrarán, seguramente, ejecutivos que están a favor de una solución y otros que estarán en contra. Sólo presentando evidencias inobjetables se podrá evitar las discusiones interminables a las que dan lugar los resultados ambiguos en estos contextos. Al hablarse de evidencias inobjetables no me refiero necesariamente a los hallazgos científicos perfectos. Estos deben ser totalmente claros para el administrador, es decir, resultados que la gente práctica y bien intencionada consideraría sin objeción, como los de frecuencia de puestos de INPPARES o la investigación de precios de APROFE.

La dimensión del objeto de estudio trae una paradoja: la producción de conocimientos particulares plantea desafíos metodológicos mayores que la investigación básica. Cuando un grupo médico evalúa el efecto de una droga, comienza con parámetros muy claros en la definición del problema, tiene para escoger entre pocos diseños estandarizados de investigación, y sabe que la evaluación de sus resultados se hará en términos de criterios científicos universalmente compartidos. Sin embargo, cuando se hace investigación de operaciones no hay parámetros fijos. Casi ninguna situación problemática viene con sus variables bien definidas ni contiene todas las condiciones requeridas para aplicar los diseños de investigación descritos en los libros. De aquí que deba apelarse a la adaptación de los diseños de investigación a las circunstancias imperantes en el programa, generando hallazgos de interpretación más difícil. Generalmente, los casos no existen en número suficiente, la asignación aleatoria de casos a tratamientos es muchas veces imposible, la deserción de casos es un hecho difícilmente controlable, y la contaminación de variables es endémica en el campo. El desafío entonces es grande, y para afrontarlo hay que mantener una rigurosidad intrínseca, que no consiste en seguir reglas sino en aplicar la crítica.

Requerimientos de la investigación operativa

La investigación operativa (cubo 3) agrega una particularidad: la del enfoque metodológico. El método experimental nació con las ciencias naturales, y los administradores de programas generalmente lo encuentran perfectamente entendible y aceptable. Sin embargo, el experimentalismo se asocia en su origen con el positivismo y, actualmente, con el post-positivismo. Recibe, por esta razón, el rechazo de algunos científicos sociales, que lo consideran artificioso al adjudicarle la propiedad de destruir la integridad de la realidad social, a fin de que el investigador pueda manejar sus

variables con claridad. Es verdad que los experimentos tienen que centrarse en la manipulación de una o dos variables. Si el científico social cree que tiene que estudiar todas las variables que le parecen relevantes, jamás podrá realizar un experimento. Pero reducirse al estudio de unas pocas, no implica necesariamente introducir artificialidad. En realidad, las críticas no surgen del cuidadoso examen de las características de las investigaciones operativas concretas; provienen de concepciones abstractas sobre experimentos de laboratorio, de poca conexión con la realidad de los programas. Los experimentos de campo, es decir, los de la investigación operativa, tienden a ser bastante naturales, pero exigen que el investigador acepte los presupuestos metodológicos del método experimental.²

En segundo lugar, el investigador debe estar dispuesto a expresar en números los resultados finales del estudio. Los estudios cualitativos tienen un sitio bien ganado en el proceso de identificación y diagnóstico de problemas. Más aun, el paradigma post-positivista admite la realización de experimentos con variables cualitativas (Guba y Lincoln, 1994). Sin embargo, para obtener conclusiones prácticas, generalmente será necesario trabajar con variables cuantitativas o reducir la información cualitativa a números, aunque estos sean sólo nominales. Los administradores, por supuesto, agradecen la presencia de números.

Finalmente, el investigador debe entender cabalmente lo esencial del diseño experimental y la cuasi experimentación. De lo contrario, correrá el riesgo de conducir su proyecto sin prestar la debida atención a las amenazas a la validez de su experimento.

Principales diseños de investigación operativa

Quienes desean evaluar los efectos de programas o intervenciones de planificación familiar y/o salud reproductiva, pueden escoger entre diversos diseños de investigación. Las secciones finales de este artículo describirán los experimentos verdaderos, cuasi experimentos y pre-experimentos más importantes, y ofrecerán criterios para elegir el más apropiado según los recursos disponibles y las limitaciones de validez interna y externa del estudio. Fisher y cols. (1991) y Cook y Campbell (1979) describen otros diseños.

Experimentos verdaderos

Para entender el diseño 1, asumamos como ejemplo que, de 80 obstetras recién egresadas de universidades de provincias, 40 son asignadas al

azar [AA] a un Curso de Actualización Anticonceptiva (X) y las otras 40 quedan a la espera de capacitación en el futuro (Grupos experimental y de control, respectivamente).

Figura 6
Diseño 1. Experimento verdadero sólo con post-test

| | | | |
|------|--------------------|---|----------------|
| [AA] | Grupo experimental | X | O ₁ |
| | Grupo control | | O ₂ |

Tres meses después del curso, se evalúa la calidad de atención de los dos grupos (O₁ y O₂) y se encuentra una diferencia de 15 puntos que es estadísticamente significativa (figura 7). Será legítimo atribuir-la al curso porque, dada la asignación aleatoria de casos a tratamientos, que permite asumir la igualdad inicial de los grupos en cualquier variable cuando hay número suficiente de casos, la única explicación que queda de la diferencia observada en el post-test, es la del curso dado a uno de los grupos de obstetras.

Figura 7
Resultados después de la intervención

| | | Puntaje promedio de calidad de atención |
|---------------------------------|--------|---|
| Grupo experimental (con curso), | N = 40 | 55 |
| Grupo control (sin curso), | N = 40 | 40 |

Los experimentos conocidos como "verdaderos" (*true experiments*) son los clásicos de las ciencias sociales y del comportamiento, que asignan aleatoriamente unidades experimentales a tratamientos para formar grupos equivalentes antes de una intervención. Las unidades pueden ser sujetos individuales (proveedores de servicios, usuarias, etc.) o agregados de sujetos individuales (familias, cohortes de usuarias, distritos, etc.). El número de sujetos individuales requeridos depende de varias consideraciones estadísticas: los efectos mínimos que quiere detectar el investigador, el número de tratamientos, la varianza del error en la población, la probabilidad de cometer un error de tipo I y la probabilidad de cometer un error de tipo II (Kirk, 1968). Barcikowski (1981), Hsieh (1988), y Murray y Hannan (1990) ofrecen cálculos respecto al número necesario de agregados.

Cuando hay un número satisfactorio de casos a la luz de las consideraciones estadísticas, puede asumirse que los dos grupos, gracias a la asignación aleatoria [AA], tendrán situaciones iniciales idénticas. Así, el modelo puede limitarse a la intervención en el grupo experimental (X) y a las observaciones posteriores en el grupo experimental y el de control (O_1 , O_2), llamadas post-tests (diseño 1). Si las unidades experimentales son agregados y se les va a tener como unidades de análisis, se puede trabajar con un número muy pequeño de ellos porque, en este caso, los datos (promedio de cada cohorte de usuarias, etc.) serán más confiables que en el caso de los sujetos individuales. Por ejemplo, en el estudio de Velásquez y cols. (1998), publicado en esta serie, 10 cohortes diarias de usuarias (es decir, las usuarias que vinieron a la clínica en 10 días específicos) recibieron un tratamiento y fueron comparadas con éxito con 10 cohortes de control.

La efectividad de la asignación aleatoria para hacer equivalentes a los grupos depende, pues, del número de unidades experimentales y otros factores estadísticos. Al incrementarse el número de unidades, las diferencias individuales (o de agregados) se cancelan entre sí y la tendencia central de cada grupo se aproxima más cercanamente al valor poblacional común respecto a cualquier variable (experiencia previa de los proveedores, ingreso de las clínicas, etc.). Al introducir un tratamiento sobre uno de los grupos aleatorios, los investigadores asumen equivalencia previa, y entonces atribuyen a la intervención las diferencias grupales observadas más tarde en una variable dependiente. La validez interna de un experimento es la medida en que se justifica este juicio de causalidad (Campbell y Stanley, 1966).³

Además del requerimiento de validez interna, el experimento verdadero debe satisfacer una condición previa: la validez estadística de los hallazgos, es decir, la extensión y significación de la covariación entre el tratamiento y la variable dependiente. En la ausencia de validez estadística, el juicio de validez interna carece de significado (Cook y Campbell, 1979). En los ámbitos programáticos, donde los investigadores en general no pueden seleccionar aleatoriamente a los sujetos, su asignación aleatoria a los tratamientos es esencial si el investigador quiere hacer inferencias estadísticas concernientes a los efectos del tratamiento. La asignación aleatoria a tratamientos hace posible la inferencia estadística, aunque los sujetos no hayan sido seleccionados al azar de una población pre-especificada (Edington, 1966).

La figura 8 ofrece un ejemplo que clarifica el papel de la asignación aleatoria (*randomization*). Dos consejeros de planificación familiar son asignados aleatoriamente a una capacitación en relaciones interpersonales (proveedor 1) o una condición de control (proveedor 2), y tres meses

más tarde se realiza una encuesta de satisfacción entre sus usuarias nuevas (N = 200), en la que las usuarias del proveedor 1 reportan mayor satisfacción que las del 2 ($p < .05$). Es decir, se encuentra una correlación significativa entre la capacitación del proveedor y la subsecuente satisfacción de las usuarias. Aquí es evidente la covariación entre tratamiento y variable dependiente (fuerte validez estadística), pero no se puede inferir causalidad (débil validez interna), pues la asignación aleatoria de sólo dos unidades experimentales a los tratamientos (proveedor 1 y sus usuarias, proveedor 2 y sus usuarias) no hace equivalentes a los grupos. Los dos proveedores son distintos y las usuarias del proveedor 1 han tenido una selección, historia, etc., que las hace diferentes de las usuarias del otro. Por tanto, la diferencia observada en satisfacción de las usuarias podría deberse a diferencias pre-existentes en rasgos de los proveedores (y/o de las usuarias), en lugar de ser un efecto del tratamiento. Aquí la asignación aleatoria de grupos es irrelevante, sin importar cual sea su tamaño, y el diseño sigue siendo un pre-experimento (sólo post-test con grupos no equivalentes, Cook y Campbell, 1979).⁴

Figura 8
Bases para juicios de covariación y causalidad en dos diseños

| Diseño | Asignación aleatoria a tratamientos | Unidades de aleatorización | Unidades de medición |
|-----------------------|-------------------------------------|--|----------------------|
| Pre-experimento | Experimental | Proveedor 1 | |
| | Control | Proveedor 2 | |
| Experimento verdadero | Experimental | Proveedor 1 → Clientes 1-5 Proveedor 2 → Clientes 6-10 . Proveedor 20 → Clientes 96-100 | |
| | Control | Proveedor 21 → Clientes 101-105 Proveedor 22 → Clientes 106-110 . Proveedor 40 → Clientes 196-200 | |

Lo que cuenta es la asignación aleatoria de suficiente número de unidades experimentales a los grupos de tratamiento. La siguiente sección de la figura 8 presenta un experimento verdadero sólo con post-test, en el que se asigna aleatoriamente 40 proveedores a los tratamientos (experimental, control) y se encuentra después una correlación significativa entre tratamiento y satisfacción promedio de las usuarias por proveedor ($N = 40$). Dado el número de unidades experimentales asignadas aleatoriamente a los tratamientos, los investigadores sí pueden asumir la existencia de equivalencia previa a la intervención entre los grupos experimental y de control y, en consecuencia, pueden concluir que sus resultados satisfacen los requerimientos de validez interna y estadística.

Si hay un número pequeño de sujetos individuales y la varianza de la variable dependiente es muy grande, será improbable que los grupos experimental y control queden igualados en la situación inicial pese a la asignación aleatoria. De allí que sea necesario considerar un pre-test en ambos grupos (O_1 y O_3 ; véase diseño 2). Supongamos que las obstetras

Figura 9

Diseño 2. Experimento verdadero con pre-test y post-test

| | | | | |
|------|--------------------|-------|---|-------|
| [AA] | Grupo experimental | O_1 | X | O_2 |
| | Grupo control | O_3 | | O_4 |

del caso anterior son 20 y no 80. El pre-test servirá para establecer la medida en que la situación de inicio de los dos grupos de 10 es comparable. Pero, sobre todo, este diseño otorga la ventaja de que cada caso sirve de control para sí mismo y el análisis puede hacerse más fino, compensando por el bajo número de casos. Por ejemplo, puede usarse la diferencia pre-test - post-test o el porcentaje de cambio del pre-test al post-test computado para cada obstetra. En realidad, este es el experimento verdadero más completo, independientemente del número de casos.

Imaginemos que las obstetras de la muestra pequeña presentan resultados como los de la figura 10. En el ejemplo anterior, la diferencia de 15 puntos en el post-test entre los dos grupos de 40 obstetras tenía significación estadística. En el presente ejemplo, observamos una diferencia de sólo 4 puntos entre los dos grupos de 10 obstetras en el post-test y no alcanzaríamos significación estadística si nos limitáramos a comparar O_2 con O_4 . El porcentaje de cambio, que toma en cuenta el pre-

test, es un indicador más fino de los efectos. La diferencia de 27 puntos porcentuales en la tasa de cambio de los grupos experimental y control sí podría representar una diferencia significativa entre los grupos, confirmando el impacto del curso pese al pequeño número de casos.

Figura 10
Resultados antes y después de la intervención

| | | <i>Puntaje promedio de calidad de atención</i> | | |
|---------------------------------|--------|--|---------|----------|
| | | Antes | Después | % Cambio |
| Grupo experimental (con curso), | N = 10 | 34 | 46 | +35 |
| Grupo control (sin curso), | N = 10 | 39 | 42 | +8 |

El experimento con asignación aleatoria y pre-test no tiene rival en la metodología de evaluación de efectos, pero no es perfecto. La razón es que hay situaciones que hacen poco recomendable el pre-test. Cuando se usan mediciones reactivas, el pre-test puede afectar al post-test hasta el extremo de invalidarlo. Por ejemplo, si las clientes simuladas visitan repetidamente al mismo proveedor para evaluar su calidad de atención, el proveedor puede sacar ventaja de la primera experiencia para ajustar el nivel de calidad en el momento que sospeche que una mujer que solicita sus servicios no es realmente una cliente. En otras situaciones, el pre-test puede tener poco sentido. Por ejemplo, el estudio puede interesarse en varias cohortes diarias de usuarias nuevas de los servicios de una clínica; en este caso, la cohorte se forma el día en que las usuarias llegan a la clínica. Por ello, Velásquez y cols. (1998) utilizaron un experimento verdadero sin pre-test.

Por otro lado, si bien los experimentos verdaderos permiten descartar la mayoría de las amenazas a la validez interna del estudio, no permiten descartar todas, y los evaluadores de programas deben estar atentos a los factores o hipótesis alternativas a la de la intervención que podrían explicar los hallazgos. En el caso del experimento verdadero, con o sin pre-test, éstas son cuatro.

La primera por considerar es la *difusión de tratamientos*. Cuando la intervención involucra materiales informativos y los grupos experimental y de control pueden comunicarse entre sí, los de control pueden aprender la información dirigida al grupo experimental. El experimento, en consecuencia, puede invalidarse, pues los resultados, que indicarían en el post-test que el grupo de control aprendió tanto o casi tanto como el experimental, no estarían reflejando la relación de causa-efecto que

se quería probar sino otra, que resulta de la contaminación del grupo de control. La mejor defensa contra esta amenaza a la validez interna del experimento es separar físicamente a los grupos, como hicieron Velásquez y cols. (1998) al trabajar con cohortes de clientes que asistían a la clínica Padres de INPPARES en diferentes días.

La *igualación compensatoria* ocurre cuando el tratamiento experimental proporciona bienes o servicios deseables que no se ofrecen al grupo de control y los administradores u otros agentes sociales que no toleran la desigualdad intervienen para asegurar una distribución más "justa" de los bienes. Por ejemplo, cuando introdujimos Depo-Provera en áreas rurales del Perú, destinándolo sólo a proveedores experimentales seleccionados aleatoriamente (León y cols., 1994), hubo coordinadoras del Programa de Planificación Familiar que pasaban los inyectables al grupo de control, porque consideraban injusto dejar sin esta oportunidad de anticoncepción a las usuarias de esos proveedores. El riesgo se puede reducir explicando a los administradores de todos los niveles el objetivo de la intervención y empleando una supervisión apropiada que permita detectar a tiempo los problemas para corregirlos.

La *rivalidad compensatoria* consiste en una reacción competitiva de los miembros del grupo que recibe el tratamiento menos favorable con la que tratan de demostrar que ellos también pueden mejorar su rendimiento. Por ejemplo, las obstetras del grupo de control del caso presentado arriba, si saben que serán comparadas con las que recibirán el curso, pueden dedicarse a estudiar por su cuenta con mayor aplicación que la que pondrían en circunstancias normales. A causa de ello, se reducirán las diferencias esperadas de la intervención. Entonces se habrá distorsionado el experimento, pues la situación puesta a prueba no reflejaría con fidelidad el efecto del curso, sino que estaría combinado con una rivalidad compensatoria que no es de interés para los evaluadores. La rivalidad compensatoria es un artefacto del estudio, pues el grupo de control existe sólo en la prueba del curso, y lo que se tiene previsto, una vez que se pruebe su eficacia, es darlo a todas las obstetras, sin excepción. Es decir, una vez que el curso demuestre su eficacia, se dará a todos los proveedores y no habrá rivalidad compensatoria.

Finalmente, está la *desmoralización por tratamiento menos deseable*, que es la reacción opuesta al caso anterior. Si los miembros del grupo de control saben que sus colegas asignados al grupo experimental tendrán un tratamiento que es más interesante o del cual se esperan importantes mejoras, pueden resentirse o desanimarse y, en lugar de mantener su rendimiento normal, lo disminuyen. Si es así, las diferencias observadas entre los grupos experimental y de control en el

post-test serán parcialmente falsas, pues expresarán no sólo el efecto positivo del tratamiento sino también el efecto negativo de no dar el tratamiento al grupo de control.

El experimento verdadero, pues, pese a constituir la mejor metodología disponible para evaluar efectos, deja abierta la puerta a posibles distorsiones, y el evaluador debe estar atento para detectarlas y corregirlas.

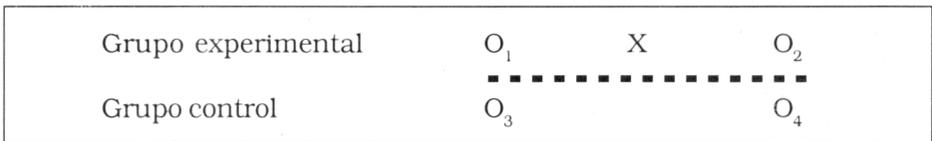
Un cuasi experimento

Hay situaciones que impiden la asignación aleatoria de casos a tratamientos y sólo permiten trabajar con grupos cuyos casos ya están autoseleccionados. Esto ocurrió con el estudio de León y Cuesta (1993), descrito anteriormente, en el que los tres grupos de clínicas establecieron por sí mismas diferentes niveles de precio para el DIU. Los grupos no fueron conformados aleatoriamente, y ese fue entonces un cuasi experimento.⁵

Volviendo al ejemplo de las obstetras, supongamos que, de las 80 del caso presentado anteriormente, 40 residen en una ciudad (digamos, Iquiura) y 40 en otra (Icaúcho). El Programa Nacional de Planificación Familiar ha dispuesto dar el curso sólo en una de las dos ciudades en el presente año, pues no hay recursos para hacerlo en ambas. Si existen recursos de evaluación para observar la calidad de atención en las dos ciudades, es decir, para efectuar pre-test y post-test en ambas, es aplicable a esta situación el diseño de grupo de control no equivalente con pre-test (diseño 3, donde la línea punteada indica que los casos no se asignaron aleatoriamente). Este es, probablemente, el diseño que con más frecuencia se usa en la investigación operativa.

Figura 11

Diseño 3. Grupo de control no-equivalente con pre-test



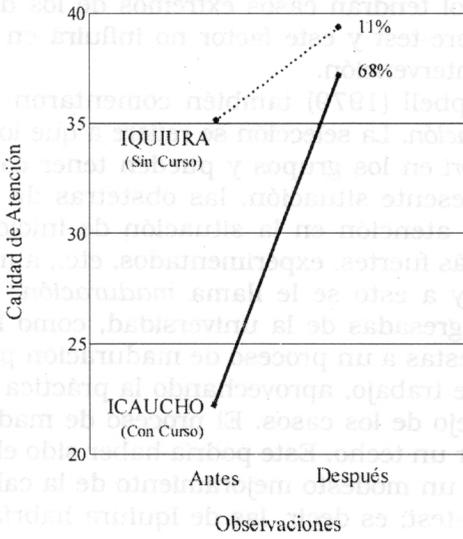
Asumamos que el programa selecciona al azar una de las ciudades y el resultado es que Icaúcho recibirá el curso e Iquiura quedará como control. El que se haya escogido la ciudad al azar no convierte este diseño en un experimento verdadero. Los grupos se han constituido obedeciendo a factores históricos que los hacen diferentes y, por ende, no se

puede asumir que sean equivalentes en la situación de inicio. Recuerde-se que lo que iguala a los grupos es la asignación aleatoria de los casos, no de los grupos, a los tratamientos. Sea que se seleccione al azar el grupo que recibirá el tratamiento o que se le asigne de otra manera, éste es un cuasi experimento.

Pese a la ausencia de asignación aleatoria de casos a tratamientos, el cuasi experimento con grupo de control no equivalente y con pre-test, ofrece resultados generalmente interpretables, aunque hay que estar alerta a las posibles distorsiones. La gráfica 3 presenta un juego de resultados de interpretación difícil, según el cual las obstetras de Icaúcho ofrecen servicios de bastante menor calidad que las de Iquiura en el pre-test. Esto podría atribuirse a diferencias en el curriculum de estudios de las universidades de las que provienen los dos grupos. Lo que más interesará a los evaluadores, sin embargo, es constatar que las de Icaúcho, luego de tomar el curso, mejoran sustancialmente en el post-test (68%) mientras que las de Iquiura mejoran poco (11%). Con un resultado como éste, probablemente atribuirán al curso dado en Icaúcho la diferencia observada en la tasa de cambio entre los grupos.

El grupo de control no equivalente con pre-test está expuesto a los cuatro problemas de validez interna discutidos en referencia al experimento verdadero y otros cinco que Cook y Campbell (1979) identifica-

Gráfica 3
Diseño de grupo de control no-equivalente: resultados de interpretación difícil



ron. Al primero lo llaman *historia local*, refiriéndose a los eventos específicos que afectan al grupo experimental, pero no al de control o viceversa. Imaginemos que ha tenido lugar un Congreso de Obstetricia en Icaúcho poco después del curso pero antes del post-test y al que no asistieron las obstetras de Iquiura. El gran mejoramiento de la calidad de atención observada en Icaúcho puede haber reflejado el efecto del curso, el efecto del congreso o el efecto combinado de ambos. Con este diseño no sabremos cuál fue la causa de esa mejora. Los experimentos verdaderos están mejor protegidos contra las hipótesis alternativas de tipo histórico. La asignación aleatoria de casos a tratamientos habría obligado a que las obstetras de Icaúcho se distribuyeran igualmente en los grupos experimental y de control, y el congreso habría afectado a los dos grupos por igual, permitiendo detectar en la diferencia de tasas de cambio entre los grupos el efecto de la capacitación.

Otra amenaza a la validez interna del diseño de grupo de control no equivalente con pre-test es la de *regresión estadística diferencial*. Mientras menos confiables sean las mediciones, mayor es la probabilidad de que los casos extremos positivos o negativos del pre-test cambien acercándose al promedio. Es decir, los puntajes muy altos caerán en el post-test y los puntajes muy bajos subirán. Si los datos de Icaúcho en el pre-test estaban muy por debajo del promedio general, es de esperar que subiesen hacia el promedio en el post-test a causa de la regresión estadística, aunque el curso no hubiese tenido efecto alguno. Así pues, podría ser un error atribuir al curso el mejoramiento de la calidad. En el experimento verdadero, dada la asignación aleatoria, los grupos experimental y de control tendrán casos extremos de los dos tipos en igual proporción en el pre-test y este factor no influirá en la evaluación de resultados de la intervención.

Cook y Campbell (1979) también comentaron la *interacción de selección y maduración*. La selección se refiere a que los casos están seleccionados *a priori* en los grupos y pueden tener características distintivas. En la presente situación, las obstetras de Icaúcho exhiben menor calidad de atención en la situación de inicio. Pero los casos pueden devenir más fuertes, experimentados, etc., a medida que transcurre el tiempo, y a esto se le llama *maduración*. Por ejemplo, las obstetras recién egresadas de la universidad, como las de este ejemplo, estarán expuestas a un proceso de maduración profesional en sus primeros meses de trabajo, aprovechando la práctica clínica para perfeccionar su manejo de los casos. El proceso de maduración, sin embargo, puede tener un techo. Este podría haber sido el caso de Iquiura, cuya curva indica un modesto mejoramiento de la calidad de atención de pre-test a post-test; es decir, las de Iquiura habrían mejorado sólo

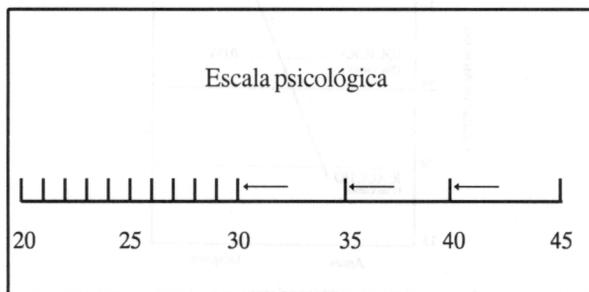
un poco porque ya estaban a un nivel alto de calidad de atención. En cambio, habiendo partido de un nivel inicial más bajo, las de Icaúcho habrían tenido más techo para mejorar. Según la hipótesis alternativa, habrían alcanzado un nivel cercano al de las iquiuranas sólo por efecto del tiempo de maduración, sin necesidad del curso. Nuevamente, el experimento verdadero estará mejor defendido en este aspecto porque los grupos generalmente serán equivalentes en la situación de inicio.

En cuarto lugar está el riesgo de *instrumentación diferencial*, o calibración del instrumento de medición de la variable dependiente. Podría ocurrir que la escala de medición de la calidad de atención sea más sensible en un segmento de la escala y menos sensible en otro. Por ejemplo, podría ser más fácil para los observadores de la calidad de atención usados en la evaluación discriminar niveles de calidad en el segmento bajo de la escala que en el segmento alto. Tal vez puedan discriminar 10 niveles de calidad entre los 20 y 30 puntos y sólo tres entre los 30 y los 45 y, en este último segmento, anclar el intervalo en su límite inferior. Así, por ejemplo, el 34 que no pueden discriminar se convertiría en 30 (véase la gráfica 4, que ofrece un modelo visual de esta hipotética situación). El incremento real podría haber sido esencialmente el mismo en los dos grupos, pero el factor métrico causaría la apariencia que tienen las curvas en la gráfica 3. Los experimentos verdaderos están menos expuestos a este tipo de problema porque la situación de inicio de los dos grupos tiende a ser igual.

Finalmente, la *mortalidad diferencial* es una diferente tasa de abandono del estudio por parte de los casos de los dos grupos. Imaginemos que en Iquiura logramos post-tests de 95% de los casos, pero en Icaúcho sólo podemos evaluar la calidad de atención de 30% de las obstetras en el post-test y que esto se debe a que las autoridades de

Gráfica 4

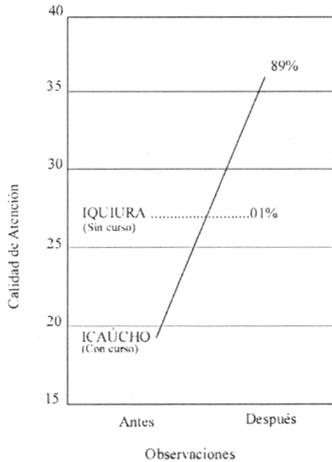
Modelo visual de la capacidad hipotética de los evaluadores para discriminar a distintos niveles de la escala de calidad de atención



Icaúcho decidieron dar de baja a las de rendimiento inaceptablemente bajo, que eran la mayoría. Su desaparición del estudio podría provocar que el post-test de Icaúcho refleje ahora una alta calidad de atención, pues sólo quedaron las mejores para el post-test. Nótese que este nivel ya existía antes del curso en el subgrupo de buenas obstetras que sobrevivieron en el empleo, pero en el pre-test había sido promediado con el de las menos competentes. En conclusión, no habría sido el curso, sino la mortalidad diferencial, el responsable de la notable mejora que muestra la curva de Icaúcho. La mortalidad diferencial es menos probable en los experimentos verdaderos.

Estos cinco problemas de validez interna pueden afectar al diseño de grupo de control no equivalente con pre-test, y el evaluador debe estar atento a ellos. El ejemplo de la gráfica 3 fue escogido para ilustrar los cinco problemas de validez interna y las dudas que nos dejaría la interpretación de los resultados, y también para que se entienda mejor cuáles son las ventajas de la asignación aleatoria de casos a tratamientos, es decir, las ventajas de los experimentos verdaderos. Sin embargo, la mayoría de los resultados del diseño de grupo de control no equivalente con pre-test tienen otro aspecto, que facilita su interpretación. Por ejemplo, si los resultados hubiesen sido como los de la gráfica 5, se podría concluir con mayor seguridad que el curso produjo un efecto positivo, pues las cinco hipótesis alternativas carecerían de fuerza. El cuasi experimento de León y Cuesta (1993) tam-

Gráfica 5
 Diseño de grupo de control no-equivalente:
 resultados más interpretables



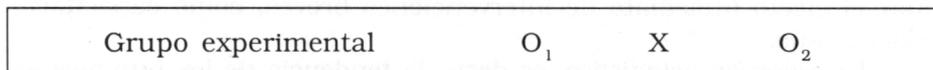
bien produjo resultados interpretables en su porción relativa al grupo de control no equivalente con pre-test (véase la gráfica 2).

El diseño de grupo de control no equivalente es bastante eficaz y puede rendir resultados generalmente interpretables y válidos. Recientemente, Heinsman y Shadish (1996) compararon los resultados de 47 cuasi experimentos psicológicos de este tipo con 51 experimentos psicológicos verdaderos. Concluyeron que, cuando está bien diseñado e implementado, el grupo de control no equivalente ofrece resultados similares a los del experimento verdadero.

Un pre-experimento

El escenario puede dejar menos opciones aún al evaluador. Imaginemos que sólo hay recursos para capacitar personal y evaluar los resultados en Icaúcho. Sin grupo de control a la mano, el programa podría tener que recurrir a un pre-experimento, como el estudio de grupo único con pre-test y post-test (diseño 4), que es uno de los diseños de investigación más usados para evaluar los efectos.

Figura 12
Diseño 4. Grupo único con pre-test y post-test

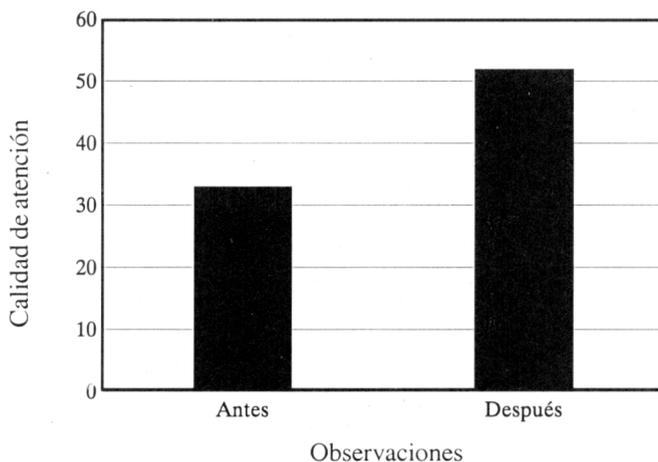


Los resultados podrían tener el aspecto de los presentados en la gráfica 6. Este diseño ha sido muy criticado por los metodólogos porque es vulnerable a múltiples factores de distorsión en la interpretación de resultados. Dada esta situación, mis comentarios se orientan a perfilar las circunstancias específicas en que el grupo único con pre-test y post-test podría usarse con beneficio pese a sus severas limitaciones generales.

El factor de *historia* ya fue definido. Se trata de eventos específicos que podrían afectar al grupo único causando las diferencias observadas entre pre-test y post-test, como el congreso realizado poco después del curso en Icaúcho. Para descartar una hipótesis alternativa de tipo histórico, se puede argumentar que no es factible para el contexto particular del estudio. Si es factible, se argumenta que hay evidencias de que realmente no operó. Si nada de esto es posible, habrá que aceptar que la hipótesis alternativa no se puede descartar. Para evitar verse en esta situación, el evaluador que usa el grupo único con pre-

Gráfica 6

Resultados en Icaúcho antes y después de la intervención



test y post-test haría bien en diseñar estudios de corta duración. Al reducir el tiempo entre el pre-test y el post-test, disminuirá la probabilidad de que acontezcan eventos contaminantes. Esta es la razón por la que este pre-experimento es generalmente apropiado para evaluar el efecto inmediato de intervenciones breves, como capacitaciones en aula.

La *regresión estadística*, es decir, la tendencia de los puntajes extremos de un pre-test a situarse más cerca del promedio en el post-test, puede controlarse evitando seleccionar casos extremos para el estudio. Por ejemplo, sería una pésima idea usar el grupo único con pre-test y post-test para evaluar la capacitación dada a personal seleccionado por su mal rendimiento. La tendencia del grupo será a presentar un mejor resultado en el post-test debido a la regresión al promedio y se tendrá el riesgo de interpretar el resultado como un efecto del curso.

La *maduración* es una fuente de hipótesis alternativas que puede controlarse añadiendo observaciones antes y después de la intervención, es decir, convirtiendo este diseño (un pre-experimento) en una serie de tiempos (un cuasi experimento). Otra estrategia para debilitar este tipo de hipótesis alternativa es considerar una duración del estudio menor que lo que necesita el fenómeno evolutivo para manifestarse. Así, por ejemplo, podrá atribuirse a un curso evaluado en semanas una mejoría de conocimientos que se sabe que tomaría meses para obtenerse de manera espontánea. La maduración no necesariamente debe verse como lineal. También existen tendencias cíclicas, como las

estacionales que León y Cuesta (1993) descubrieron que afectaban la demanda del DIU en Ecuador.

La experiencia de haber sido sometido al pre-test puede tener consecuencias. Por ejemplo, si se trata de un test de conocimientos, la persona se familiarizará con las preguntas y logrará un mejor resultado en el post-test. Para descartar un efecto falso, el investigador puede emplear en el post-test una prueba equivalente a la usada en el pre-test, es decir, un instrumento de estructura similar, pero que no tenga exactamente las mismas preguntas. Esta es la razón por la que el Test de Conocimientos de Planificación Familiar (León, 1995) tiene dos formas equivalentes.

También se presenta la *interacción de las pruebas con el tratamiento*. Tomemos el caso del investigador que, al prepararse para evaluar el impacto de una campaña masiva para hacer más atractivos a la población los servicios de planificación familiar, conduce como pre-test una encuesta a domicilio sobre el uso de servicios de planificación familiar. Las preguntas formuladas al encuestado pueden despertar su curiosidad respecto a los servicios y hacerle prestar más atención a la campaña masiva, con el resultado de movilizarlo a solicitarlos. Es decir, el pre-test habrá interactuado con la campaña y, sin embargo, los efectos probablemente serán interpretados como efectos directos y únicos de la campaña. Para evitar este riesgo se puede apelar a mediciones no reactivas, es decir, sistemas de recolección de datos que no afectan al sujeto, como datos de archivos y otros registros de sus acciones.

La *interacción de tratamiento y mortalidad* se produce cuando la intervención misma causa una deserción selectiva entre los participantes en el estudio. Por ejemplo, un curso difícil puede desalentar a los participantes menos hábiles, llevándolos a desertar. El post-test, limitado a los participantes sobrevivientes, puede presentar puntajes más altos que el pre-test a causa de la mayor habilidad de los sobrevivientes y no necesariamente como efecto del curso. En este caso, para ser más apropiado, el análisis debería limitarse a la comparación pre-test/post-test de los sobrevivientes.

Otro riesgo es el cambio de *instrumentación* entre el pre-test y el post-test. Por ejemplo, los observadores de la calidad de atención pueden ser muy exigentes y calificar bajo en el pre-test, pero a medida que ganan experiencia desarrollan estándares más laxos. Es decir, darán mejores calificaciones en el post-test. La recomendación obvia será la de usar instrumentos de observación más confiables y válidos.

Si bien el grupo único con pre-test y post-test puede ser útil en algunas circunstancias, deja abiertos muchos flancos. Por esta razón, siempre que sea posible, será preferible usar otros diseños. El grupo

de control no equivalente con pre-test, teniendo sus propias debilidades, está mucho mejor protegido. En esta jerarquía de diseños, por supuesto, el experimento verdadero con pre-test tiene una posición superior a la de cualquier otro en virtud de su solidez metodológica.

Validez externa

La validez externa de un experimento consiste en el grado de legitimidad para generalizar los resultados a otras situaciones, contextos, o poblaciones. El juicio de validez interna de un experimento siempre será relevante en la investigación diseñada para evaluar efectos, pues en esta clase de investigación siempre estaremos haciendo una inferencia de causa-efecto. En cambio, el juicio de validez externa sólo es relevante cuando el investigador quiere generalizar sus resultados más allá del ámbito del estudio realizado.

La investigación operativa se desarrolla en cualquiera de dos marcos relativos a la aplicabilidad de los hallazgos. Puede hacerlo en un marco limitado, definido por la situación específica del estudio que constituye el foco de interés; por ejemplo, se quiere determinar si un mejoramiento logístico introducido en la red de servicios de salud de la ciudad de Puno realmente sirvió para que las usuarias sean mejor atendidas. Aquí no pretendemos generalizar los resultados más allá del contexto en el que se ejecutó el estudio, lo cual es legítimo. Aquí, entonces, no hay problemas de validez externa. Pero hay situaciones que no interesan tanto por sí mismas como en virtud de su carácter representativo de situaciones más generales a las que quisiéramos extender los resultados. Por ejemplo, podríamos querer probar en Tumbes, Huancavelica y Pucallpa unas guías de salud reproductiva con el propósito de confirmar que mejoran la calidad del servicio dado por los proveedores en todo el país. La generalización a nivel nacional de nuestros hallazgos en estas tres ciudades puede bordear la legitimidad o puede estar completamente injustificado.

Figura 13

Comparación entre la validez interna y externa de un experimento

| <i>Validez interna</i> | <i>Validez externa</i> |
|-------------------------|------------------------------|
| ■ Experimento verdadero | ■ Muestreo representativo |
| ■ Cuasi experimento | ■ Muestreo de heterogeneidad |
| ■ Pre-experimento | ■ Grupo disponible |

Los problemas de validez externa de un experimento tienen una estructura similar a los de validez interna, en la medida en que para ambos casos contamos con modelos teóricos rigurosos que no siempre es posible seguir (véase la figura 13). Obtendremos la mayor validez interna si usamos un experimento verdadero; la validez interna de nuestro estudio estará generalmente comprometida si usamos un cuasi experimento; y sólo en circunstancias especiales podremos generar conclusiones válidas si usamos un pre experimento para evaluar efectos.

En el caso de la validez externa de un experimento, la máxima rigurosidad estará dada por 1) la definición precisa de un universo o población a la cual queremos aplicar las conclusiones del estudio; 2) la obtención de una muestra representativa de esa población, *v.g.*, una muestra aleatoria; y 3) la conducción del experimento en esa muestra de casos. Por ejemplo, nuestro universo podría definirse por el total de obstetras graduadas que trabajan en puntos de entrega de servicios del Ministerio de Salud. Pero igualmente podremos definir universos de subsistemas organizacionales, distritos geográficos, u otras unidades.

¿Qué ocurre si no podemos obtener una muestra estrictamente representativa? Cook y Campbell (1979) propusieron un modelo de *muestreo deliberado de heterogeneidad*. Aquí lo que se busca es definir, para el universo de interés, diversas clases de casos y lograr que la muestra contenga un amplio rango de instancias de cada clase. Por ejemplo, si no podemos obtener una muestra aleatoria de obstetras, deberemos tratar de conseguirla de diversas regiones del Perú, de diversas edades, etc. Nuestra certidumbre respecto a la validez de la generalización al universo evidentemente se reducirá, dependiendo de la naturaleza de la variable de interés, la fisonomía de los resultados y otros factores, pero en muchas ocasiones tendremos bases suficientes para formular conclusiones prácticas. ¿Qué pasa, por ejemplo, si encontramos tanto en Tumbes como en Pucallpa y Huancavelica que las obstetras logran aprender sólo 20% del contenido de un curso dado en un día, 60% de uno dado en 3 días, y 100% de uno dado en una semana? Las tres ciudades tal vez no sean estrictamente representativas del país, pero con esta consistencia de resultados nos quedarán pocas dudas y podremos recomendar que el curso se dé en una semana en todo el país si lo que se quiere es impartir 100% de los conocimientos. Evidentemente, correremos el riesgo de encontrar luego excepciones a la regla, pero tampoco se puede paralizar las decisiones programáticas porque el diseño no se ajusta perfectamente al modelo más riguroso. El administrador debe tomar decisiones entre alternativas concretas, y tomarlas con base en un estudio como éste siempre será mejor que tomarlas con base en intuiciones y preferencias personales.

Si no podemos siquiera escoger los grupos y debemos trabajar con los que estén disponibles, quedaremos ciertamente muy limitados en nuestra capacidad para generalizar los hallazgos. Quiero enfatizar que la validez externa de un experimento no es una cuestión rígida que sólo varía entre el SÍ y el NO. Siempre será una cuestión de juicio extrapolar las implicaciones de un estudio a otras situaciones. En las áreas de la investigación psicológica y médica se realizan estudios de causa-efecto bajo severas limitaciones de muestreo de casos y, sin embargo, contribuyen con hallazgos útiles y generalizables. Los psicólogos experimentales de las universidades norteamericanas toman como conejillos de Indias a los estudiantes que tienen a la mano y contribuyen al conocimiento publicando sus resultados en revistas científicas sumamente exigentes. Los médicos realizan investigaciones sobre nuevas drogas con voluntarios atraídos por un pago, presidiarios, etc., es decir, muestras sesgadas, y así hacen avanzar las fronteras del conocimiento. Estas, por supuesto, no son situaciones ideales, pero ilustran la posibilidad de efectuar aportes dentro de las limitaciones de la realidad que impiden ceñirse rígidamente a un modelo ideal.

Problemas de los experimentos verdaderos

Este capítulo, sin embargo, no debe terminar con una opinión que puede ser malinterpretada como invitación al libertinaje metodológico. Es preferible hacerlo con una nota de invitación a la rigurosidad.

Los investigadores están de acuerdo en que el experimento verdadero representa la mejor opción metodológica para evaluar efectos. Sin embargo, las investigaciones de planificación familiar y salud reproductiva presentan un uso muy escaso de experimentos verdaderos. Recientemente, Bauman y cols. (1994) hicieron notar que de 143 estudios de investigación operativa completados entre 1974 y 1986, sólo uno podía ser considerado un experimento verdadero. Si se revisa las publicaciones periódicas más conocidas del área, como *Studies in Family Planning* e *International Family Planning Perspectives*, desde 1987 a la fecha, se podrá constatar que no ha habido mucho progreso metodológico en la última década.

Bauman y cols. (1994) formularon una crítica del área e identificaron 16 razones, o seudorrazones, del escaso uso de los experimentos verdaderos:

■ *Contaminación.* Este concepto ya fue examinado. La interacción entre miembros de los grupos experimental y de control puede esparcir la intervención. Pero este riesgo puede minimizarse separando los casos físicamente.

■ *Cambios inesperados del programa.* Ningún programa está libre de cambios imprevistos. Lo que deben hacer los investigadores es documentar los cambios, de manera que puedan replicarse a gran escala si se les encuentra efectivos.

■ *Influencia de rasgos personales.* Los programas o intervenciones son implementados por personas, cuyas características individuales pueden influir los resultados favorable o desfavorablemente. Si se demuestra que hay una influencia personal, se puede concluir que la intervención tendrá éxito con la participación de personas similares.

■ *Espera para la aparición de cambios.* Algunos efectos, como los del mejoramiento del acceso a los servicios sobre la tasa de fecundidad, ciertamente toman años, pero otros, como los de la capacitación sobre la calidad de atención, se detectan en semanas. En cualquier caso, diseños menos eficientes que los experimentos verdaderos tampoco ahorrarán tiempo al evaluar el impacto de los mismos factores.

■ *Sesgo de medición por el personal del programa.* Este no es un problema serio si el personal del programa comete errores que no son graves y se distribuyen igualmente en los grupos experimental y de control. De lo contrario, se busca personal ajeno al programa.

■ *Las mediciones pueden ser reactivas.* Si el personal se siente observado, su comportamiento no será el habitual. Pero se pueden utilizar técnicas no reactivas o minimizar el hecho de que se esté conduciendo un estudio.

■ *El programa no es generalizable a gran escala.* Para que lo sea, será conveniente definir sus objetivos con la participación de autoridades que tienen una idea de lo que es viable en el ámbito nacional. En cualquier caso, para saber si es generalizable a gran escala habrá que evaluarlo también a gran escala.

■ *Pocos casos están disponibles.* Con imaginación se puede encontrar más casos. Además, hay que tener en cuenta que, cuando se trabaja con agregados, como clínicas, distritos, etc., bastan unos 20 o 30 para asignar aleatoriamente a las condiciones del estudio. Si hay pocas regiones, se trabaja con un número mayor de provincias; si hay pocas provincias, se trabaja con un número mayor de distritos o comunidades.

■ *Los casos no representan a la población.* Esto no será así si se emplean las técnicas apropiadas de muestreo.

■ *Las replicaciones son necesarias.* Esto es verdad para cualquier diseño, verdaderamente experimental, cuasi experimental o pre-experimental.

■ *Es difícil mantener la equivalencia de los grupos.* No siempre es tan difícil. En cualquier caso, se puede verificar si los grupos se están diferenciando en aspectos distintos de los esperados.

■ *Son difíciles de hallar áreas comparables.* Si no hay regiones comparables, se trabaja con provincias comparables; si no las hay, se desciende a los distritos, y así sucesivamente.

■ *El costo del experimento verdadero es muy alto.* No hay evidencias de que el costo sea mayor que el de los cuasi o pre experimentos, excepto un estudio que lo encontró más costoso en 1%. Por otro lado, los diseños menos eficientes requieren con frecuencia acciones correctivas adicionales, no presupuestadas. En cualquier caso, si el experimento verdadero cuesta más es porque sus resultados valen más.

■ *Los hallazgos llegan demasiado tarde.* En efecto, pueden llegar cuando ya no son aplicables. Si esto es previsible, mejor será no conducir el estudio. Pero los cuasi experimentos están en las mismas condiciones.

■ *Obligan a excluir de beneficios a la gente.* Por ejemplo, se dice que los casos asignados al grupo de control no recibirán los beneficios del grupo experimental. Esto presenta un problema ético si el programa o la intervención retiran un beneficio real. Pero frecuentemente lo que se estudia es una condición nueva, todavía no disponible para el público, y sólo probándola se sabrá si vale. Además, se puede dar el tratamiento al grupo de control apenas terminado el estudio. También se aduce que una intervención puede tener efectos negativos. Sin embargo, los efectos que se buscan son positivos para las personas, y sólo poniéndolos a prueba se sabrá a ciencia cierta si lo son o si causan resultados adversos.

■ *El investigador suele ser vencido por consideraciones políticas.* Para implementar un programa nuevo hay que vencer resistencias como las de la tradición o los intereses creados y, generalmente, los investigadores se estrellan contra estas barreras. Por lo regular triunfan las consideraciones políticas sobre las metodológicas. Por ejemplo, las autoridades pueden no aceptar la asignación aleatoria porque prefieren que una intervención se realice en un lugar específico. Pero hay también muchos casos en los que un rechazo inicial ha sido superado a fuerza de constancia.

En resumen, no hay argumentos a priori para descartar los experimentos verdaderos, y la recomendación es usarlos tanto como sea posible, pues representan la mejor alternativa metodológica para evaluar los efectos de programas o intervenciones. Sólo si la situación del programa hace realmente imposible la asignación aleatoria de casos a tratamientos es cuando el investigador debe considerar la posibilidad de un cuasi experimento. Y al pre-experimento debe recurrirse sólo cuando ya no queda mejor alternativa.

Notas

¹ El personal del programa es protagonista central de las operaciones.

² Los presupuestos principales del método experimental son los de distribución de la variable de interés entre los sujetos de la investigación, de la suma cero de los errores aleatorios y de la partición de la varianza en efectos experimentales y otros.

³ Hay, sin embargo, amenazas a la validez interna del experimento verdadero que no son controladas cabalmente por la asignación aleatoria de sujetos a tratamientos. Ellas serán descritas más adelante.

⁴ Está muy extendido el concepto erróneo de que la asignación aleatoria de *grupos* a tratamientos define el experimento verdadero. De allí la necesidad de aclarar este asunto aquí.

⁵ Pese a no haber asignado aleatoriamente las clínicas a los *grupos*, León y Cuesta (1998) emplearon pruebas estadísticas al comparar sus resultados. Violaciones como esta del modelo estadístico son la regla más que la excepción en la literatura psicológica y de las ciencias sociales.

Referencias

Amin, S. (1997). *Measuring the impact of micro-credit programs on fertility*. Seminario Organizado por el Population Council. New York: Population Council, 15-16 de diciembre.

Anderson, J. (1998). "Modelos conceptuales". Exposición en el *Taller de Investigación Operativa del Proyecto 2000*. Ocucaje: Ministerio de Salud y USAID/Perú, 12 de enero.

Aravindan, G. R., C. Pineau, C. W. Bardin, y C. Y. Cheng (1996). "Ability of trypsin in mimicking germ cell factors that affect Sertoli cell secretory function." *Journal of Cellular Physiology*, 168: 123-133, 1996.

Barcikowski, R. S. (1981). "Statistical power with group mean as the unit of analysis." *Journal of Educational Statistics*, 6: 267-285.

Bauman, K. F., C. I. Viadro y A. O. Tsui (1994). "Use of true experimental designs for family planning program evaluations: Merits, problems, and solutions." *International Family Planning Perspectives*, 20, 3: 108-113.

Benson, J., V. Huapaya, M. Abernathy y T. King (1998). *Mejoramiento de calidad y reducción de costos en un modelo integrado de atención post-aborto: resultados preliminares*. Documentos de Trabajo, núm 18. Serie INOPAL III, Population Council. México, D.F., en prensa.

Bertrand, J. T., R. J. Magnani y J. C. Knowles (s/f). *Handbook of Indicators for Family Planning Program Evaluation*. Chapel Hill, NC.: The Evaluation Project.

Bongaarts, J. y J. Bruce (1995). "The question of access: A response." *Studies in Family Planning*, 26, 4, 243-244.

Bruce, J. (1990). "Fundamental elements of the quality of care: A simple framework." *Studies in Family Planning*, 1990, 21, 2, 61-91.

Campbell, D. C. y J. C. Stanley (1966). *Experimental and Quasi-experimental Designs for Research*. Chicago: Rand-McNally.

Carrón, J. M., M. M. Melián y F. R. León (1994). "Developing tools of low-cost use to improve the quality of care of rural CBD in Paraguay." *Informe Final del Contrato No. CI91.96A del Proyecto INOPAL II*. Lima: Population Council.

Cobián, E. y S. Reyes (1998). *Percepciones masculinas de las necesidades y servicios de planificación familiar y salud reproductiva: Un estudio cualitativo*. Documentos de Trabajo, núm 17. Serie INOPAL III, Population Council. México, D.F., en prensa.

Collaborative Group on Hormonal Factors in Breast Cancer (1996). "Breast cancer and hormonal contraceptives: Collaborative reanalysis of individual data on 53,297 women with breast cancer and 100,239 women without breast cancer from 54 epidemiological studies." *Lancet* 347, 9017: 1713-1727.

Conway, M., M. T. Pizzamiglio y L. Mount (1996). "Status, communality, and agency: Implications for stereotypes of gender and other groups." *Journal of Personality and Social Psychology*, 71, 1: 25-38.

Cook, T. D. y D. T. Campbell (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally.

Del Valle, D., M. R. Gárate, J. R. Foreit, D. Li y F. Suárez (1993). *Encuesta de Disponibilidad de Servicios de Planificación Familiar*. Lima: Ministerio de Salud, Instituto Nacional de Estadística e Informática, Asociación Benéfica PRISMA, y Population Council.

Donabedian, A. (1988). "The quality of care: How can it be assessed?" *Journal of the American Medical Association*, 260, 12: 1743-1748.

Edington, E. S. (1966). "Statistical inference and nonrandom samples." *Psychological Bulletin*, 66: 485-487.

Fisher, A. A., J. E. Laing, J. E. Stoeckel y J. W. Townsend (1991). *Handbook for family planning operations research design*, 2a. ed. New York: Population Council. (Primera edición: 1983.)

Foreit, J. R., J. Rosen, M. Ramos, E. Mostajo y R. Monge (1990). "The impact of service delivery frequency on family planning program output and efficiency." *Studies in Family Planning*, 21, 4: 209-215.

Foreit, J. R., M. R. Gárate, M. C. Herrera, A. Brazzoduro y F. Guillén (1992). "Comparison of male and female CBD distributors in Peru." *Studies in Family Planning*, 22, 1: 58-62.

Foreit, K. G., J. R. Foreit, G. Lagos y A. Guzmán (1993). "Effectiveness and cost-effectiveness of postpartum IUD insertion in Lima, Peru." *International Family Planning Perspectives*, 19, 1: 19-24.

Fortney, J. A. (1996). "Reproductive morbidity: A conceptual framework." Manuscrito presentado al *Seminar on Innovative Approaches to the Assessment of Reproductive Health*. Manila, Philippines: IUSSP Committee on Reproductive Health and Population Institute of the University of Philippines, 1996.

Guba, E. G. y Y. S. Lincoln (1994). "Competing paradigms in qualitative research". En N. K. Denzin y Y. S. Lincoln (Eds.), *Handbook of qualitative research*. Thousand Oaks, Cal.: SAGE Publications.

Heinsman, D. T. y W. R. Shadish (1996). "Assignment methods in experi-

mentation: When do nonrandomized experiments approximate answers from randomized experiments?" *Psychological Methods*, 1, 2: 154-169.

Hsieh, F. Y. (1988). "Sample size formulae for intervention studies with the cluster as unit of randomization." *Statistics in Medicine*, 8: 1195-1201.

ICPD (International Conference on Population and Development). *Program of Action of the Conference*. El Cairo: Naciones Unidas, 1994.

INOPAL I (1990). *Final Technical Report: Operations Research to Improve Family Planning and Maternal-Child Health Service Delivery Systems in Latin America and the Caribbean*. México: Population Council.

INOPAL II (1995). *Informe Final: Investigación Operativa y Cooperación Técnica en Planificación Familiar y Salud Materno Infantil en América Latina y el Caribe*. Lima: Population Council.

Koblinsky, M. A. (1995). "Beyond maternal mortality: Magnitude, interrelationships, and consequences of women's health, pregnancy-related complications and nutritional status on pregnancy outcomes." *International Journal of Gynecology & Obstetrics*, 48, June Supplement: S21-S32.

León, F. R. (1995). *Test de Conocimientos de Planificación Familiar para Proveedores No Profesionales*. Lima: Population Council.

León, F. R. y A. Cuesta (1993). "The need for quasi-experimental methodology to evaluate pricing effects." *Studies in Family Planning*, 24, 6: 375-381.

León, F. R., R. Hurtado, R. Monge, A. Mongrut, T. Watanabe, M. E. León, J. Parra e I. Ramos (1994). *Mejorando la Calidad y Acceso a Servicios de Planificación Familiar en Ambitos Rurales y Semi-Rurales del Perú*. Lima: INANDEP (Instituto Andino de Estudios en Población y Desarrollo).

Mensch, B. S., R. A. Miller, A. A. Fisher, J. Mwita, N. Keyonzo, F. Y. Mohamed Ali y C. Ndeti. (1994). "Family planning in Nairobi: A situation analysis of the City Commission clinics." *International Family Planning Perspectives*, 20, 2: 48-54.

Miller, R., A. Fisher, K. Miller, L. Ndhlovu, B. Ndugga Maggwa, I. Askew, D. Sanogo y P. Tapsoba (1997). *The Situation Analysis approach to assessing family planning and reproductive health services*. New York: Population Council.

Mostajo, P. y F. R. León (1992). "A diagnosis of post model and territorial distribution of family planning service delivery points in Metropolitan Lima." *Informe Final del Sub-Contrato C119-16A del Proyecto INOPAL II*. Lima: Population Council.

Murray, D. M. y P. J. Hannan (1990). "Planning for the appropriate analysis in school-based drug-use prevention studies." *Journal of Consulting and Clinical Psychology*, 58: 458-468.

Ramos, I, J. A. Noble, L. Carbajal, F. Iturrizaga, E. Caballero, A. Padilla y M. R. Gárate (1998). "Continuidad en la planificación familiar: Un seguimiento de usuarias de establecimientos públicos." Lima: Population Council.

Reyes, J., L. H. Ochoa, V. Sandoval, H. Raggars y S. Rutstein (1997). *Encuesta demográfica y de salud familiar 1996*. Lima: Instituto Nacional de Estadística e Informática y Macro International Inc.

Ross, J. A. (1995) "The question of access." *Studies in Family Planning*, 26, 241-242.

Ross, J. A., J. Donayre y R. McNamara (1987). "Perspectives on operations research." *International Family Planning Perspectives*, 13, 4: 128-136.

Rosenberg, M. J., R. Foldes, D. R. Mishell Jr., L. Speroff, M. S. Waugh y R. Burkman (1996). "Performance of the TCu380A and Cu-Fix IUDs in an international randomized trial." *Contraception*, 53, 197-203.

Sánchez, J., E. Gotuzzo, J. Escamilla, C. Carrillo, I. A. Phillips, C. Barrios, W. E. Stamm, R. L. Ashley, J. K. Kreiss y K. K. Holmes (1996). "Gender differences in sexual practices and sexually transmitted infections among adults in Lima, Peru." *American Journal of Public Health*, 86, 8:1098-1107.

Snowden, R. K. I. Kennedy, F. R. León, V. C. Orense, H. W. Perera, R. Phillips, I. Askew, A. Flynn y L. J. Severy (1988). "Physician's views of periodic abstinence methods: A study in four countries." *Studies in Family Planning*, 19, 4: 215-226.

Velásquez, A., L. Jiménez, A. Reckemmer, M. E. Planas, F. R. León, R. Durand y A. Calderón (1998). *Incrementando la utilización de servicios de salud reproductiva en una clínica de Lima*. Documentos de Trabajo, núm 4. Serie INOPAL III, Population Council. México,D.F.

Reconocimientos

La segunda parte de este artículo (Principales diseños de investigación operativa) proviene de una conferencia del autor preparada para el Taller de Investigación Operativa que condujo el Instituto de Estudios de Población de la Universidad Cayetano Heredia (diciembre de 1996), como parte de su proyecto de institucionalización de la investigación operativa en universidades y programas con el auspicio del Proyecto INOPAL III del Population Council, a su vez bajo el contrato Núm. CCP-95-00-00007-00 con la Oficina de Población de la Agencia de los Estados Unidos para el Desarrollo Internacional. La primera parte proviene de una conferencia preparada por el autor para una reunión con profesores y alumnos de Ciencias Sociales de la Universidad Católica (junio de 1997), que también fue parte del mismo proyecto de IEPO. El autor agradece los comentarios de Jane Bertrand a una versión previa de la primera parte.

Otros títulos de la serie

Documento de trabajo núm. 1. *La anticoncepción de emergencia como elemento de la atención a víctimas de violación*

Ricardo Vernon, Raffaella Schiavon, Silvia E. Llaguno

Documento de trabajo núm. 2. *Educación sobre salud sexual y reproductiva en áreas indígenas de Guatemala a través de maestros bilingües*

Gloria Cospín, Ricardo Vernon

Documento de trabajo núm. 3. *Oferta sistemática de servicios de planificación familiar y salud reproductiva en Guatemala*

Ricardo Vernon, Emma Ottolenghi, Federico León, Jorge Solórzano

Documento de trabajo núm. 4. *Incrementando la utilización de servicios de salud reproductiva en una clínica de Lima*

Aníbal Velásquez, Lissette Jiménez, Adolfo Rechkemer, María E. Planas, Federico R. León, Rubén Durand, Alicia Calderón

Documento de trabajo núm. 5. *Estrategias de información sobre anticoncepción de emergencia*

Martha Givaudan, Ricardo Vernon, Carmen Fuertes, Susan Pick

Documento de trabajo núm. 6. *Encuesta de necesidades de información, educación y comunicación (IEC) sobre salud reproductiva en seis etnias en México*

Javier Cabral, Ángel Flores, Francisco Huerta, Carmen Baltazar, Fabiola García, Concepción Orozco, Carlos Brambila

Documento de trabajo núm. 7. *Costos de las consultas de atención integral en salud reproductiva en Guatemala*

Carlos Brambila, Jorge Solórzano

Documento de trabajo núm. 8. *Uso de listados para detectar y referir mujeres a los servicios de salud reproductiva*

Luis Améndola, Delmy Euceda, Rebecka Lundgren, Irma Mendoza, Norma Ali

Documento de trabajo núm. 9. *La respuesta del consumidor a los cambios de precios de los servicios de planificación familiar y salud reproductiva*

Teresa de Vargas, John H. Bratt, Varuni Dayaratna, James R. Foreit, Daniel H. Kress

Documento de trabajo núm. 10. *Estrategias para involucrar a los hombres en el cuidado de la salud reproductiva: de la administración de la granja a la administración de la familia*

James Foreit, Rebecka Lundgren, Irma Mendoza, Dolores María Valmaña, Judy Canahuati

Documento de trabajo núm. 11. *Sí, se puede: cómo mejorar la calidad de la atención post-aborto en un hospital público. El caso de Oaxaca, México*

Ana Langer, Cecilia García-Barrios, Angela Heimburger, Lourdes Campero, Olivia Ortiz, Carmen Díaz, Vilma Barahona, Francisca Ramírez, Beatriz Casas, Beverly Winikoff, Karen Stein

Documento de trabajo núm. 12. *Investigación operativa en planificación familiar y salud reproductiva*

Federico R. León

Documento de trabajo núm. 13. *Tres estrategias para promover el sostenimiento de las clínicas de CEMOPLAF en Ecuador*

John H. Bratt, James Foreit, Teresa de Vargas

Se terminó de imprimir en
Solar, Servicios Editoriales, S.A. de C.V.
Tel.: 515 1657.
El tiraje consta de 2,200 ejemplares.