

Indicators



**EVALUACIÓN
DE PROGRAMAS
DE PLANIFICACIÓN
FAMILIAR**



**CON
ADAPTACIONES
PARA SALUD
REPRODUCTIVA**

**Bertrand
Magnani
Rutenberg**

PROCESO

PRODUCTO

RESULTADO:

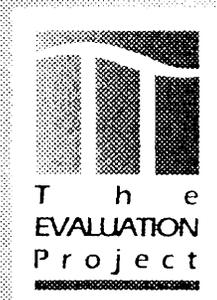
EFFECTO

IMPACTO

**EVALUACIÓN
DE PROGRAMAS
DE PLANIFICACIÓN FAMILIAR**

**CON
ADAPTACIONES
PARA SALUD
REPRODUCTIVA**

Jane T. Bertrand
Robert J. Magnani
Naomi Rutenberg



Setiembre 1996

USAID Contrato Número: DPE-3060-C-00-1054-00

Reconocimientos

Reconocemos el apoyo de la Oficina de Población de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID Contrato Número: DPE-3060-C-00-1054-00) y agradecemos haber hecho posible esta publicación. Las opiniones expresadas aquí corresponden a los autores y no a la agencia patrocinadora.

Los autores desean agradecer a varios miembros del Proyecto EVALUATION por sus críticas constructivas a diversas versiones de este manual: Amy Tsui, David Guilkey, Albert Hermalin, John Ross, John Stover, David Hotchkiss, Jennifer Strickler, y Lisanne Brown. Krista Stewart, Asesora Técnica del Proyecto de la USAID, también hizo sugerencias muy útiles a versiones anteriores. También queremos extender nuestro agradecimiento a varios revisores externos: Jerry Bowers, John Casterline, Allan Keller, Barbara Janowitz, F.M. Mburu, Virginia Paine, y Jack Reynolds. Algunos colegas hicieron contribuciones sobre aspectos específicos del tema de la salud reproductiva (Capítulo VII): Gina Dallabetta, Erin Eckert, Susan Hassig, Cate Johnson, Marge Koblinsky, Miriam Labbok, Therese McGinn, Barbara Mensch, Nancy Mock, Leo Morris y Chloe O'Gara.

Varias misiones de la USAID y programas nacionales han ofrecido al personal del Proyecto EVALUATION la oportunidad de probar distintos enfoques de la evaluación de la planificación familiar en el campo. Estamos particularmente agradecidos con nuestros colegas de la USAID y la familia de programas de planificación familiar en Brasil, Ecuador, Honduras, Gana, India, Kenia, Marruecos, Nigeria, Perú, Filipinas y Tanzania por su apoyo y participación en varias actividades de evaluación desde 1992.

Apreciamos el apoyo de Allison Cohn, Tracy Saunders y Martha Scherzer en la edición de la versión final de este manuscrito. Por el diseño y la diagramación del documento, damos las gracias a Sue Koenigshofer y Erin Hancock de SCK Design.

La traducción al español del manual y su diagramación estuvo a cargo de Luis Rosero-Bixby y Helena Ramírez.

LISTA DE SIGLAS

AOE	Atención Obstétrica Esencial
APP	Años de Protección de la Pareja
APPE	Años de Protección Pareja Estandarizados
ARI	Ayuda a las Personas a Lograr sus Intenciones Reproductivas
CAP	Conocimiento, Actitud y Práctica
CBD	Distribución Basada en la Comunidad
CDC	Centros de Control y Prevención de Enfermedades de Atlanta, EEUU
CDIE	Centro para el Desarrollo y la Evaluación (de USAID)
COPE	Técnica Eficiente para el Proveedor Orientada hacia el Cliente
CPS	Encuesta de Prevalencia Anticonceptiva
DHS	Encuestas de Demografía y Salud
ETS	Enfermedad(es) de Transmisión Sexual
GIS	Sistemas de Información Geográfica
IEC	Información-Educación-Comunicación
IPPF	Federación Internacional de Paternidad Responsable
MPE	Malnutrición Proteico-Energética
MS	Ministerio de Salud
MVIC	Máximo-verosimilitud con Información Completa
ONG	Organización no Gubernamental
PF	Planificación Familiar
PF-PN	Programa Nacional de Planificación Familiar
PF-SA	Servicios de Apoyo a la Planificación Familiar
PPS	Puntos de Provisión de Servicios
SAM	Módulo de Disponibilidad del Servicio (de la DHS)
SIDA	Síndrome de Inmunodeficiencia Adquirido (AIDS en inglés)
SIG	Sistemas de Información de Gestión
SMI	Salud Materno-infantil
SR	Salud Reproductiva
TBN	Tasa Bruta de Natalidad (CBR en inglés)
TGF	Tasa Global de Fecundidad (TFR en inglés)
USAID	Agencia para el Desarrollo Internacional de los EEUU
VIH	Virus de Inmunodeficiencia Humano
WFS	Encuesta Mundial de Fecundidad

TABLA DE CONTENIDOS

	Reconocimientos	2
	Lista de siglas	3
	Capítulo	
I	Visión de conjunto de la evaluación	6
	¿Por qué evaluar?	
	Objetivos del manual	
	Destinatarios de este manual	
	Planificación familiar o salud reproductiva	
	Alcance y punto focal de las evaluaciones	
	¿Por qué otro manual de evaluación?	
	Organización de este manual	
II	Definición del alcance de la evaluación	12
	Determinación de metas y objetivos de programas	
	Descripción de cómo “debe” funcionar el programa	
	Establecimiento de los objetivos de la evaluación	
	Esquema del alcance de la evaluación en el plan de evaluación	
III	Enfoque metodológico: Seguimiento (monitoreo) de programas	26
	Aclaración del propósito principal del seguimiento	
	Identificación de los componentes que deben recibir seguimiento	
	Definición de indicadores relevantes	
	Identificación de fuentes de datos	
	Diseño de un formato para la presentación de los resultados	
	Resumen del enfoque metodológico	
IV	Enfoque metodológico: Valoración del impacto	44
	Una advertencia amigable	
	Visión de conjunto de los temas	
	Criterios para la selección del enfoque metodológico	

	Enfoques preferidos	
	Enfoques alternativos	
	Resumen	
V	Desarrollo de un plan de implementación	74
	Definición de las instituciones y personas responsables de la evaluación	
	Establecimiento de un cronograma de actividades específicas	
	Presupuesto de la evaluación	
VI	Diseminación y utilización de los resultados	80
	Preparación de la presentación de resultados para la audiencia correspondiente	
	Identificación de resultados con implicaciones programáticas importantes	
	Establecimiento de un foro para presentar los resultados que implican una acción	
	Presencia permanente en el proceso	
VII	Adaptaciones a otras intervenciones de salud reproductiva	84
	Prevención de ETS/VIH	
	Embarazo seguro	
	Lactancia	
	Nutrición de las mujeres	
	Servicios de salud reproductiva para adolescentes	
VIII	Resumen: Lista de pasos del diseño de un plan de evaluación	98
	Referencias	101
<u>Apéndice</u>		
	El modelo multinivel en formato de regresión	108

Capítulo I

Visión de conjunto de la evaluación

- ¿Por qué evaluar?
- Objetivos de este manual
- Destinatarios de este manual
- Planificación familiar o salud reproductiva
- Alcance y punto focal de las evaluaciones
- ¿Por qué otro manual de evaluación?
- Organización de este manual

VISIÓN DE CONJUNTO DE LA EVALUACIÓN

¿POR QUÉ EVALUAR?

La evaluación es la aplicación de procedimientos de investigación de las ciencias sociales para valorar y mejorar las políticas y programas sociales, desde sus primeras etapas de definición y diseño hasta su desarrollo e implementación (Rossi & Freeman, 1993). Los resultados de la evaluación deben proveer información para la administración de programas, el planeamiento estratégico, el diseño de nuevos proyectos o iniciativas y la fijación de recursos.

La evaluación de los programas de planificación familiar incluye tanto el seguimiento del programa como la valoración del impacto. El seguimiento se usa para determinar si el programa es llevado a cabo apropiadamente en sus distintos niveles y a qué costo. También marca el cambio ocurrido a través del tiempo en la entrada de recursos, la producción y el uso de los servicios. La valoración del impacto determina la medida en que este cambio puede ser atribuido a la intervención del programa (causa y efecto).

Los resultados del seguimiento del programa son indispensables para la gestión del programa porque informan al administrador si el programa está bien orientado, donde están los problemas y cuáles resultados inesperados han ocurrido. La evaluación de los procesos usados para implementar el programa permite hacer correcciones a mitad del camino. De hecho, el tipo de seguimiento de programas descrito en el Capítulo III de este manual está integralmente vinculado a los sistemas de información de gestión (SIG), enfatizando la fuerte relación entre el seguimiento y la administración.

Los resultados de la evaluación también son insumos importantes para el planeamiento estratégico y el diseño de programas. Los indicadores del desempeño del programa, de los productos y de los resultados en la población, describen la situación vigente de la demanda de servicios y el ambiente del programa. Los resultados que vin-

culan los insumos y las actividades a los productos del programa y a los cambios en la población sirven para demostrar qué ha funcionado en el pasado y sugerir posibles direcciones para el futuro. Las intervenciones exitosas se pueden incrementar o replicar en nuevos proyectos o etapas de programas, mientras las actividades que no producen resultados pueden ser descartadas. La evaluación también puede servir para explorar por qué algunas intervenciones no funcionaron.

Resumiendo, las personas responsables de implementar programas y las que financian programas deberían exigir que la evaluación sea parte integral de cada intervención. Para lograr un mayor beneficio, la evaluación debe formar parte del diseño del programa desde el principio y proveer información a los administradores durante el transcurso de la actividad. Los resultados de la evaluación pueden ayudar a los administradores a saber qué están haciendo bien, a identificar deficiencias que pueden ser corregidas y a tomar decisiones informadas sobre la dirección futura de sus programas. En la situación actual de restricciones presupuestarias, la evaluación apunta hacia un uso racional de los limitados recursos —humanos y materiales— para lograr resultados.

OBJETIVOS DEL MANUAL

Este manual prepara al usuario para:

- Distinguir entre los principales tipos de evaluación de programas, seguimiento de programas y valoración de impacto;
- Evaluar críticamente las ventajas y desventajas de los métodos alternativos de valoración de impacto;
- Valorar y seleccionar los tipos de evaluación más adecuada para un escenario determinado (i.e., aquella que responde las preguntas más importantes, y que es factible y práctica de ejecutar);

- Identificar los indicadores y fuentes de datos apropiados para la evaluación; y
- Preparar un plan de evaluación definiendo el diseño del estudio, los indicadores y las fuentes de datos que guiará la implementación correspondiente.

DESTINATARIOS DE ESTE MANUAL

Este manual está dirigido a profesionales de la salud con distintos niveles de capacitación y experiencia en la evaluación de programas, incluyendo:

- los administradores de programas,
- los especialistas en evaluación
- el personal de las agencias donantes.

Este manual supone que el usuario cuenta con capacitación o experiencia en técnicas de investigación social; no revisa procedimientos básicos de recolección y análisis de datos.¹

¿PLANIFICACIÓN FAMILIAR O SALUD REPRODUCTIVA?

Hasta hace poco, el término "planificación familiar" (PF) se interpretaba principalmente como "servicios anticonceptivos." Aunque algunos programas también ofrecían servicios de infertilidad para parejas que no podían lograr el embarazo, la mayor parte de los servicios estaba relacionada con la anticoncepción.

Sin embargo, alrededor de la Conferencia Internacional de Población y Desarrollo en el Cairo en 1994, hubo bastante presión dentro de la comunidad relacionada con el tema de la población (especialmente de grupos de mujeres) para ampliar la constelación de servicios a la atención de otras necesidades de salud reproductiva (SR) de las mujeres. Estas incluyen el embarazo seguro, las enfermedades de transmisión sexual y el síndrome de inmunodeficiencia adquirido (SIDA), la lactancia, la nutrición de las mujeres y otras. Además, las adolescentes y los hombres han sido identificados como audiencias objetivo importantes para las intervenciones de los programas. La importancia relativa de los diferentes servicios o poblaciones objetivo varía por países.

Este manual se basa en la vasta experiencia de la comunidad relacionada con el tema de la población en la evaluación de programas de planificación familiar. Específicamente, ofrece

lineamientos relevantes para desarrollar un plan de evaluación para un programa nacional de planificación familiar, con énfasis en los servicios anticonceptivos.

¿Se aplican por igual estos principios a los programas de otras áreas de la salud reproductiva? En lo referente al seguimiento de programas, la lógica y las técnicas metodológicas son, en buena medida, similares entre los distintos tipos de intervenciones de PF y SR. Las diferencias están en la selección de los indicadores, su medición y la definición de la población objetivo, no en los diseños de estudio o las técnicas de análisis utilizados.

Por lo tanto, el enfoque para la evaluación descrito en este manual generalmente es aplicable a otras áreas de la salud reproductiva, ya sean programas verticales (por ejemplo, un programa nacional para promover la lactancia), o programas integrados con componentes múltiples (por ejemplo, lactancia, embarazo seguro, planificación familiar y prevención de ETS/VIH). Sin embargo, hay algunas diferencias notables que aparecen resumidas en el Capítulo VII.

En lo referente a la valoración del impacto, la lógica y las técnicas metodológicas también son semejantes. En principio, los tres "métodos preferidos" presentados en el Capítulo IV son tan aplicables a otras intervenciones de salud reproductiva como a la planificación familiar. Sin embargo, el Capítulo IV se centra casi exclusivamente en la planificación familiar por varias razones. Primero, porque la cuestión del impacto en los programas de planificación familiar ha recibido más atención, y durante más años, que otras intervenciones de SR. Consecuentemente, se ha puesto más atención al desarrollo y refinamiento de los métodos de evaluación de la planificación familiar que a otros programas de SR. Los avances metodológicos descritos en el Capítulo IV han sido desarrollados en relación a los esfuerzos para valorar el impacto de la planificación familiar sobre la fecundidad. No existe un trabajo semejante hasta ahora relacionado con otros resultados de SR. Además, uno de los resultados asociados con la planificación familiar - disminución de la fecundidad - puede ser medido con validez a través de la información reportada por los entrevistados en encuestas grandes. En cambio, existen dificultades metodológicas importantes

¹ Sin embargo, aquellos interesados en un texto útil sobre los antecedentes pueden consultar García-Nuñez (1992).

para medir otros resultados de SR, como la mortalidad materna, la prevalencia de infecciones VIH y las tasas de aborto. El énfasis en la fecundidad en el Capítulo IV no tiene como propósito disminuir la importancia de otras intervenciones de salud reproductiva. Más bien, refleja el hecho de que la fecundidad se presta más fácilmente que la mayoría de los otros resultados de salud reproductiva para mediciones precisas, basadas en informes de los entrevistados de las encuestas por muestreo.

En suma, este manual aprovecha la amplia experiencia en evaluación de los programas de planificación familiar, mucha de la cual es aplicable al espectro más amplio de intervenciones de salud reproductiva. El manual está concebido como complemento a otras dos publicaciones del Proyecto EVALUATION, el Manual de Indicadores para la Evaluación de Programas de Planificación Familiar (Bertrand et al., 1994) y el de Indicadores de Evaluación de Programas de Salud Reproductiva (Bertrand & Tsui, 1995). Estos dos documentos ofrecen un menú de indicadores para las intervenciones de evaluación en las áreas de la planificación familiar, el embarazo seguro, la lactancia, la prevención de ETS/VIH, la nutrición de las mujeres y los servicios de salud reproductiva para adolescentes. A diferencia de los anteriores, este manual ofrece lineamientos para el diseño de un plan de evaluación (que incorporará esos indicadores) para el seguimiento y evaluación de las intervenciones.

ALCANCE Y PUNTO FOCAL DE LAS EVALUACIONES

Las evaluaciones varían mucho en cuanto a su alcance y campo de acción. Por ejemplo, el área objeto de la evaluación podría definirse como:

- todo el país;
- toda una región o estado; o
- una ciudad o lugar específico.

La evaluación puede centrarse en diferentes componentes del programa:

- insumos,
- procesos,
- productos, y
- resultados.

Las mediciones pueden ser:

- poblacionales (por ejemplo, en una muestra aleatoria de la población general), o
- del programa (por ejemplo, de los clientes o participantes en un determinado programa).

Pueden usarse distintas técnicas para recolectar y analizar los datos:

- cuantitativa, o
- cualitativa.

La población objetivo variará según las circunstancias y para los distintos tipos de intervenciones:

- todas las mujeres en edad reproductiva (por ejemplo, planificación familiar);
- todos los adultos sexualmente activos (por ejemplo, planificación familiar integrada - prevención de ETS/VIH); o
- jóvenes de 10 –19 años (i.e., programas para adolescentes).

Este manual se centra en la evaluación de programas con alcance nacional, aunque muchas de las técnicas también pueden usarse para evaluar programas o proyectos más pequeños. Considera la pregunta clave de la mayoría de los administradores de programas y del personal de agencias donantes: ¿ha logrado el programa sus objetivos en lo referente a producir cambios en la población? Va más allá, preguntando en qué medida se pueden atribuir los cambios observados al programa.

Aunque obtener resultados es lo más importante, una evaluación exhaustiva también debe examinar los procesos involucrados en la realización del programa. Históricamente, la evaluación de programas de planificación familiar ha tendido a centrarse mucho en los productos inmediatos del programa de tipo cuantitativo (por ejemplo, el número de nuevos aceptantes, los años de protección pareja [APP] o los resultados finales en la población (por ejemplo, el nivel de prevalencia anticonceptiva, la tasa global de fecundidad [TGF]). Sin embargo, este enfoque de evaluación trata al programa como una “caja negra”. Si los resultados esperados no son obtenidos, ofrece poca explicación sobre las razones. No se sabe, por ejemplo, cuáles factores contribuyeron a obtener resultados tan pobres: ¿un acceso inadecuado al servicio, una atención de mala calidad en los puntos de provisión de servicios (PPS) del sistema, falta de información entre la población objetivo, falta de suministros en el

exitoso, se tiene poco conocimiento sobre qué contribuyó al éxito.

En suma, una evaluación exhaustiva examina no solo los resultados cuantitativos que señalan el progreso hacia los objetivos del programa; sino que evalúa también el funcionamiento interno del programa en términos de áreas funcionales (administración, capacitación, suministros y aspectos logísticos, información, educación y comunicación [IEC], investigación y evaluación) y suficiencia del programa (acceso a los servicios y calidad de la atención). Este manual se centra tanto en los procesos como en los resultados, con un énfasis particular en los resultados. Estos siguen siendo el interés principal de muchos administradores de programa y agencias donantes, especialmente en un ambiente con recursos decrecientes y responsabilidades crecientes.

¿POR QUÉ OTRO MANUAL DE EVALUACIÓN?

La comunidad relacionada con el tema de la población se ha interesado en la evaluación de los programas de planificación familiar desde sus orígenes en la década de los sesenta. Como consecuencia, varios métodos de evaluación han sido propuestos y refinados (Bogue, 1970; Reynolds, 1972; Sherris et al., 1985; Lloyd y Ross, 1989; Manuales de NNUU, 1979, 1982, 1986; García-Nuñez, 1992; Buckner et al., 1995). ¿Por qué, entonces, necesitamos un nuevo documento sobre cómo evaluar los programas de PF? Este documento difiere de otros textos de evaluación por lo siguiente:

- Hace referencia a una necesidad declarada cada vez más frecuentemente por las agencias donantes y los administradores de programas: ¿cómo se diseña un plan de evaluación para un programa nacional?
- Reconoce que en la comunidad internacional de agencias donantes la acción focalizada de la planificación familiar se ha desplazado de un sentido directamente definido como servicios anticonceptivos a una definición más amplia que comprende otros aspectos de la salud reproductiva.
- Incorpora desarrollos y refinamientos metodológicos producidos recientemente por el Proyecto EVALUATION, financiado por la USAID, en los lineamientos para la aplicación en el campo.

- Actualiza el “estado del arte” en los diseños más apropiados para valorar el impacto de los programas de planificación familiar y señala los asuntos metodológicos en términos que los administradores y el personal de evaluación puedan entender. Estos términos, sin embargo, resultan también satisfactorios para la gran comunidad científica que se preocupa por el rigor metodológico.
- Refleja la experiencia práctica del personal del Proyecto EVALUATION en la preparación de estrategias de evaluación para países específicos; como tal, considera el inevitable conflicto entre el rigor científico y las limitaciones prácticas.

ORGANIZACIÓN DE ESTE MANUAL

Este manual está organizado de acuerdo con los elementos de un plan de evaluación cuyo esquema se presenta en la Gráfica I-1. Un plan de evaluación describe lo que se va a hacer, cómo, por quién, cuándo, con cuales fondos y por qué; sirve como plan de acción para la implementación. Idealmente, será preparado junto con el diseño del programa (o antes de su implementación). En lo posible, el plan debe incluir insumos de los planificadores y los administradores del programa, quienes deben trabajar en contacto con los especialistas que llevarán a cabo la investigación de evaluación.

Cada capítulo trata una sección del esquema modelo para un plan de evaluación, que aparece en la Gráfica I-1. La información tiene el propósito de ofrecer una guía para completar un plan de evaluación para un programa de PF, confeccionado para una situación particular. Los capítulos cubren lo siguiente:

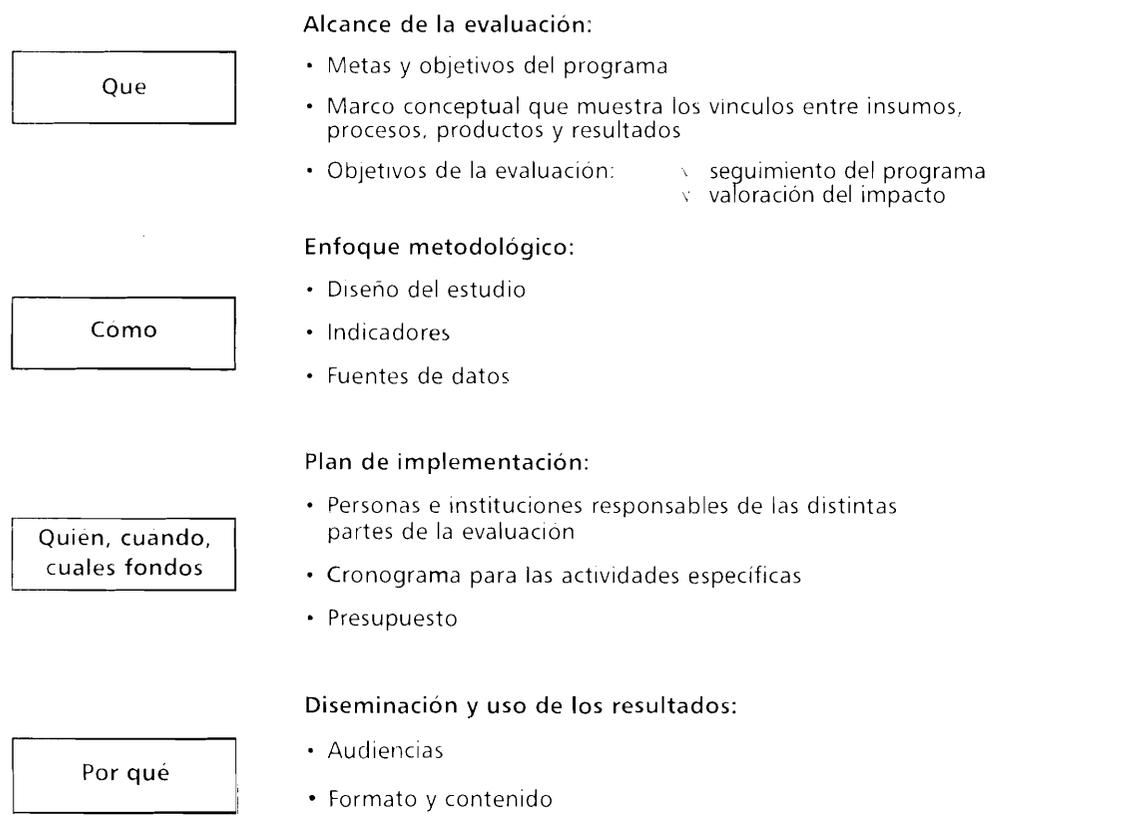
- Capítulo II** Definición del alcance de la evaluación
- Capítulo III** Enfoque metodológico: seguimiento de programas
- Capítulo IV** Enfoque metodológico: valoración del impacto
- Capítulo V** Desarrollo de un plan de implementación
- Capítulo VI** Diseminación y utilización de los resultados
- Capítulo VII** Adaptaciones a otras intervenciones de salud reproductiva
- Capítulo VIII** Resumen: lista de pasos del diseño de un plan de evaluación

Debido a que demostrar impacto es la cuestión clave en una evaluación, dedicamos un capítulo aparte (IV) para este tema. En contraste con textos anteriores que ofrecen un menú de ocho "métodos clásicos" para evaluar el impacto de los programas de planificación familiar sobre la fecundidad, esta sección examina en forma crítica las alternativas y se inclina claramente por el uso de tres métodos en particular. Los lectores notarán que la terminología de esta sección es más cuantitativa que en otros capítulos; sin embargo, hemos tratado de presentar los conceptos en un lenguaje que pueda ser comprendido por personas con un conocimiento básico del diseño de estudios de investigación en ciencias sociales. El Capítulo IV puede leerse en dos planos. Para aquellos con dominio limitado de la estadística, el capítulo identifica los tres métodos preferidos ("preferidos" porque los cambios observados pueden ser atribuidos al programa) y explica

sus ventajas y desventajas. Da también al lector la oportunidad de familiarizarse con los elementos para tomar decisiones informadas respecto al tipo de evaluación a realizar, aunque el lector no sea directamente responsable de los aspectos técnicos del trabajo. Para lectores con más conocimientos de estadística, el Capítulo IV explica las razones para llamar "preferidos" a tres métodos específicos y resume los enfoques y procedimientos de las estimaciones estadísticas requeridas en estos métodos. Al ver el Capítulo IV, los administradores de programas y el personal de las agencias donantes pueden pensar que está dirigido solo a especialistas en evaluación. Pero son precisamente estas dos categorías de profesionales en salud quienes están frecuentemente bajo la presión de "demostrar impacto". Los lectores que se sientan desmotivados por el lenguaje estadístico del Capítulo IV pueden tratar de centrar su atención en los conceptos y no en las estadísticas.

Grafica I-1

Esquema modelo para un plan de evaluación



Capítulo II

Definición del alcance de la evaluación

- Determinación de las metas y los objetivos del programa
- Descripción de cómo "debería" funcionar el programa
- Establecimiento de los objetivos de la evaluación
- Esquema del alcance de la evaluación

DEFINICIÓN DEL ALCANCE DE LA EVALUACIÓN

DETERMINACIÓN DE LA META Y OBJETIVOS DEL PROGRAMA

Definición del programa

Las distintas personas tienen distintas ideas sobre qué es un "programa". En este manual estamos interesados principalmente en la evaluación de los programas nacionales. Sin embargo, muchas de las técnicas descritas aquí son igualmente aplicables a intervenciones más pequeñas. Los recuadros II-1 y II-2 definen los términos "programa" y "proyecto" utilizados en este manual.

En la mayoría de los países en desarrollo existen ahora muchas organizaciones dedicadas a la provisión de planificación familiar y otros servicios de salud reproductiva. Aunque cada organización debe dar seguimiento a su propio desempeño y resultados, los encargados de tomar decisiones de alto nivel y el personal de las agencias donantes están igualmente interesados en evaluar los cambios a nivel nacional (por ejemplo, en prevalencia anticonceptiva) que resultan de los esfuerzos colectivos de todos los programas y fuentes que ofrecen servicios relevantes. El conjunto de servicios relevantes suele ser llamado "el programa nacional", aún sin una entidad coordinadora oficial. Por lo tanto, en un país el "programa nacional" puede incluir los servicios del Ministerio de Salud (MS), los afiliados a la Federación Internacional de Paternidad Responsable (IPPF), otras organizaciones no gubernamentales (ONGs), un programa subsidiado de mercadeo social y el sector comercial con fines de lucro.

Nata del traductor: los términos "metas" y "objetivos" son la traducción del inglés "goals" y "objectives", respectivamente. En ciertos textos de evaluación en español estos dos términos se usan completamente a la inversa: "goals"=objetivos y "objectives"=metas.

Definición de la meta y objetivos del programa

Idealmente, el programa tendrá metas y objetivos bien articulados.

La meta es una declaración, usualmente general y abstracta, de un estado deseado hacia el cual está dirigido un programa (Rossi y Freeman, 1993). Una meta puede establecerse en términos de las actividades de una agencia completa o puede ser específica para un país particular. Por ejemplo, una de las metas de la USAID es:

- estabilizar la población mundial y proteger la salud humana (USAID, 1995).

Algunos ejemplos de metas específicas para países son:

- una mejora sostenible del nivel de vida, y
- el empoderamiento político, social y económico.

Los objetivos, en cambio, son declaraciones específicas y operacionalizadas con el detalle de los logros deseados en un programa (Rossi & Freeman, 1993). Algunos ejemplos de objetivos a nivel de población relevantes para la planificación familiar son:

- reducir la tasa global de fecundidad a 4.0 hijos para el año X;
- aumentar la prevalencia anticonceptiva en un 50% para el año X; y
- lograr un intervalo medio de 36 meses entre los nacimientos al final de un proyecto quinquenal.

Algunos países tienen objetivos cuantificables para sus programas (como en el ejemplo anterior). En otros casos, el objetivo declarado indicará la dirección deseada del cambio sin cuantificar su magnitud (por ejemplo, aumentar la prevalencia anticonceptiva durante la vida del programa). Cualquier caso puede ser "puesto a prueba", aunque los resultados son menos ambiguos cuando el nivel de aumento esperado ha sido cuantificado.

Los programas nacionales generalmente tienen metas que especifican los resultados para la población, tales como los cambios en la prevalencia anticonceptiva, la mortalidad infantil, la fecundidad total y otros. Los programas institucionales (por ejemplo, el programa del afiliado a la IPPF) y los proyectos de

los programas suelen compartir las mismas metas, pero reconocen que no podrán medir su contribución al cambio esperado en los resultados de la población; por lo tanto, sus objetivos generalmente son establecidos en términos de los resultados esperados de las actividades a nivel del programa.

DEFINICIÓN: Programa

Los diferentes tipos de actividad organizada propia de las instituciones de planificación familiar pueden clasificarse como se indica a continuación. Este manual se centra en el primer grupo, pero muchos de los principios y las herramientas de la evaluación se aplican a los dos.

Programa Nacional de Planificación Familiar

Definición: Todas las actividades organizadas para promover la planificación familiar en el sector público, el sector voluntario privado y el sector comercial en un país determinado.

Este conjunto de actividades puede ser coordinado por un cuerpo gubernamental o para-estatal como el Consejo Nacional de Planificación Familiar de Zimbabwe o el BKKBN (Badan Koordinasi Keluarga Berencana Nasional), el Consejo Nacional de Coordinación de la Planificación Familiar, de Indonesia. Más frecuentemente, las distintas organizaciones tales como el Ministerio de Salud (MS), los afiliados a la Federación Internacional de Paternidad Responsable, el sector comercial privado y otras organizaciones no gubernamentales (ONG) trabajan con cierta autonomía a pero comparten uno o más objetivos comunes, generalmente: mejorar las condiciones sociales y el bienestar humano ayudando a las parejas a ejercer su derecho a escoger el número y espaciamiento de sus hijos, mejorar la salud de mujeres y niños y retardar el crecimiento de la población. Idealmente, hay un documento formal para definir las metas nacionales y los mecanismos operativos para lograrlas, que sirve de base para evaluar el programa.

El programa de una institución

(Programa institucional)

Definición: El conjunto de actividades que una organización pública o privada ejecuta para lograr sus objetivos de PF.

Algunos ejemplos son el programa del MS, el programa del afiliado a la IPPF y el programa de mercadeo social (cuando no es parte de los anteriores). Estos programas suelen tener objetivos más específicos que el programa nacional. Por ejemplo, en contraste con el programa nacional que busca aumentar la prevalencia anticonceptiva general, el sector público puede estar específicamente preocupado por proveer métodos de planificación familiar a poblaciones de bajos ingresos.

Los evaluadores a veces enfrentan situaciones en las que los objetivos del programa no están establecidos en términos medibles. En este caso, parte del proceso de establecer los objetivos es ayudar a indicar los objetivos del programa en términos que se presten para la evaluación; o sea, "operacionalizar" los objetivos.

DESCRIPCIÓN DE CÓMO "DEBE" FUNCIONAR EL PROGRAMA

Componentes del programa

En el concepto más amplio, un programa de planificación familiar puede ser visto en términos de cuatro elementos distintos: insumos, procesos (o actividades), productos y resultados (Veney y Gorbach, 1993).

- Los **insumos** del programa son el conjunto de recursos (i.e., personal, establecimientos, espacio, equipo y suministros) que constituyen los materiales en crudo del programa.
- Los **procesos** del programa se refieren al conjunto de actividades en las cuales se utilizan los insumos en busca de los resultados esperados de éste. Los procesos del programa incluyen todas las operaciones de entrega del servicio² (gestión, capacitación, suministros y aspectos logísticos, información-educación-comunicación (IEC), investigación y evaluación) que el programa realiza para proveer los servicios de planificación familiar.
- Los **productos** del programa son los resultados obtenidos a nivel de programa a través de la ejecución de las actividades que usan los recursos de éste. Hay tres tipos de productos del programa:
 - ✓ los productos del área funcional, como el número de personas capacitadas o el número de charlas de IEC;
 - ✓ los productos del servicio, como el acceso a los servicios y la calidad de la atención; y
 - ✓ la utilización del servicio, como los años

² También conocidas como "áreas funcionales"

de protección pareja (APP) y el número de aceptantes nuevos.

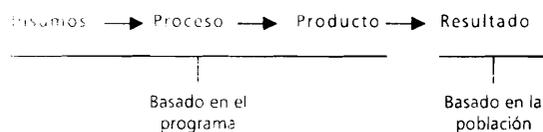
- Los **resultados** del programa son el conjunto de efectos esperados en la población debido a las actividades del programa y a la generación de productos del programa. Estos pueden dividirse en dos componentes: resultados intermedios³, y resultados a largo plazo.

Los resultados intermedios son los efectos en la población que están clara y estrechamente vinculados a las actividades y a los productos del programa. El resultado intermedio más conocido de un programa de planificación familiar es el uso de anticonceptivos. Los cambios en los resultados intermedios generalmente ocurren dentro de un período de 2 a 5 años desde el inicio del programa.

Los resultados a largo plazo son los efectos en la población que se dan a largo plazo y que se originan en los resultados intermedios. Algunos ejemplos de resultados a largo plazo de programas de planificación familiar son los cambios en la fecundidad y en la condición de la salud materno-infantil. Aunque las tasas de salud y fecundidad pueden cambiar abruptamente respondiendo a fuerzas externas, por lo general existe un lapso de tiempo considerable (5 a 10 años) entre el inicio del programa y la observación de los cambios en estas tasas.

Grafica II-1

Relación entre los componentes del programa



Los insumos (recursos del programa) son incorporados a los procesos (actividades del programa), los cuales generan productos (frutos del programa) y finalmente, resultados (cambios en la conducta de la población).

³ Alternativamente, se les llama "resultados de corto a mediano plazo".

DEFINICIÓN: Proyecto

Proyecto (dentro de una institución)

Definición: El conjunto de actividades con objetivos específicos que contribuyen a los objetivos generales de la institución. Los distintos proyectos de una institución, que pueden ser financiados con distintas fuentes, constituyen colectivamente su programa. Un proyecto suele tener atributos específicos en términos de:

- población objetivo (por ejemplo, adolescentes, hombres);
- método promovido (por ejemplo, métodos de acción prolongada);
- tipo de mecanismo de entrega de servicio utilizado (por ejemplo, distribución basada en la comunidad [CBD], obstetras tradicionales); y
- constelación de otros servicios sociales y de salud ofrecidos concurrentemente (por ejemplo, atención prenatal).

Uno también puede haber escuchado referencias al "programa de población" de una agencia donante. Por ejemplo, las agencias multilaterales como el Fondo de Población de las Naciones Unidas (FNUAP) y la IPPF, y agencias de desarrollo bilaterales de ciertos países (EEUU, Japón, Alemania, Inglaterra, Suecia, Canadá, entre otros) apoyan una cantidad de actividades en el sector poblacional en distintos países. Una agencia donante suele ofrecer fondos para uno o más proyectos dentro de una institución y en casos aislados, para el programa completo de población. Aunque las actividades financiadas por un donante pueden diferir mucho en cuanto a su naturaleza (por ejemplo, realización de un censo, apoyo de viajes de observación para encargados clave de la toma de decisiones, compra de suministros), el conjunto de actividades relacionadas con la población de una agencia donante suele llamarse su "programa de población". Este manual considera la evaluación de la financiación de donantes en la medida que estos fondos estén canalizados en un programa o proyecto de planificación familiar en un país en vías de desarrollo.

Los programas y los proyectos tienen una estrecha relación. El conjunto de proyectos múltiples de una institución constituye integralmente el programa institucional. Los diferentes programas institucionales integran lo que llamamos "el programa nacional". A su vez, estos programas generalmente están financiados parcialmente por el gobierno y posiblemente por una o más agencias donantes.

Definición del alcance de la evaluación

Los primeros tres – insumos, procesos y productos – se relacionan con las actividades y los resultados a nivel de programa. Los insumos, los procesos y los productos son medidos con datos del programa y de los establecimientos (ver Gráfica II-1). Los datos por programa provienen de la recolección rutinaria de información (i.e., estadísticas de servicio, registros clínicos de clientes y otros registros clínicos, registros administrativos, embarques de suministros y ventas) así como otra información recogida en el lugar donde se prestan los servicios (por ejemplo, encuestas a proveedores, observación de la interacción proveedor-cliente, inspecciones de ventas, observadores incógnitos) o de un estudio de seguimiento de clientes.

Los resultados (cambios intermedios y a largo plazo en la población) son medidos con

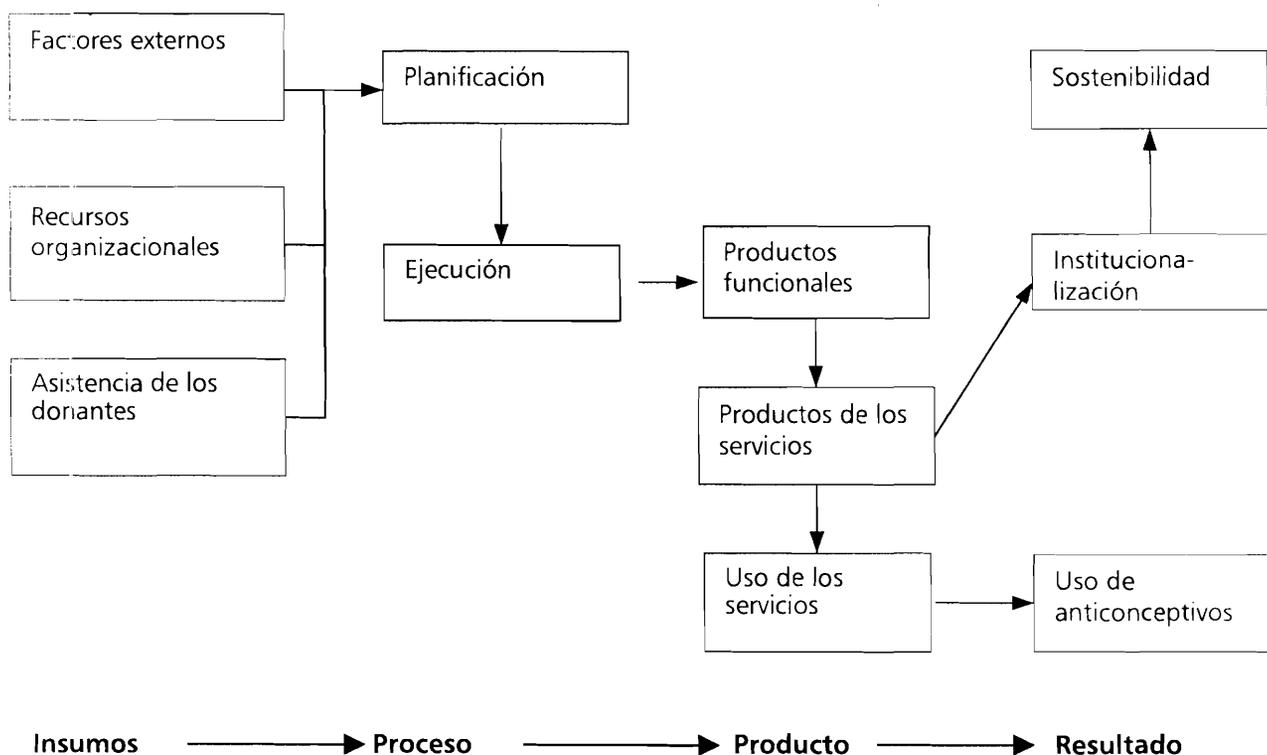
datos de la población recogidos de un área de captación o grupo social que el programa buscaba beneficiar. Puede ser un país, una región o un sub-grupo particular de la población (por ejemplo, los adolescentes).

Descripción de los vínculos causales

Una vez que los objetivos son establecidos, es importante definir cómo se espera que las actividades del programa logren estos objetivos. La secuencia causal esperada se muestra en su forma más simple en la Gráfica II-1 en la página anterior. El diagrama de flujo (modelo conceptual) en la Gráfica II-2 señala el contenido (factores, actividades, resultados, etc.) que están incluidos bajo cada una de las amplias categorías de insumo, proceso, producto y resultado.

Gráfica II-2

Componentes del impacto del programa



Fuente: Tsui y Gorbach, 1996

La Gráfica II-2 es útil para mostrar la relación entre los componentes del programa y los términos insumo-proceso-producto-resultado. El diagrama en la Gráfica II-2 es atractivo por su simplicidad, pero enmascara la complejidad de los programas reales. Por ejemplo, la "planificación" y la "implementación" comprenden un conjunto de actividades interrelacionadas de diferentes áreas funcionales (gestión, capacitación, suministros, aspectos logísticos, IEC). Al planear los vínculos entre el programa y los resultados esperados, es importante identificar los (a menudo múltiples) componentes de los "procesos".

La Gráfica II-2 muestra solo una parte de la situación: la contribución de las actividades de los programas (el "lado de la oferta") a los resultados deseados. En la realidad, el éxito de los programas también está determinado por los factores de la demanda (sociales, económicos y otras variables ajenas al programa que afectan la demanda del servicio). La Gráfica II-3 ilustra un marco conceptual completo utilizado por el Proyecto EVALUATION para describir rutas por las cuales los factores de la demanda y los factores de la oferta contribuyen a los productos del programa y a los efectos en la población. En la Gráfica II-3, los factores de la demanda se muestran en el lado superior izquierdo; los factores de la oferta se encuentran en el lado inferior izquierdo y están más elaborados en la Gráfica II-4. El texto del recuadro II-3 sobre el "Marco conceptual" explica con más detalle estas relaciones.

Las Gráficas II-2 (que muestra los componentes del programa en términos de insumo-proceso-producto-resultado) y la Gráfica II-4 (que detalla el entorno de oferta de planificación familiar) están muy relacionadas y en ambos casos excluyen los factores de la demanda en los hogares (o sea, si la gente desea servicios de planificación familiar).

Estos diagramas sirven como mapas para visualizar cómo uno espera llegar del punto A al punto B. Además, permiten a los funcionarios del programa y a los evaluadores aclarar la forma como los múltiples y variados ingredientes (como liderazgo del programa, regulaciones de políticas, estilo de gestión, calidad de capacitación del personal, rango de productos o anticonceptivos disponibles, frecuencia reposición de inventarios, cantidad y calidad de IEC y calidad de la atención) se juntan en un determinado programa, así como conceptualizar estos elementos en una forma sistemática que facilite la evaluación.

El modelo en las Gráficas II-3 y II-4 describe los vínculos esperados para un programa vertical de planificación familiar dirigido a la reducción de la fecundidad a largo plazo. Sin embargo, como se dijo en el Capítulo I, el "modelo" para los programas de planificación familiar ha evolucionado considerable-

MARCO CONCEPTUAL:

El impacto de los programas de PF sobre la fecundidad

El marco conceptual es la base para identificar indicadores de programa apropiados y especificar las trayectorias a través de las cuales los insumos del programa producen productos programáticos y, finalmente, cambios en la conducta de la población objetivo. Un marco conceptual describe los vínculos entre los insumos de los programas de PF y el cambio en la fecundidad, mostrado en las gráficas II-3 y II-4. El Manual de Indicadores para Evaluación de Programas de PF señala indicadores clave para cada cuadro del diagrama. El marco puede (y debe) ser modificado para evaluar los programas de PF con distintos objetivos o programas que tratan de alcanzar estos objetivos mediante una constelación diferente de servicios.

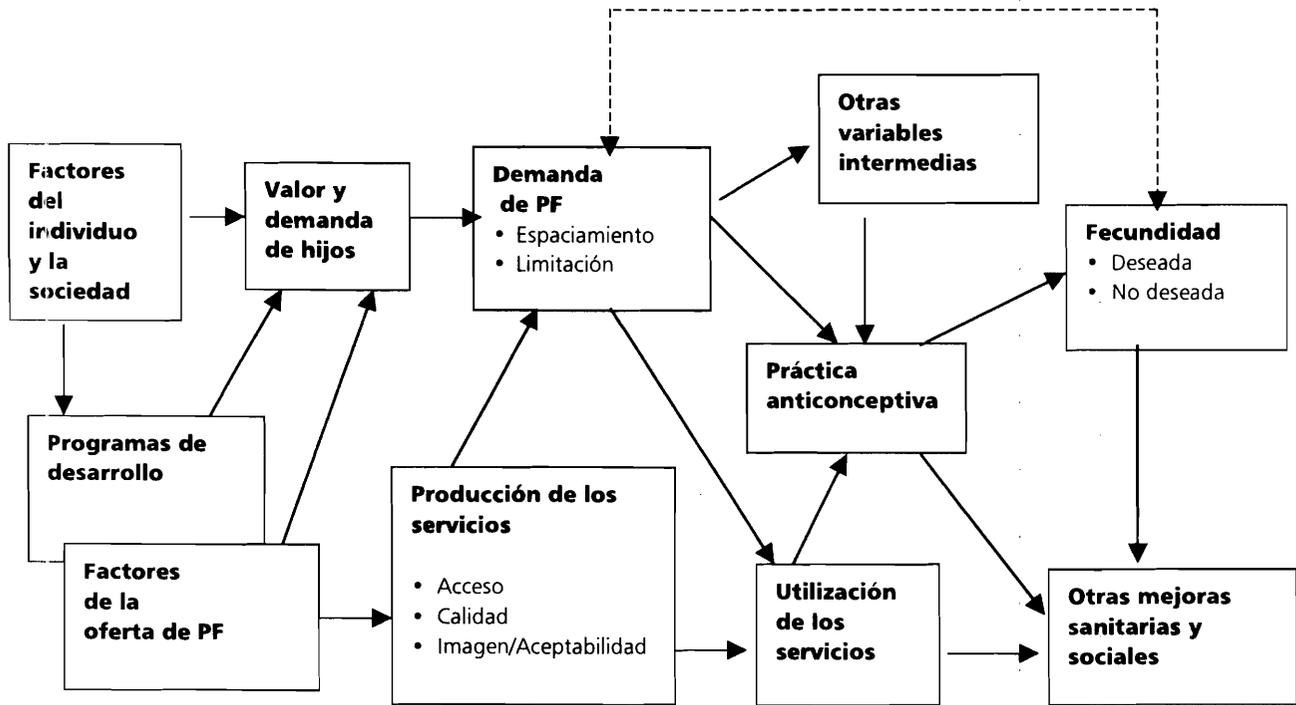
El marco reconoce que la fecundidad y otros impactos son consecuencia, tanto de la demanda, como de la oferta de servicios de PF. La demanda de hijos y la demanda de servicios de PF está afectada por varios factores políticos, socioeconómicos, culturales e individuales. Por eso, un aumento en la disponibilidad de servicios de PF es más probable que se transforme en mayores niveles de uso en un país donde estos otros factores ejercen una influencia positiva (en vez de negativa) sobre la demanda.

El entorno de la oferta de PF (Gráfica II-4) también está definido por los sistemas políticos y administrativos dentro de los cuales opera el programa. El apoyo político para el programa de PF, su financiamiento y el ambiente legal y regulatorio afectan la organización y el éxito del mismo. Los insumos al programa de PF en forma de personal, establecimientos y espacio, equipo y suministros, etc., se transforman mediante las actividades programáticas. Estas actividades programáticas consisten en el planeamiento y la implementación de las principales funciones del programa de PF: gestión, capacitación, distribución de anticonceptivos y suministros afines, esfuerzos de IEC e investigación y evaluación. En forma conjunta, los resultados de estas áreas funcionales crean los principales productos del programa: accesibilidad, calidad y servicios reputados de PF. Estos productos atraen clientes al programa y junto con la demanda de PF, determinan el impacto del programa sobre la población objetivo.

Los costos totales de los insumos usados para producir productos y resultados programáticos particulares pueden calcularse. Por ejemplo, los costos unitarios de personal, suministros y otros insumos pueden ser multiplicados por las cantidades utilizadas, sumados y luego relacionados a los productos o servicios producidos por el programa.

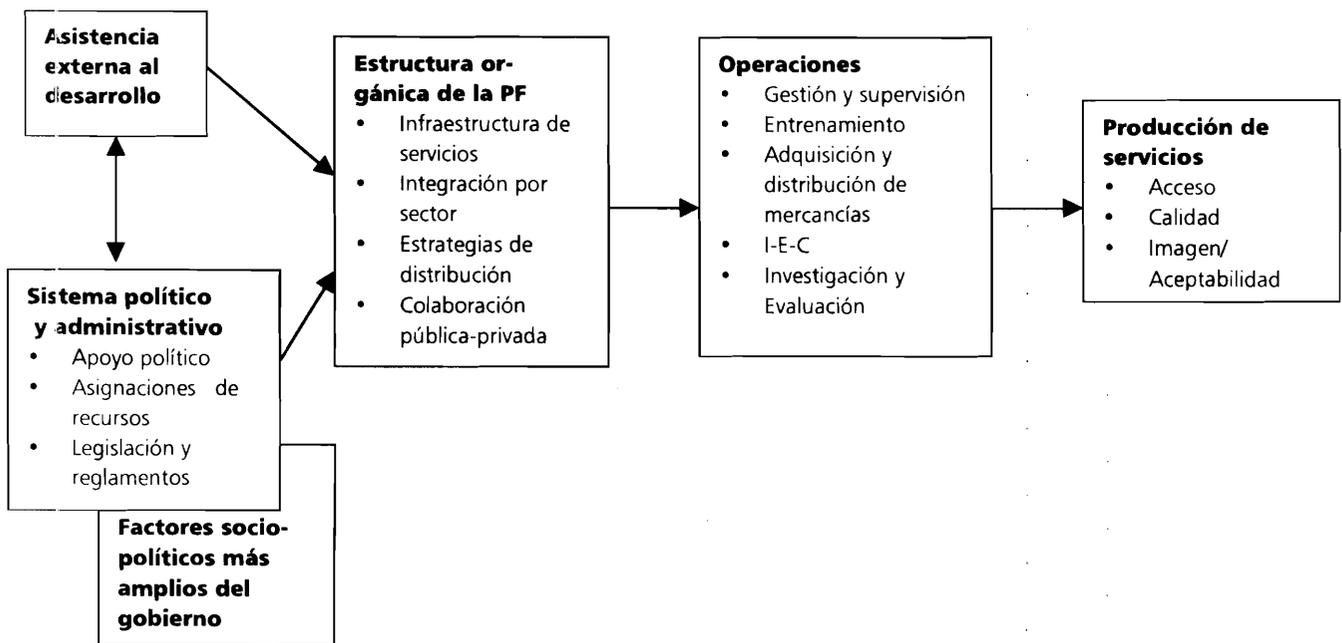
Gráfica II-3

Marco conceptual de la demanda de planificación familiar e impacto del programa sobre la fecundidad



Gráfica II-4

Marco conceptual de los factores de la oferta de planificación familiar



mente durante los últimos años para incluir otros servicios de salud reproductiva (relacionados con la maternidad segura, la lactancia, las enfermedades de transmisión sexual [ETS], los servicios post-aborto). Los resultados esperados no se relacionan solo con la fecundidad, sino también con la condición de la salud materno-infantil y el logro de las preferencias reproductivas individuales (según mediciones del índice ALIR⁴).

⁴ ALIR es una sigla de "Ayuda a las Personas a Lograr sus Intenciones Reproductivas" (Jain y Bruce, 1994). Mide el grado en el que los miembros de una población objetivo logran sus intenciones reproductivas (por ejemplo, tener otro hijo, evitar otro embarazo).

ESTABLECIMIENTO DE LOS OBJETIVOS DE LA EVALUACIÓN

Tipos alternativos de evaluación

Como lo indica el esquema de la Gráfica II-5, hay distintos tipos de evaluación, cada uno con un propósito diferente. Al diseñar una estrategia de evaluación, el evaluador necesita identificar las preguntas claves que desea contestar y, por lo tanto, el tipo de evaluación que se debe realizar.

Uno de los principales objetivos de este manual es diferenciar claramente entre el seguimiento del programa y la valoración del impacto. Como se refleja en las preguntas ilustrativas de la Gráfica II-5, el seguimiento del programa considera ciertas preguntas

Gráfica II-5

Tipos de evaluación

Tipos

Valoración de necesidades

Seguimiento de programas

Valoración del impacto

Pregunta(s) considerada(s)

¿Qué debería incluir el programa y cuál es la mejor forma de entrega para satisfacer las necesidades del grupo objetivo?

Insumos

- ¿Estuvieron disponibles los insumos (por ejemplo, equipo, suministros, materiales, personal) para el programa en las cantidades y en los tiempos especificados en el plan del programa?

Procesos

- ¿Se efectuaron las actividades programadas conforme al plan?
- ¿En qué medida fueron efectuadas apropiadamente?

Productos

- ¿Ocurrieron los cambios esperados al nivel de programa, en términos de:
 - √ acceso a los servicios
 - √ calidad de la atención
 - √ utilización del servicio

Resultados

- ¿Ocurrió el cambio esperado en la población (no necesariamente atribuible al programa)?

Costos

- ¿Cuál fue el costo extra de: expandir una actividad, producir una unidad mayor de producto y lograr el cambio ocurrido?

¿Cuál y cuánto cambio ocurrió (al nivel del programa o de la población) que sea atribuible al programa?

Definición del alcance de la evaluación

diferentes, una de las cuales es: ¿se dio un cambio? Sin embargo, sin la valoración del impacto, uno no puede contestar la pregunta: ¿se dio el cambio debido al programa? El resto de este capítulo compara y contrasta el seguimiento del programa frente a la valoración del impacto.

Propósitos del seguimiento del programa

El seguimiento⁵ se refiere a un conjunto de técnicas de evaluación, cada una de las cuales mide algún aspecto del desempeño del programa. El seguimiento del programa tiene dos propósitos distintos:

- mejorar los programas identificando aquellos aspectos que funcionan de acuerdo con el plan y aquellos que requieren correcciones a mitad de camino; y
- rastrear los cambios en los servicios ofrecidos

(frutos del servicio) y los resultados deseados.

El seguimiento incluye medir la situación actual y el cambio a través del tiempo en cualquiera de los componentes del programa.

A nivel del programa:

- Insumos
- Productos: Productos funcionales, productos del servicio (o idoneidad del servicio), utilización del servicio

A nivel de la población:

- Resultados: Resultados intermedios (por ejemplo, prevalencia anticonceptiva), resulta-

⁵ El seguimiento, o monitoreo, de las actividades relacionadas con las variables a nivel del programa también se llama "evaluación del proceso".

Gráfica II-6

Uso del marco conceptual para identificar los aspectos que debe considerar el seguimiento de programas

Componente del programa	Cantidad del componente	Calidad del componente	Costo del componente
Insumos	¿Que tipo y nivel de recursos fue asignado a esta intervención?	¿Había personal calificado disponible para implementar las actividades?	¿Cuál fue el costo unitario de cada recurso? ¿el costo total de cada recurso? ¿el costo total del programa?
Productos Áreas funcionales (por ejemplo, capacitación)	¿Cuántas personas fueron capacitadas, por categoría de personal?	¿Fue capaz el personal capacitado de desempeñar tareas en forma competente 6 meses después de la capacitación?	¿Cuál fue el costo por participante-día de la capacitación?
Productos Frutos del servicio	Acceso: ¿Aumentó el número de PPS proveedores de servicio?	Calidad: ¿Mejoró la calidad de la atención con el tiempo?	¿Cuál fue el costo agregado de aumentar los números de PPS? ¿de mejorar la calidad de la atención?
Productos Utilización del servicio	¿Aumentó el número de aceptantes nuevos o APP?	¿Ha aumentado el porcentaje de clientes que regresan por citas de seguimiento?	¿Cuál fue el costo agregado asociado con el aumento de aceptantes nuevos? ¿con el aumento en APP?
Resultados Resultados intermedios	¿Se dio un cambio en la conducta clave (por ejemplo, prevalencia anticonceptiva) en la población objetivo?	¿Se dio un cambio en la conducta clave (por ejemplo, provisión de servicio de calidad) en la población objetivo?	¿Cuál fue el aumento en los costos asociado con el cambio en la prevalencia anticonceptiva?
Resultados De largo plazo	¿Lograron las mujeres sus intenciones reproductivas?	¿Cambiaron las tasas de fecundidad con el tiempo?	¿Cuál fue el costo de lograr el cambio en la fecundidad?

dos a largo plazo (por ejemplo, tasa global de fecundidad).

Algunos lectores pueden sorprenderse al ver el cambio en la población categorizado como seguimiento más que como valoración del impacto, puesto que en muchos programas el "impacto buscado" es una caída en la fecundidad, la mortalidad o la morbilidad. Sin embargo, no es imposible atribuir causa y efecto o determinar el porcentaje de cambio atribuible al programa basado en un simple seguimiento de las tendencias en los datos. Aun cuando estas variables se refieren a los cambios finales que el programa desea producir, el rastreo de estas tendencias cae bajo la amplia categoría de seguimiento del programa (en ausencia de un diseño de estudio o un análisis estadístico que determine la causalidad).

Idealmente, el seguimiento del programa incluirá tanto las técnicas de investigación cuantitativas como las cualitativas. La recolección de datos puede incluir cualquiera de las técnicas estándar usadas en la investigación de las ciencias sociales (encuestas, grupos focales, entrevistas en profundidad, observación, informantes clave y otras) además del análisis de los datos del programa (por ejemplo, estadísticas de servicio).

La Gráfica II-6 ilustra cómo el marco conceptual puede ser usado para identificar y organizar las preguntas clave que deben ser consideradas respecto a la cantidad, calidad y costo de los programas. Por simplicidad, se muestra solo un aspecto en cada cuadro de la Gráfica II-6. En una verdadera evaluación, uno seleccionaría las cajas que le interesan e identificaría los indicadores (variables) para medir cuán bien funciona el programa en cada aspecto. Una evaluación exhaustiva del programa trataría de cubrir buena parte de estas cajas. En cambio, un estudio especial podría centrarse en una sola, como la calidad de la atención en un grupo de PPS.

El seguimiento de programas es mucho más común que la valoración del impacto en la evaluación de programas nacionales de planificación familiar⁶. A nivel del programa, los administradores usan las tendencias de los datos del programa (número de clientes nuevos, cantidad de APP, número de consultas según propósito de la consulta, número de contactos con adolescentes y otros) para valorar el progreso e identificar las áreas que requieren mejoras. Además, rastrean tendencias en la población durante un periodo de tiempo (en prevalencia anticonceptiva, combinación de métodos, tasa global de fecundidad, duración media de intervalos entre nacimientos, y otras) para valorar el progreso hacia los obje-

tivos intermedios y de largo plazo. En general, aunque el seguimiento de productos y resultados no demuestra causa y efecto, suele existir un supuesto de asociación razonable.

Propósito de la valoración del impacto

El propósito de la valoración del impacto es medir el grado de cambio atribuible a un programa o intervención determinado. A diferencia del seguimiento, que simplemente rastrea el cambio, la valoración del impacto se ocupa de la cuestión de causalidad.

La palabra "impacto" se escucha frecuentemente en las organizaciones cuya misión es producir un cambio de conducta. A menudo el personal del programa usará este término para describir el cambio deseado que espera lograr mediante una intervención, y de esta forma describe su propio marco conceptual que vincula los insumos de los procesos con los productos y los resultados. Por ejemplo, el personal del proyecto de mercadeo social puede especular respecto al impacto de sus actividades de promoción sobre las ventas. El personal de IEC espera que su campaña nacional tenga el impacto deseado de poner a los clientes en contacto con los servicios.

Es fácil describir este tipo de relación causa-efecto, pero en general es más difícil demostrar el impacto en forma empírica. La razón, como se mencionó antes, es que aun cuando el cambio deseado ocurre, no necesariamente o exclusivamente puede ser atribuido a la intervención del programa. Por ejemplo, el director del programa de planificación familiar puede citar el aumento en la prevalencia anticonceptiva a través del tiempo como "prueba del impacto del programa nacional". En lenguaje común, muchos aceptarán esta declaración al pie de la letra. En términos técnicos, sin embargo, la palabra "impacto" indica que la evaluación se basa en un diseño de estudio, demostrando no solo el cambio, sino causa-efecto.

⁶ Existen varios estudios del tipo investigación de operaciones que usan diseños cuasi-experimentales para vincular algún aspecto de la entrega del servicio con un cambio en el programa o en la población. Muchos de estos estudios están diseñados para demostrar causa y efecto. Sin embargo, generalmente son realizados en un área geográfica específica, y por lo tanto, no constituyen una evaluación del programa nacional.

Esta diferencia se ilustra en la Gráfica II-8. El diagrama está incluido para hacer énfasis en el punto de que una evaluación puede demostrar el cambio esperado, pero no es necesariamente atribuible al programa.

En resumen, la evaluación de la planificación familiar a menudo se limita solo al seguimiento de programas. Pero si uno hace la valoración del impacto, el análisis debería incluir mediciones

del desempeño del programa. De hecho, las variables usadas para valorar el impacto (y por lo tanto, la información recolectada) son las mismas que los indicadores usados para el seguimiento del programa. Lo que las diferencia es el diseño de la evaluación y las técnicas analíticas utilizadas.

Este manual estimula a los evaluadores a distinguir entre el seguimiento de programas y la valoración del impacto en el diseño de los planes de evaluación y en los informes de resultados. Además, invita a los miembros de la comunidad de evaluadores a renovar esfuerzos para llevar a cabo valoraciones de impacto en el sentido riguroso del término y provee una guía metodológica para hacerlo.

A modo de examen previo de los argumentos más detallados que se ofrecen en el Capítulo IV, la Gráfica II-9 resume los "métodos preferidos" para valoración del impacto y los datos requeridos para cada uno.

Factores para decidir cuándo valorar el impacto

En la práctica, la pregunta NO es si se debe hacer seguimiento de programas O valoración de impacto. Más bien se trata de realizar la valoración de impacto además del seguimiento de programas. La decisión depende básicamente de lo siguiente:

- la etapa en que se encuentra la implementación del programa;
- la disponibilidad de recursos para la recolección y el análisis de datos adicionales, y
- la necesidad de demostrar el impacto (por ejemplo, para justificar la continuación del financiamiento).

Etapa en que se encuentra la implementación del programa

El plan de evaluación para un programa determinado debería definirse de entrada. (En el caso de la USAID, generalmente es al inicio de un ciclo de cinco años de financiación.) Una vez que el programa está en marcha, normalmente es demasiado tarde para aplicar en forma retroactiva un diseño de evaluación que permita una medición confiable del impacto. Además, el desarrollo de un plan de evaluación desde el principio del programa asegura que los datos necesarios están siendo o serán recolectados.

ESQUEMA DE RESULTADOS:

Enfoque de la USAID para el seguimiento de programas

La USAID actualmente da seguimiento al desempeño de programas usando un esquema de resultados como el mostrado en el ejemplo de la Gráfica II-7 de la misión USAID en Marruecos.

En este sistema, una misión desarrolla uno o más objetivos estratégicos. El objetivo estratégico es el resultado más ambicioso que la unidad operativa, junto con sus asociados, está dispuesto a asumir. Constituye la norma sobre la cual se va a juzgar su desempeño. En los sectores de población y salud, los objetivos estratégicos pueden ir, desde una caída en la fecundidad y la mortalidad, hasta un aumento en el uso de algunos servicios de planificación familiar, sobrevivencia infantil y salud reproductiva o materna.

Seguidamente desarrolla un esquema de resultados para cada objetivo estratégico. Este esquema ilustra las rutas causales que llevan al logro de los objetivos, así como a los resultados requeridos en etapas anteriores para su logro. Este esquema también es útil para comunicar las premisas subyacentes de la estrategia. El esquema de resultados forma parte esencial del plan estratégico que debe ser desarrollado por cada misión o unidad operativa en el extranjero que usa fondos del programa. Además de los objetivos estratégicos y el esquema de resultados, el plan señala los enfoques que van a usarse para lograr los objetivos, los indicadores para medir resultados y la frecuencia de los reportes de cada uno.

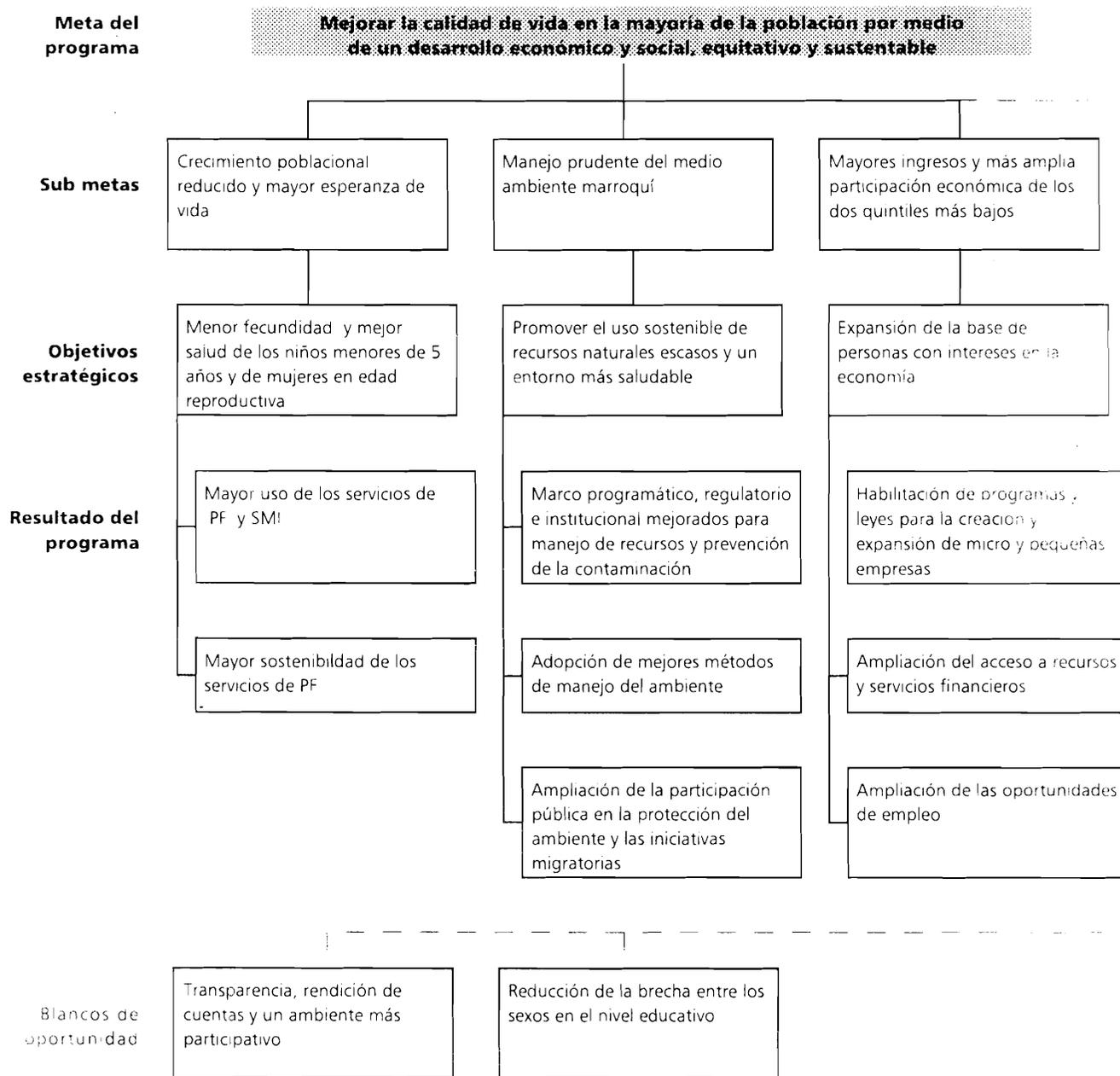
Cada objetivo estratégico y cada resultado debe contemplar al menos un indicador de desempeño. Estos indicadores deben ser claros, precisos y objetivamente medibles. Estos indicadores son medidos en un punto de partida, o línea de base, y posteriormente en uno o más puntos durante el ciclo del proyecto. Además, el esquema de resultados debería incluir cualquier resultado clave producido por otros asociados, como por ejemplo, las ONG, el gobierno del país anfitrión, otros donantes y clientes.

Las unidades operativas documentan su progreso hacia el logro de los resultados establecidos en un informe anual que contiene una parte descriptiva y cuadros de los indicadores de desempeño.

Grafica II-7

Ejemplo de un sistema de seguimiento para un programa nacional

Programa de USAID en Marruecos



Fuente: USAID/Morocco Country Program Strategy

Para los evaluadores que diseñaron un plan de evaluación antes del inicio del programa, la "etapa de implementación" no es problema. Pero para aquellos que reciben la tarea cuando el programa (o el ciclo de financiación vigente) ya está en marcha, las opciones son más limitadas. Bajo condiciones menos que óptimas, el evaluador tiene las siguientes opciones:

- limitar la evaluación solo al seguimiento del programa;
- tratar de demostrar vínculos entre los componentes del programa y las medidas de producto o resultado, aún cuando no sea posible mostrar causa-efecto en forma inequívoca (ver Capítulo IV); o
- identificar diseños más solventes que puedan ser usados en etapas posteriores de actividad.

Disponibilidad de fuentes de datos existentes y potenciales

La disponibilidad de datos también debe considerarse al tomar la decisión de realizar análisis de impacto. Los requerimientos de datos para los tres "métodos preferidos" son algo diferentes, como se resume brevemente en la Gráfica II-9 y se describe en detalle en el Capítulo IV. Esto es a menudo un factor determinante para la decisión.

ESQUEMA DEL ALCANCE DE LA EVALUACIÓN

Finalmente, es importante presentar las decisiones de qué evaluar en forma escrita. Como se dijo en el Capítulo I, la descripción del alcance de la evaluación debería incluir:

- La metas y los objetivos del programa
- El marco conceptual que señale los vínculos entre los insumos, los procesos, los productos y los resultados
- Los objetivos de la evaluación
- Seguimiento del programa
- Valoración del impacto

No hay necesidad de presentar con gran detalle estos tres temas de un plan de evaluación (aunque es importante hacer la diferencia entre el seguimiento de programas y la valoración del impacto). Más bien, lo relevante es la claridad conceptual.

Con una descripción clara de los tipos de evaluación a realizar, uno sigue con los detalles del enfoque metodológico. El capítulo siguiente (III) ofrece lineamientos para el seguimiento de programas.

Ciertos conceptos del Capítulo III (por ejemplo, la selección de indicadores, las fuentes de datos) son igualmente aplicables a la valoración del impacto. Sin embargo, los problemas de diseñar una medición de impacto son tan diferentes que se discuten por separado en el Capítulo IV.

Gráfica II-8

La secuencia de los tipos de evaluación

Proceso	Resultados	Impacto
Cuán bien funcionó el programa: • ¿Cuáles y cuántas actividades se ejecutaron? (valoración cuantitativa) • ¿Cuán bien fueron implementadas? (valoración cualitativa)	Ocurrió el cambio esperado: • ¿a nivel del programa? (productos) • ¿a nivel de la población? (resultados)	¿En qué medida puede el cambio ser atribuido a la intervención? Basado en un modelo teórico y demostrado por: • un diseño experimental o cuasi-experimental; o • análisis multinivel longitudinal

← Seguimiento al desempeño del programa → ← Valoración del impacto →

Gráfica II-9

Métodos preferidos para valorar el impacto de programas de PF

Los textos y manuales anteriores sobre la evaluación del impacto de los programas de PF han descrito varios métodos o enfoques (United Nations, 1979, 1982, 1985; Hermalin, 1982; Sherris et al., 1985; Buckner et al., 1995). Sin embargo, el consenso desarrollado por el Proyecto EVALUATION es que solo tres enfoques demuestran adecuadamente la causalidad (i.e., que el cambio observado es atribuible al programa). Estos tres enfoques preferidos se exponen con detalle en el Capítulo IV y son los siguientes:

Experimentos aleatorizados: el diseño con pretest, postest y grupo control

Este diseño es considerado como la "norma de oro" para valorar el impacto, porque, si se implementa apropiadamente, responde a la pregunta: ¿Qué habría sucedido sin el programa? Comparando el cambio que ocurre en las poblaciones experimentales con las poblaciones control, uno puede medir qué proporción del cambio es atribuible al programa. La principal limitación de este método se relaciona con su factibilidad y la aceptación política de realizar experimentos.

Datos requeridos: Se deben obtener datos de dos grupos: aquellos que reciben y aquellos que no reciben la intervención; además, los sujetos (o grupos de sujetos, como pueblos) deben ser asignados aleatoriamente al grupo experimental o al grupo control. Los datos (sean estadísticas de servicio o datos de encuestas) son recolectados tanto antes ("pre") como después ("pos") de la intervención para los dos grupos.

Aunque este diseño suele ser criticado por la dificultad de ejecutarlo, un enfoque prometedor es la "localización" de las intervenciones. Específicamente, si uno se encuentra al principio del ciclo de un proyecto que comprenderá nuevas intervenciones o un estudio piloto antes de la expansión a gran escala, es posible asignar aleatoriamente los establecimientos o áreas que reciben y no reciben la intervención y luego comparar los resultados de los dos grupos. Este enfoque puede ser particularmente útil si los recursos son muy limitados como para iniciar las intervenciones en todas las áreas simultáneamente.

Cuasiexperimentos: "diseño de grupo control no equivalente"

Este enfoque es similar al diseño de grupo control pretest-postest expuesto anteriormente, excepto que los sujetos (o las unidades administrativas como los pueblos) NO son asignados aleatoriamente al grupo experimental o al de control. Mas bien, el investigador selecciona o asigna los pueblos con base en la similitud de sus características socio-demográficas u otras características clave. Por lo demás, los datos requeridos son los mismos del diseño aleatorio.

Modelos de regresión longitudinal de nivel múltiple

Esta metodología usa técnicas estadísticas de las ciencias sociales, aplicadas a los tipos de datos disponibles para la evaluación de los programas de PF. Brevemente, está diseñada para demostrar empíricamente que los insumos del programa resultaron en cambios al nivel de programa (por ejemplo, más y mejores servicios) Y que estos productos propiciaron un cambio en la conducta deseada, como la práctica anticonceptiva (o sea, el resultado intermedio). En algunos casos, el resultado a largo plazo es preferido al intermedio, o bien, se usan ambos (por ejemplo, la tasa de fecundidad).

Datos requeridos: Los datos requeridos para este tipo de análisis son las encuestas de la población objetivo (que puede ser el país completo) en dos fechas distintas realizadas en los mismos conglomerados, así como datos sobre la red de entrega de servicios en esas dos fechas. En algunos casos, es posible hacer análisis similares de un estudio realizado en una fecha determinada si la información sobre las variables clave puede ser reconstruida retrospectivamente para una fecha anterior. Las encuestas de tipo DHS ofrecen información sobre la población objetivo. Los datos sobre la red de entrega del servicio pueden ser obtenidos del módulo de disponibilidad de servicios (SAM) de la DHS o de un estudio de Análisis de Situación. Pocos países tienen actualmente datos por hogares y por establecimientos para el mismo conjunto de conglomerados muestrales. Sin embargo, muchos países ya tienen una encuesta DHS con un módulo de disponibilidad de servicio y están planeando una segunda DHS; en este caso, agregando el segundo SAM se obtendría los datos necesarios para este método.

Nota: Los métodos anteriores se exponen en un orden que facilita su presentación, no necesariamente en orden de preferencia. Las condiciones bajo las cuales los distintos métodos serían preferidos se discuten en el Capítulo IV.

Capítulo III

Enfoque metodológico: seguimiento de programas

- Aclaración del propósito principal del seguimiento
- Identificación de los componentes que van a recibir seguimiento
- Definición de indicadores relevantes
- Identificación de fuentes de datos
- Diseño de un formato para la presentación de resultados
- Resumen del enfoque metodológico

ENFOQUE METODOLÓGICO: SEGUIMIENTO DE PROGRAMAS

Este capítulo supone que el evaluador ha completado el primer paso del desarrollo de un plan de evaluación: la definición del alcance de una evaluación (i.e., solo dar seguimiento al desempeño del programa, solo medir el impacto o ambas).

Si la decisión incluye el seguimiento de programas, los próximos pasos del desarrollo de un plan de evaluación consisten en definir:

- el propósito principal del seguimiento,
- los componentes del programa que deben recibir seguimiento,
- el o los diseños de estudio,
- los indicadores,
- las fuentes de datos, y
- el formato para la presentación de los resultados.

ACLARACIÓN DEL PROPÓSITO PRINCIPAL DEL SEGUIMIENTO

Como se dijo en el Capítulo II, el seguimiento de programas tiene dos propósitos principales:

- mejorar los programas identificando aquellos aspectos que funcionan de acuerdo con el plan y aquellos que requieren correcciones a mitad de camino, y
- rastrear (y demostrar) los resultados a nivel del programa o de la población.

Por ejemplo, ciertas técnicas de evaluación están diseñadas específicamente para mejorar el desempeño. En un esfuerzo por mejorar la calidad de la atención en los programas de planificación familiar, algunos países han experimentado con la "Técnica Eficiente para el Proveedor Orientada hacia el Cliente" (COPE), que es un instrumento de autoevaluación diseñado para uso local (AVSC International, 1995). Los datos NO son agregados a un ni-

vel superior, sino que más bien son analizados por los proveedores del servicio para identificar los cambios que se pueden introducir localmente para corregir problemas identificados por el ejercicio. Esta técnica cualitativa representa un enfoque prometedor para mejorar los programas de abajo hacia arriba, pero no satisface las necesidades de un oficial de un programa regional para el rastreo de resultados de los PPS en la red del programa.

En cambio, hay técnicas que dan seguimiento a los logros de un programa pero ofrecen relativamente poca información sobre sus ventajas y desventajas. Un ejemplo son los informes de rutina de las estadísticas de servicios (número de aceptantes nuevos, número de visitas clínicas, número de años de protección de pareja y otros) recolectados para los PPS y agregados a nivel central. Este tipo de información es valiosa para rastrear las tendencias a través del tiempo, aunque por sí solos no indican por qué el programa está o no está logrando los resultados deseados.

Los administradores de programas y las agencias donantes están generalmente interesados en ambos tipos de seguimiento. Las agencias donantes casi siempre quieren datos cuantitativos sobre los "resultados", pero cada vez más se interesan por saber que el programa tiene formas de obtener datos, generalmente cualitativos (por ejemplo, grupos focales, entrevistas profundas, listas de observación y otros), los cuales serán utilizados directamente para mejorar el programa. De este modo, el seguimiento de programas suele consistir en una combinación de actividades de evaluación que en su conjunto proveen información sobre el programa como un todo.

IDENTIFICACIÓN DE LOS COMPONENTES QUE DEBEN RECIBIR SEGUIMIENTO

La decisión de a cuáles componentes de un programa se debe dar seguimiento depende, en parte, del propósito principal de la evaluación: mejorar el programa, rastrear resultados

Gráfica III-1

Propósito del seguimiento y componentes de interés

Propósito	Componentes del programa
Mejorar el programa	<p><i>Productos funcionales</i> Número y calidad de actividades realizadas en distintas áreas de gestión, supervisión, capacitación, suministros, logística, IEC, sistemas de registro.</p> <p><i>Productos del servicio</i> Acceso a los servicios, calidad de la atención e imagen del programa.</p>
Rastreo de resultados	<p><i>Utilización del servicio</i> Resultados producidos al nivel de programa (por ejemplo, número de aceptantes, número de visitas, APP, etc.).</p> <p><i>Resultados</i> Cambios intermedios o de largo plazo al nivel de la población (por ejemplo, prevalencia anticonceptiva, intervalo medio entre nacimientos, TGF).</p>

o ambas (Ver Gráfica III.1). Para el administrador del programa, no es un asunto de "uno u otro", él necesitará ambos. En cambio, las agencias donantes a menudo están más interesadas en rastrear los resultados del programa (aunque en general promueven bastante entre las agencias ejecutoras la realización de evaluación dirigida a la identificación de áreas que requieren mejorar).

La evaluación de las distintas operaciones del servicio ("áreas funcionales") es particularmente útil cuando se hace con suficiente antelación en el proceso de ejecución para permitir las correcciones a mitad de camino. La utilización del servicio generalmente se rastrea continuamente durante la ejecución del proyecto. Los resultados, en cambio, son generalmente medidos en (dos o más) intervalos para determinar el cambio a través del tiempo.

Los siguientes puntos del seguimiento deben ser tomados en cuenta:

- El seguimiento de programas emplea ideal-

mente tanto las técnicas cuantitativas como las cualitativas.

- No es práctico tratar de hacer una evaluación detallada de TODOS los aspectos del programa. Más bien, es importante priorizar los puntos para los cuales la información será más útil a la institución y crucial para el éxito del programa.
- El uso de un marco conceptual para definir las rutas para lograr los resultados deseados es igualmente aplicable a los proyectos y a los programas nacionales (ver recuadro III-1).
- Las actividades de evaluación suelen ser de diversa naturaleza y periodicidad. Algunas pueden hacerse de rutina (por ejemplo, la recolección y el registro de estadísticas de servicio), otras periódicamente (por ejemplo, visitas simuladas o con observadores incógnitos para valorar la calidad de la atención) y otras una sola vez (por ejemplo, el

métodos anticonceptivos).

- El personal administrativo clave debería tomar el liderazgo en la decisión de a cuáles aspectos del programa dar seguimiento, no los especialistas en evaluación⁷. De esta forma, la evaluación servirá realmente a las necesidades de la institución.

Para llegar a la decisión de cuáles aspectos específicos del programa se deben evaluar, es útil identificar el ámbito de temas posibles y luego priorizarlos. La Gráfica II-6 del capítulo anterior ofrece un marco útil para considerar opciones. Esta gráfica la hemos repetido en la Gráfica III-2 dejando espacios en blanco, los cuales pueden ser utilizados por una organización para identificar y priorizar los aspectos a evaluar.

La tarea es, entonces, identificar los aspectos a evaluar más apremiantes para el programa en estudio. Por ejemplo, casi todos los administradores querrán una retroalimentación regular sobre los "resultados" que reflejen el desempeño del establecimiento (por ejemplo, el número de aceptantes nuevos, el número de APP). Además, a menudo hay asuntos específicos que deben ser atendidos (por ejemplo, la necesidad de dar seguimiento a la calidad). Los administradores también querrán saber cuán productivo es su personal para generar productos, como visitas a clientes o APP; por ejemplo, ¿Cuántos clientes reciben servicios cada día? ¿Cuál fue el costo del recurso humano para producir un APP el año anterior? ¿Cómo varió el costo por APP por tipo y ubicación de la clínica?

DEFINICIÓN DE LOS INDICADORES RELEVANTES

El propósito de los indicadores

Los indicadores son variables que miden los distintos aspectos de un determinado programa: los insumos, los procesos, los productos y los resultados.

Se puede asignar un valor numérico a un indicador (un porcentaje, un promedio, un rango, un número absoluto) o una calificación de sí/no (por ejemplo, "presencia" frente a "ausencia"). En algunos casos, el valor conlleva una interpretación ampliamente compartida (por

ejemplo, una tasa de prevalencia anticonceptiva de 75% es "alta"). En otros casos, el valor (número) es muy útil en un sentido relativo, en comparación con programas similares o el mismo programa en una etapa anterior (por ejemplo, tendencias observadas en los APP).

A riesgo de sobre simplificar, se puede decir que el seguimiento de programas consiste en la medición de cuán bien funciona el programa en una o dos de las cajas del marco conceptual (Gráficas II-3 y II-4). El marco conceptual ilustra cómo debería funcionar el programa teóricamente para lograr los resultados deseados en el programa y en la población. El seguimiento del programa cuantifica lo que realmente ocurre en cada nivel (de insumos, procesos, productos y resultados). Aunque algunos pueden considerar que el marco conceptual es un mero ejercicio académico, de hecho este marco es muy práctico, dado que identifica las áreas para las cuales el evaluador podría seleccionar indicadores.

Un menú de posibles indicadores para evaluar programas de planificación familiar se ofrece en el Manual de Indicadores para la Evaluación de Programas de Planificación Familiar (Bertrand et al., 1994). En la medida que los programas de planificación familiar se extienden para incluir otros aspectos de la salud reproductiva, el posible número de indicadores adicionales relevantes aumenta (en Bertrand y Tsui, 1995, se describen indicadores para: embarazo seguro, ETS/SIDA, nutrición de mujeres, lactancia y servicios de salud reproductiva para adolescentes).

En una evaluación concreta, uno debería priorizar los indicadores basados en los objetivos específicos del programa y seleccionar un conjunto manejable para satisfacer las necesidades particulares de la situación. Es, en suma, esencial identificar los aspectos claves que deben ser considerados en la evaluación y seleccionar los indicadores correspondientes.

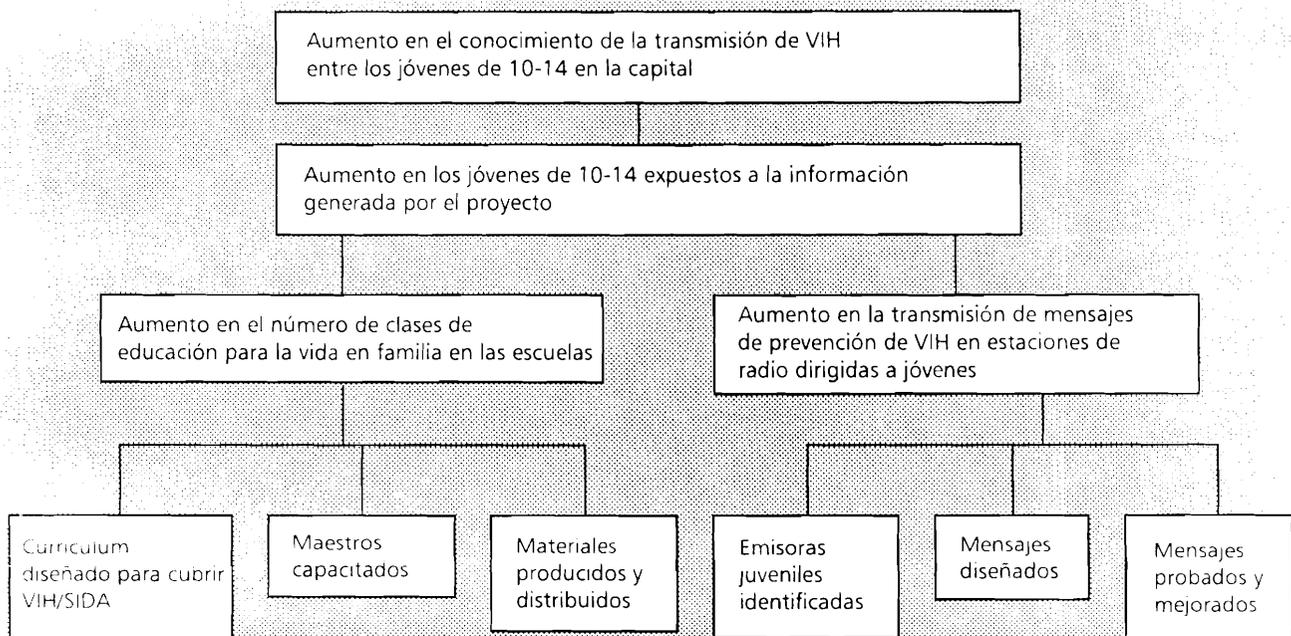
⁷ Es importante, sin embargo, para el especialista en evaluación, involucrarse de cerca en estas discusiones, para ofrecer información que pueda influir la decisión (por ejemplo, los costos y el tiempo aproximado requerido para las distintas actividades de recolección de datos, las fuentes alternativas de información, los sesgos inherentes a los distintos métodos de recolección de datos y otros).

Los esquemas conceptuales son útiles para la evaluación de proyectos y programas

El Capítulo II describió en detalle el valor de un esquema conceptual en el diseño de la evaluación de un programa nacional. Sin embargo, es igualmente útil contar con un "mapa vital" que describa las rutas para el logro de los resultados deseados en proyectos más pequeños. Aunque el cambio esperado generalmente está expresado en términos de una práctica o una conducta, eso puede no ser el caso en algunos programas, como se muestra en el ejemplo siguiente sobre juventud y prevención del SIDA.

En este ejemplo el supuesto implícito es que el conocimiento

obtenido en una edad temprana influirá de hecho la conducta de los jóvenes cuando llegan a estar a riesgo de contraer una infección VIH. El alcance de la evaluación y el período del proyecto podrían, empero, impedir la verificación de este supuesto. Más bien, como se refleja en el esquema siguiente, el proyecto está diseñado para aumentar el conocimiento sobre la transmisión del VIH y la evaluación se centraría en este resultado. Cada resultado intermedio puede recibir, a su vez, seguimiento para asegurar que el proyecto está siendo implementado de acuerdo con el diseño.



Gráfica III-2

Esquema para identificar y priorizar los aspectos del desempeño del programa que deben recibir seguimiento

Componente del programa	Cantidad del componente	Calidad del componente	Costo del componente
Insumos			
Productos Áreas funcionales: Gestión Capacitación Suministros/logística IEC Investigación/evaluación			
Productos Productos del servicio	Acceso:	Calidad:	
Productos Utilización del servicio			
Resultados Resultados intermedios			
Resultados Resultados a largo plazo			

Características de un buen indicador

Los buenos indicadores comparten algunas características importantes. Posiblemente la más importante es la **validez**: ¿Mide el indicador lo que se supone que debe medir? Por ejemplo, las preguntas de encuestas sobre el tamaño ideal de familia no están pensadas usualmente como medidas válidas de la demanda de fecundidad debido a los sesgos de las respuestas. Por otra parte, la intención declarada de tener más hijos se considera un indicador válido de demanda puesto que tiende a estar menos influenciado por los errores de las respuestas. Las ventas de anticonceptivos en un programa de mercadeo

social puede ser un indicador válido del éxito del mercadeo, pero sería un indicador inválido del nivel de uso de anticonceptivos, si los usuarios están usando los productos subvencionados para sustituir otros métodos.

Otra dimensión de la validez es que el indicador debe tener una conexión cercana con la intervención, o al menos esta conexión debe ser defendible. Por ejemplo, la tasa de mortalidad materna, en general no es una medida válida del impacto de un programa de planificación familiar sobre la salud femenina. Aunque los programas de planificación familiar ciertamente contribuyen a reducir la mortalidad materna, existen otros factores

(atención prenatal, sistema de referencia, acceso a la atención hospitalaria, transporte) que también influyen en la tasa. Un indicador con mayor validez del impacto de la planificación familiar en la salud de las mujeres puede ser el número de nacimientos de mujeres que pertenecen a categorías de alto riesgo.

La **confiabilidad** es otra característica deseable en un buen indicador. La confiabilidad se refiere al error aleatorio de medición en un indicador. El error de medición puede surgir de errores aleatorios propios del muestreo, un error no de muestreo o una medición subjetiva del indicador. Por ejemplo, debido al error de muestreo, una encuesta nacional como la DHS no provee estimaciones confiables de la práctica anticonceptiva en áreas pequeñas. Las estadísticas de servicio pueden dar medidas más confiables de la práctica anticonceptiva en estas áreas (aunque pueden no ser válidas si la población en el área obtiene anticonceptivos fuera de los servicios oficiales). Las encuestas por muestreo generalmente proveen estimaciones de aborto no confiables debido al sesgo de las respuestas, en este caso, la renuencia de las entrevistadas a reportar abortos. Además, los indicadores que se basan en juicios subjetivos, como por ejemplo, la calidad de liderazgo del programa, pueden no ser confiables, debido a que los diferentes evaluadores pueden usar distintas normas para medir esa característica.

Un indicador debe ser definido en **términos claros y precisos**. El indicador debe ser definido operacionalmente para que otras personas puedan saber precisamente lo que se está midiendo. Por ejemplo, ¿qué es un aceptante de planificación familiar? ¿Cómo se define proveedor competente? ¿Se basa el conocimiento de un método de planificación familiar en el conocimiento espontáneo o en conocimiento con ayuda? La definición del indicador debería también especificar la población en la que se va a medir el indicador: ¿todas las mujeres o mujeres casadas? ¿Mujeres de 15-44 o de 15 a 49 años? ¿Todos los usuarios de la planificación familiar o solo los usuarios de métodos modernos? ¿Todas las mujeres del país o solo las del área del proyecto? Finalmente, el significado de las medidas demográficas como tasas y razones debería estar claramente especificado.

De igual manera, es preferible escoger indicadores que sean comparables entre los distintos grupos poblacionales de la población y entre distintos enfoques de programa. En

igualdad de condiciones, una tasa de prevalencia anticonceptiva basada en mujeres de 15-49 años es preferible a una basada en mujeres de 15-44 años, porque será más comparable a tasas de otros programas. Un programa con diferentes enfoques de entrega de servicio (por ejemplo, clínico, CBD y mercadeo social) debería tratar de identificar al menos uno o dos indicadores de utilización del servicio que sean apropiados para las tres modalidades, como los APP, de manera que los resultados puedan ser comparados entre los distintos enfoques de entrega de servicios. En la medida de lo posible, los indicadores deberían reflejar el desempeño relativo a alguna norma o "denominador común". Por ejemplo, podría ser interesante conocer el número de APP para un método específico del sector comercial, pero tal vez es más útil conocer el "porcentaje de usuarios de un método específico obtenido del sector comercial" (para métodos de reabastecimiento).

Los indicadores deben ser insesgados. Describen la situación en un momento dado. Si, por ejemplo, un programa está tratando de valorar su progreso en reducir el desabastecimiento, el indicador apropiado es el "porcentaje de PPS que se encontraron desabastecidos en los últimos 12 meses" (que podría ser rastreado a través del tiempo), no "la disminución en el porcentaje de PPS que estuvieron desabastecidos."

Los indicadores deben ser recolectados a tiempo. El indicador debe proveer una medida para un período reciente o al menos, para el período durante el cual ocurre la intervención; también, debería estar disponible en intervalos adecuados.

Los indicadores poblacionales casi nunca están disponibles anualmente y suelen referirse a un período de varios años antes de la encuesta (por ejemplo, las estimaciones de fecundidad de la DHS generalmente se refieren a períodos de tres años anteriores a la encuesta). Aunque los datos de rutina del programa, como las estadísticas de servicio, podrían ser una buena fuente de datos actualizados, tiende a haber un atraso considerable en su disponibilidad. Finalmente, algunos datos poblacionales se recolectan ahora periódicamente mediante métodos como el Análisis de Situación o el Módulo de Disponibilidad del Servicio de la DHS. Estos instrumentos proveen medidas oportunas, pero no continuas, del funcionamiento del programa.

Factores que afectan la selección de los indicadores

En un mundo ideal, el evaluador identificaría sistemáticamente los indicadores considerados como los más útiles para una determinada evaluación y procedería a recolectar o adquirir los datos requeridos. Sin embargo, si en el mundo real el tiempo y los recursos humanos y financieros son limitados, otros factores intervienen en la selección de los indicadores. Los siguientes son factores comunes que intervienen en la decisión.

- Disponibilidad de datos requeridos para medir el indicador

Ejemplo: Para valorar los efectos de los programas de planificación familiar sobre los resultados en la fecundidad y en la salud a nivel mundial, sería extremadamente útil contar con datos de todos los países sobre las fuentes de financiamiento (agencias donantes, impuestos locales, tarifas para los clientes) de la planificación familiar y sobre los costos de proveer servicios de planificación familiar. Sin embargo, tales datos no existen actualmente de manera fácilmente accesible. Además, no está claro que todos los gobiernos estén dispuestos a abrir sus registros financieros a los evaluadores externos para efectos de recolectar esta información.

- Tiempo asignado a la evaluación

Ejemplo: Los administradores de programa podrían querer saber si su nuevo enfoque en la consejería sobre el uso del NOR-PLANT® a los clientes trae como resultado mayores tasas de continuación. Sin embargo, si la evaluación del programa de consejería se limita a la vida del proyecto (parte del cual posiblemente ha pasado), es imposible comprobar los efectos a largo plazo de la consejería.

- Apoyo financiero disponible para la evaluación

Ejemplo: Muchos directores de IEC quisieran conocer el porcentaje de la población objetivo que fue alcanzada por una determinada campaña y la reacción del público a los mensajes. Pero pueden no contar con los recursos para realizar una encuesta entre la población. Podría haber un conflicto entre el costo, por un lado y la validez, confiabilidad y oportunidad, por otro.

- Requisitos de las agencias donantes

Ejemplo: El indicador "años de protección de la pareja" se ha convertido en la medida más usada de la utilización del servicio en los programas financiados por la USAID, porque la USAID (y la IPPF) exige a las agencias receptoras de su asistencia reportar este resultado.

Uso de indicadores múltiples

Para los indicadores bien establecidos como la Tasa Global de Fecundidad (TGF) y la Tasa de Prevalencia Anticonceptiva (CPR), un solo indicador suele ser suficiente. Sin embargo, a veces es aconsejable usar dos o más indicadores para medir un determinado resultado. Una situación de este tipo es cuando la calidad de los datos es dudosa; un resultado goza de mayor credibilidad si la misma tendencia puede ser demostrada en dos o más indicadores.

Además, cuando se introducen nuevos indicadores de programa, es útil tener indicadores alternativos para una determinada categoría de resultados (por ejemplo, medir la calidad de la interacción cliente-proveedor). Esto ofrece al evaluador un plan alternativo en caso de que los datos de una fuente no se concreten o sean considerados inválidos (por ejemplo, los entrevistados no entendieron bien la pregunta). Finalmente, los indicadores de una cierta área funcional suelen medir una cadena de eventos y el uso de múltiples indicadores puede ser importante para desarrollar una comprensión de la dinámica a lo largo de la cadena. Por ejemplo, un programa IEC (1) genera cierto número de mensajes a través de cierto número de canales y (2) ofrece orientación a clientes potenciales y reales que buscan servicios esperando que los miembros de la población objetivo (3) escuchen los mensajes sobre planificación familiar, (4) entiendan los mensajes principales, (5) reaccionen positivamente a los mensajes, (6) discutan los mensajes con otros, (7) desarrollen una disposición favorable hacia el comportamiento promovido, como es el uso de anticonceptivos, (8) se conviertan en aceptantes, y (9) continúen la práctica. En cada paso de este proceso el porcentaje que sigue esta cadena puede potencialmente disminuir; para el evaluador es indispensable identificar el patrón de esta respuesta por parte de la población objetivo.

En suma, la selección de los indicadores se basa en el propósito de la evaluación: saber más sobre un área funcional, producto o resultado del programa y en algunos casos para satisfacer los requisitos de la agencia donante. La selección

de los indicadores está determinada por las necesidades e intereses específicos de quienes realizan la evaluación. Pueden conducirse distintos tipos de evaluación en forma escalonada durante los cinco o más años que dure el proyecto.

Operacionalización de los indicadores

“Operacionalización de los indicadores” significa identificar la forma como una cierta conducta o concepto puede ser medido. En el mejor escenario posible, un indicador está conceptualmente claro y se presta fácil e inequívocamente para la medición. Un ejemplo sería el número de personas entrenadas en un año determinado, por categoría de personal (médico, enfermera, especialista en suministros y logística y otros).

Desafortunadamente, muy pocos indicadores son tan simples y directos. Más bien, aún después de que el evaluador ha identificado los indicadores que deben usarse, tiende a enfrentar uno o más de los problemas siguientes al operacionalizarlos.

- La medición de un indicador requiere juicios subjetivos

Muchos están de acuerdo en que uno de los factores más importantes para los programas de planificación familiar exitosos actualmente es la “calidad del liderazgo del programa”. Para usar este indicador, es obligatorio definir las características que constituyen el “liderazgo”, pero la valoración final sigue siendo subjetiva.

De igual manera, los indicadores que requieren un juicio de “presencia” o “ausencia” pueden ser complicados si se debe establecer “cuánto” constituye la “presencia”. Por ejemplo, con relación al indicador “ausencia de restricciones arbitrarias para los usuarios”, uno podría tener un programa modelo con una pequeña excepción. ¿Debería bastar esta excepción para alterar la valoración que uno hace?

Si el evaluador se enfrenta con el problema de los juicios subjetivos, debe decidir primero si retiene el indicador. Si lo retiene, debe aclarar los criterios usados para llegar a la calificación o valoración final.

- Las reglas de medición están claras pero las aplicaciones finales difieren del enfoque recomendado

Por ejemplo, el Manual de Indicadores para

la Evaluación del Programa de Planificación Familiar (Bertrand et al., 1994) recomienda definir el “número de aceptantes nuevos para la institución” como “nuevos solamente una vez”. Vale decir que si una persona abandona el programa de la institución por varios años y luego vuelve, NO sería “nueva” para la institución. Pero, puesto que algunas organizaciones no guardan registros por más de cinco años, puede no ser factible adherirse a la definición recomendada. En tales casos, el evaluador debe tener muy claro en qué se distingue la medición utilizada de la práctica estándar o recomendada.

- El indicador está conceptualmente claro pero la balanza en que se lo mide no lo está.

En apariencia, el “costo del suministro mensual de anticonceptivos como porcentaje del salario mensual” luce claro y medible. Pero, cuando uno aplica el indicador, pueden surgir ciertas preguntas. Por ejemplo, los distintos métodos anticonceptivos tienen distintos costos. ¿Cuáles costos deberían estar incluidos: el costo promedio de todos los métodos disponibles? ¿El costo promedio ponderado según la proporción que usa los distintos métodos? Además, ¿cuáles números deberían usarse si el costo de los métodos varía durante el transcurso del año? ¿Qué número debería usarse en el denominador si las cifras de salarios mensuales promedio están desactualizadas?

Es útil revisar cómo otros investigadores y evaluadores han manejado situaciones similares, en la medida que los evaluadores tengan acceso a informes de otros. Si no se tiene esta información, lo correcto es: documentar cualquier decisión que se haga para operacionalizar los indicadores para una evaluación específica.

Para cada indicador que pueda ser utilizado en el seguimiento de programas, es necesario identificar una o más fuentes de datos de las cuales obtener información, como se describe en la sección siguiente.

IDENTIFICACIÓN DE FUENTES DE DATOS

La evaluación de un programa nacional de planificación familiar normalmente comprende datos de la población y del programa. Aunque existan muchas fuentes de datos (ver Gráfica III-3), la mayor parte de la evaluación de la pla-

nificación familiar se limita a unas pocas fuentes de datos. En el diseño de una evaluación, es esencial hacer un inventario de los datos que ya existen e identificar la recolección adicional de datos que es necesaria para proveer respuestas a las preguntas en estudio.

¿Qué está primero: identificar los indicadores que se van a usar o seleccionar las fuentes de datos para medirlos? Los dos procesos están estrechamente relacionados; la selección de los indicadores suele estar determinada por la disponibilidad de datos o la factibilidad de recolectar información adicional con un costo mínimo para el programa (por ejemplo, estadísticas rutinarias del servicio).

Sin embargo, algunos tipos de evaluación NO pueden ser llevados a cabo usando los datos existentes, tales como las valoraciones de la calidad de la atención que requieren los estudios especiales. En tales casos, uno hace un esquema de los indicadores de interés y luego identifica la recolección de datos requerida para obtener la información.

Aunque potencialmente existe un gran número de fuentes de datos para evaluar los programas de planificación familiar, la mayor parte de la evaluación de programas se basa en las siguientes fuentes de datos:

Componente a medir	Fuente(s) de datos
Productos (del programa)	Registros de programa, especialmente estadísticas del servicio Encuestas de establecimientos Datos sobre el sector comercial Estudios especiales
Resultados (en la población)	Encuestas de hogares tipo DHS

Datos a nivel del programa

Existen muchos tipos diferentes de información que constituyen los datos del programa. Estos pueden ser agrupados en cuatro categorías principales.

Registros del programa y estadísticas del servicio para proveedores del sector público y ONGs.

“Registros del programa” se refiere a todos los tipos de información generados por una o más divisiones del programa y guardados en archivo en una oficina central o regional. Algunos ejemplos útiles para el seguimiento de programas (especialmente, productos funcionales) son: el número de visitas de supervisión hechas a trabajadores de CBD, el número de personas entrenadas por año por tipo de personal, el número de los distintos productos de comunicación elaborados en un año determinado, la cantidad de productos de comunicación distribuidos y otros.

Las estadísticas del servicio son subcategorías de los datos del programa. Incluyen cualquier tipo de información recolectada de rutina y registrada en relación al uso de un servicio. Algunos de los indicadores más comunes para los programas de planificación familiar son

- el número de aceptantes nuevos;
- el número de visitas a las PPS;
- los años de protección de pareja; y
- características del usuario.

Generalmente, estos datos son recolectados en cada PPS, luego agregados en una oficina central para dar seguimiento a las tendencias a través del tiempo y por sub-unidad dentro del sistema.

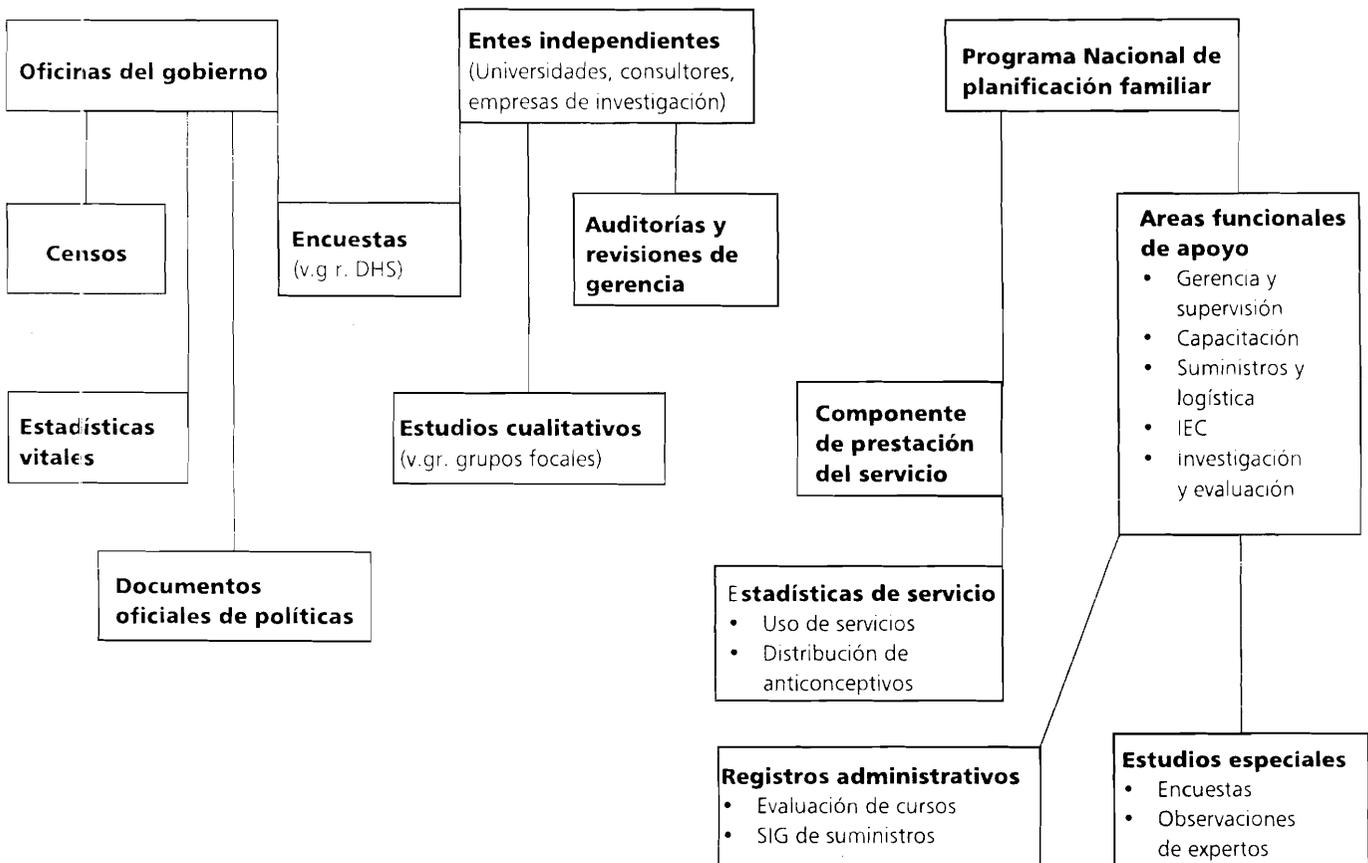
Encuestas de establecimientos

El objetivo primario de las encuestas de establecimientos es describir la disponibilidad, funcionamiento y calidad de las actividades de planificación familiar y de salud. Esta información puede obtenerse entrevistando personas informadas o visitando el establecimiento y observando sus operaciones. Los datos de establecimientos también son requeridos para estudios que vinculan la información del programa (por ejemplo, calidad de la atención) con, por ejemplo, los resultados en la población (por ejemplo, prevalencia anticonceptiva), ver Mensch et al. (1994).

Hay dos tipos principales de encuestas de establecimientos en proyectos de la USAID. La primera, conocida como “Análisis de Situación”,

Gráfica III-3

Fuentes y tipos de datos para la evaluación de un programa nacional



fue desarrollada en el contexto del proyecto de Asistencia Técnica e Investigación de Operaciones en África (OR/TA) (Fisher et al., 1992) y ha sido replicado en varios países alrededor del mundo. La segunda es el Módulo de Disponibilidad del Servicio (SAM), desarrollada e implementada bajo el programa DHS.

Los dos tipos de encuesta se diferencian entre sí por (1) los instrumentos de recolección de datos utilizados y (2) en el universo de PPS que describen. En el Análisis de Situación, los

instrumentos para la recolección de datos incluyen una serie de módulos (un inventario de suministros en el PPS, observación de la interacción proveedor-cliente, entrevistas de salida con clientes, entrevistas con proveedores del servicio, entre otros). Los resultados descriptivos ofrecen a los administradores centrales información importante sobre las ventajas y desventajas del programa a través del país (o área geográfica bajo estudio). El Análisis de Situación también se usa para obtener datos sobre la calidad de la atención en los

programas de planificación familiar.

EL SAM es una encuesta a comunidades, cuya conducción está asociada a una encuesta de hogares en la DHS (pocas DHS incluyen un SAM). Para cada conglomerado muestral utilizado en el estudio, los informantes clave proveen una lista de establecimientos existentes dedicados a la planificación familiar y la salud. Los equipos de campo obtienen luego la información en cada (1) hospital, (2) clínica, (3) centro de salud, (4) farmacia y (5) doctor privado más cercanos, dentro de un radio de 30 kilómetros del centro del conglomerado⁸. La información recolectada en cada sitio cubre la afiliación institucional (privada o pública), el número y tipo de personal, la infraestructura (equipo, tipo de materiales de construcción), tipos de servicios proveídos, tipos de métodos anticonceptivos disponibles, etc. Esta información es potencialmente útil para valorar la disponibilidad y suficiencia de los servicios de planificación familiar para una población determinada y puede ser particularmente importante para vincular los cambios (mejoras) en el entorno de la oferta de la planificación familiar con los cambios en la prevalencia a través del tiempo.

La diferencia en el muestreo utilizado en los dos tipos de encuestas es la siguiente: el Análisis de Situación se basa en una muestra aleatoria de PPS en un país (los que pueden estar desproporcionadamente localizados en áreas urbanas), mientras los datos de los SAM son recolectados para una muestra aleatoria de mujeres en el país⁹. Entonces, el Análisis de Situación mide los PPS promedio, mientras la SAM mide los servicios disponibles para las mujeres promedio en un país determinado. En realidad, las encuestas de establecimientos han estado bastante subutilizadas para los propósitos de evaluar los programas de planificación familiar. Su utilidad potencial en la valoración del impacto (en relación a las entrevistas individuales) se discute en el Capítulo IV.

Los estudios de costos también pueden conducirse en los establecimientos. Estos proveen información sobre el costo de ofrecer diferentes servicios incluyendo visitas a los aceptantes y visitas de seguimiento por método. Usando la información sobre las visitas y el número de APP por método, estos datos pueden ser agregados para determinar el costo por APP. Además, el costo de expandir las visitas a aceptantes y de seguimiento,

así como el número de APP pueden ser calculados considerando el volumen de capacidad subutilizada.

Información sobre la provisión de planificación familiar en el sector comercial

Las fuentes de datos sobre el sector comercial generalmente no forman parte de las estadísticas del programa, las cuales provienen del sector público y los grandes proveedores no gubernamentales (ONG). No hay una central de estadísticas del servicio de planificación familiar en el sector comercial. Esto se debe, en parte, a la ausencia de un cuerpo central encargado de estos datos y en parte, porque los proveedores comerciales están compitiendo entre ellos y pueden estar reacios a compartir información sobre el volumen y calidad de sus servicios.

La información sobre las actividades de planificación familiar en el sector comercial privado han sido medidas en algunas SAM y Estudios de Análisis de Situación, así como en encuestas centradas exclusivamente en los proveedores privados. Además, algunas fuentes de datos exclusivas de este sector están disponibles para medir la disponibilidad de métodos y servicios en el sector. Estas incluyen:

- datos sobre cargamentos de anticonceptivos para distribuidores y vendedores;
- ventas de los minoristas;
- inspecciones de locales de ventas minoristas; y
- reportes de detallistas sobre la disponibilidad de servicios y métodos de planificación familiar provenientes de doctores y clínicas.

Los proveedores en el sector comercial y los patrocinadores de los programas de mercadeo social y del sector privado también están interesados en la calidad de los servicios proveídos. Algunas técnicas utilizadas para valorar la calidad de los servicios son: (1) estudios de clientes incógnitos para determinar si los minoristas están pro-

⁸ Es preferible que el equipo de recolección de datos efectivamente visite cada sitio; sin embargo, en las primeras DHS o cuando los fondos no estaban disponibles, informantes clave proveían esta información en algunos de los estudios.

⁹ Sin embargo, con una ponderación apropiada, los resultados del SAM pueden ser presentados para una población equivalente a la descrita normalmente en el Estudio de Análisis de Situación.

moviendo los productos de mercadeo social y para valorar la calidad de la información ofrecida; (2) entrevistas a la salida hechas a los consumidores para valorar su satisfacción con los servicios recibidos; y (3) encuestas a la población.

Estudios especiales

Los estudios especiales generalmente son conducidos para responder a una necesidad específica. Pueden utilizar métodos de investigación cuantitativos o cualitativos. La lista de posibles estudios especiales es larga, algunos ejemplos ilustrativos son los siguientes:

- un seguimiento de clientes de esterilización para determinar su nivel de satisfacción con el procedimiento;
- grupos focales entre adolescentes que asisten a un determinado programa para valorar si éste responde a sus intereses y necesidades;
- una revisión administrativa de los documentos del programa; y
- mapeo de una comunidad para mostrar dónde viven las parejas elegibles y cuál método anticonceptivo usan.

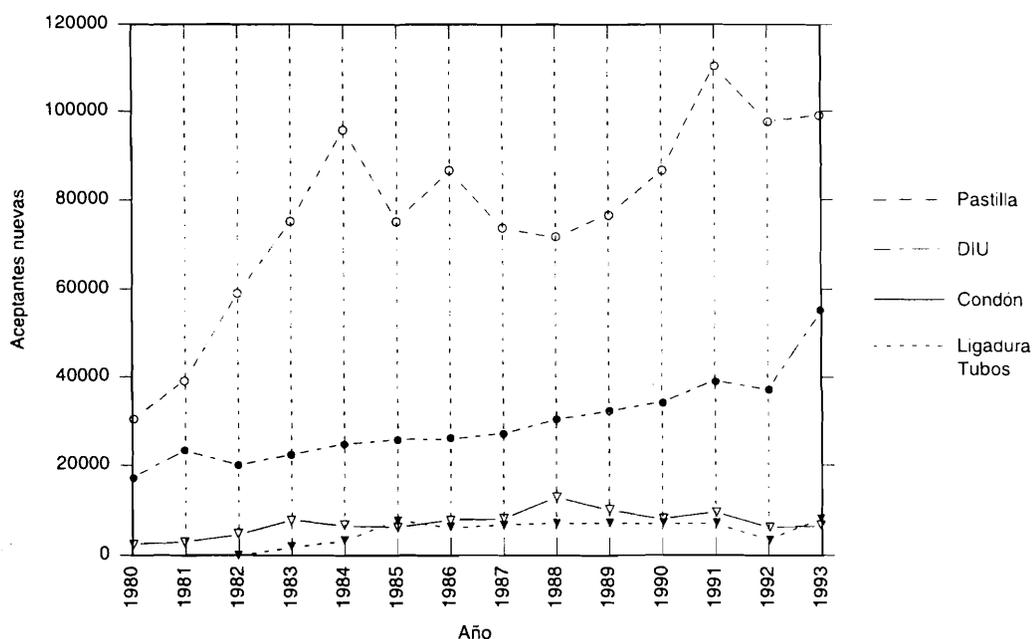
Datos de la población

El instrumento primario para recolectar datos de la población para programas de planificación familiar es la encuesta tipo DHS. Siguiendo la tradición de la Encuesta Mundial de Fecundidad (WFS) y la Encuesta de Prevalencia Anticonceptiva (CPS), la Encuesta Demográfica y de Salud (DHS) generalmente se realiza en una muestra representativa de mujeres en edad reproductiva de un país. Recientemente, cierto número de encuestas DHS también han incluido una muestra de hombres (ya sea una muestra independiente de hombres o una muestra de esposos de las mujeres entrevistadas en la DHS). El principal cuestionario de la DHS que comprende unas 250 preguntas ofrece información detallada sobre fecundidad y planificación familiar, además de información sobre salud materno-infantil, utilización de servicios de salud y temas afines (Robey et al., 1992).

Usamos el término "encuestas tipo DHS" para subrayar que hay otras encuestas similares a la DHS en contenido y tipo de población estudiada. Los Centros de Control y Prevención de Enfermedades (CDC) han conducido ciertas encuestas de salud reproductiva en algunos países de América Latina y otras regiones del mundo. De igual manera, algunos países han realizado sus propias encuestas nacionales sobre fecundidad y temas afines.

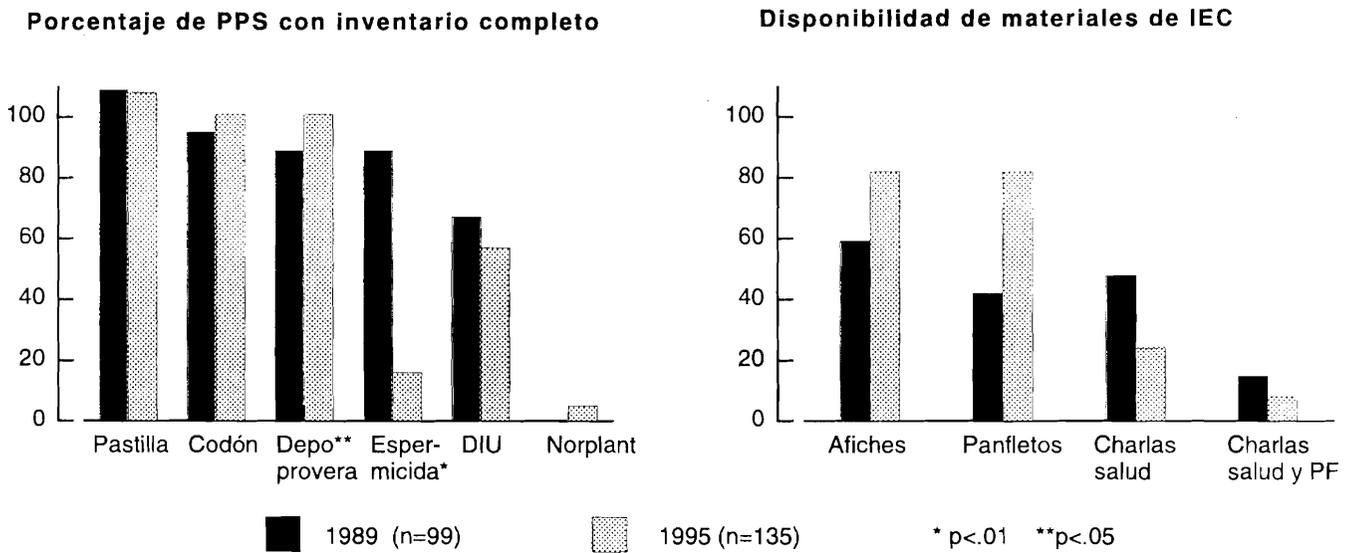
Gráfica III-4

Nuevas Aceptantes por Método y Año: 1980-1993



Gráfica III-5

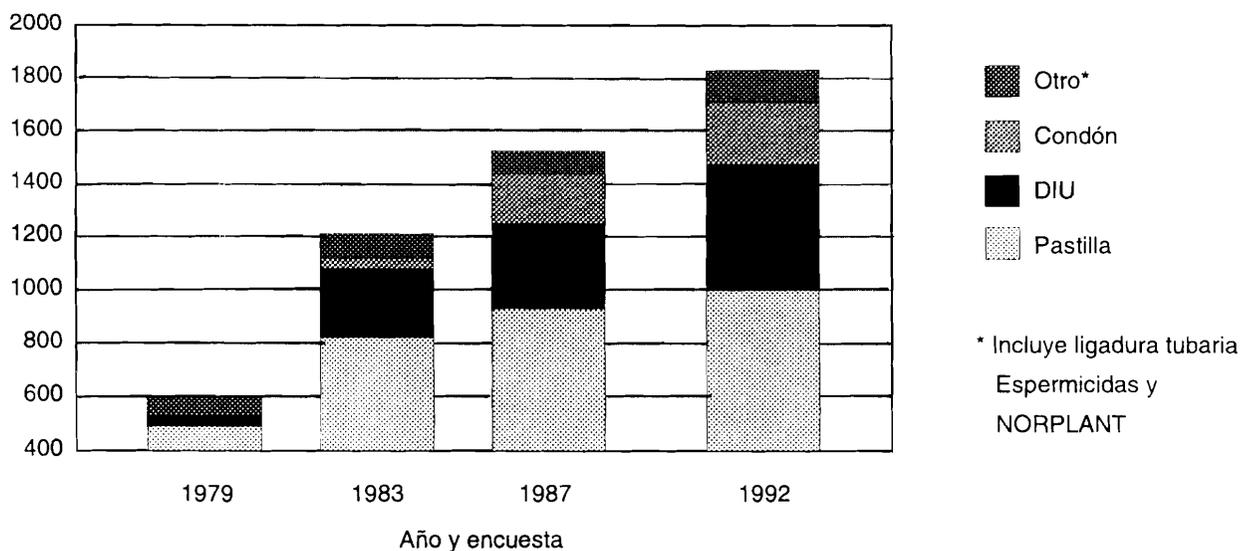
Cambios en los productos del servicio según dos análisis de situación. Kenia 1989, 1995



Fuente: Miller, et al., 1996

Gráfica III-6

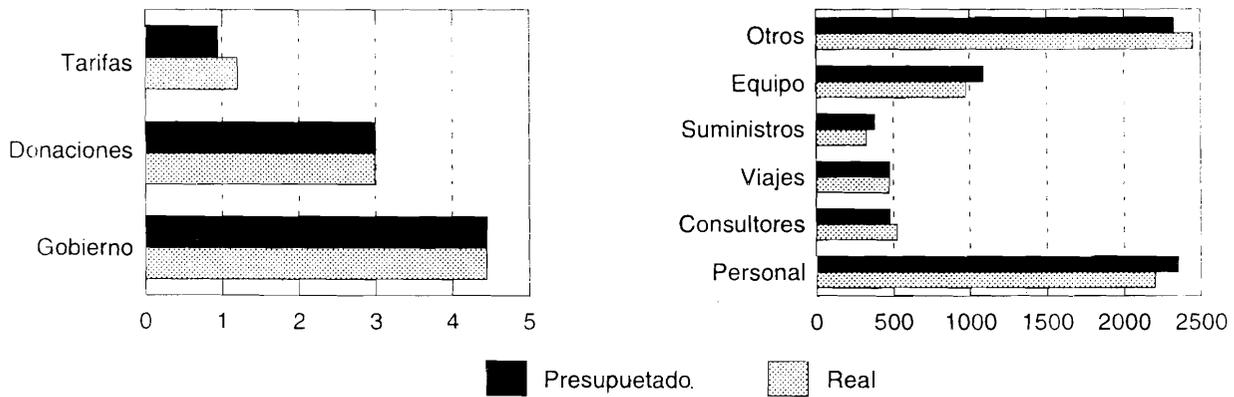
Composición de los Métodos según Encuestas Nacionales



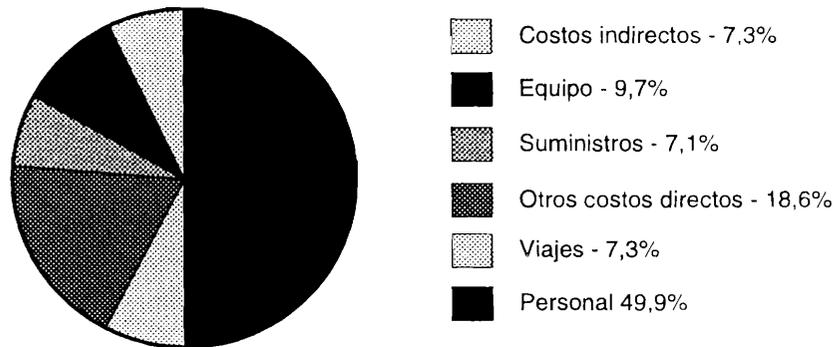
Grafica III-7

Ejemplo: Información de Costos Basada en los Registros del Programa

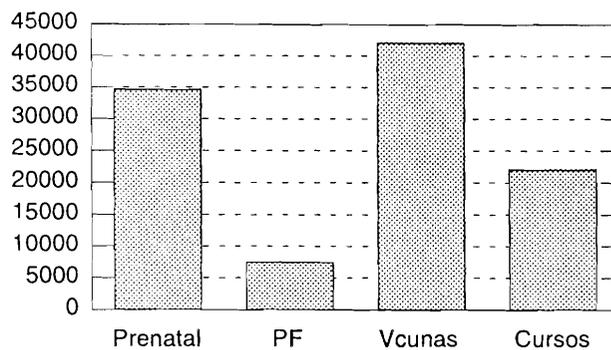
Ingresos y gastos presupuestados vs. reales



Distribución de los gastos del programa



Distribución de los gastos por servicio/actividad



Fuente: Reynolds, J. 1993: 9

DISEÑO DE UN FORMATO PARA LA PRESENTACIÓN DE RESULTADOS

Frecuentemente, los evaluadores y otros investigadores recolectan mucha más información que la necesaria. Una forma de evitar esto es especificar cómo la información va a ser procesada y presentada con bastante anticipación a la recolección de datos. Este enfoque permite al evaluador ver cómo será el informe y valorar si el grado de detalle se justifica. Por otra parte, este paso puede también mostrar las carencias de información clave que debe ser incluida en la recolección de datos.

Los investigadores suelen referirse a este paso como "el diseño de cuadros simulados". Consiste en producir una serie de cuadros y gráficas que indican las variables exactas que

deben presentarse en cada una y el tipo de dato que debe incluirse (números absolutos, porcentajes, promedios, medianas, etc.), pero no los datos reales (los cuales no están todavía disponibles o procesados).

Los datos pueden presentarse en forma tabular o gráfica. Una ventaja de presentar los resultados en cuadros es que uno puede incluir valores precisos para un número relativamente grande de indicadores. Sin embargo, el "mar de números" puede desanimar a algunos lectores a la hora de examinar las tendencias. En cambio, las gráficas tienden a destacar las tendencias en los datos (contribuyendo a su comprensión), pero pueden impedir la presentación de valores específicos. Las técnicas básicas para presentar resultados son similares, independientemente de la fuente de datos. Las gráficas III-4 a III-7 son ejemplos

Gráfica III-8

Ejemplo: Visión de conjunto de la metodología para el seguimiento de un programa concreto

Componentes a seguir	Fuente de datos	Indicadores	Frecuencia de informes	Organización responsable
Acceso a servicios	Expedientes del programa	Número de PPS que ofrecen PF en el área	Trimestral	MS y asociación de PF
Calidad de la atención	Encuesta en los establecimientos (v.g. análisis de situación)	Ejemplo; <ul style="list-style-type: none"> • Núm. de métodos en inventario • Proveedor da información relevante y veraz 	Ocasional	Agencia especializada en esta clase de estudios
Utilización de los servicios	Estadísticas de servicio	Número de aceptantes, APP	Trimestral	MS y asociación de PF
Costos	Registros financieros y otros documentos del programa	Costo por aceptante, visita de control y de discontinuación, según método	Estudio único	Agencia especializada

Enfoque metodológico: Seguimiento de programas

de fuentes comunes de datos para indicadores típicos que uno puede rastrear:

estadísticas del servicio recolectadas de rutina: el número de usuarios nuevos por tipo de método (Gráfica III-4);

- datos de establecimientos (Análisis de Situación): porcentaje de PPS con inventario de suministros a mano, disponibilidad de materiales de IEC (Gráfica III-5);
- medidas de resultados en la población (DHS): cambios en la composición de métodos anticonceptivos a través del tiempo (Gráfica III-6); y
- registros financieros del programa: gastos en comparación con cantidades presupuestadas, gastos por tipo y por actividad del servicio (Gráfica III-7).

En el diseño de cuadros y gráficas simulados, el evaluador debe incluir toda la información que pretende presentar. Con este ejercicio, uno puede valorar rápidamente si los resultados de la evaluación serán suficientemente exhaustivos sin llegar a abrumar al usuario potencial.

RESUMEN DEL ENFOQUE METODOLÓGICO

No hay una regla fija sobre la manera en que los temas tratados en este capítulo deban ser resumidos en el plan de evaluación. Pero, es importante que el documento presente una visión de conjunto de los elementos cubiertos aquí:

- propósito(s) principal(es) del seguimiento;
- componentes que deben recibir seguimiento (basado en el "mapa" del marco conceptual);
- indicadores para cada ejercicio de recolección;
- fuente(s) de datos para cada ejercicio de recolección (indicando cuáles existen ya y cuáles deben ser recolectados); y
- formato para la presentación de los resultados.

Capítulo IV

Enfoque metodológico: Valoración del impacto

- Una advertencia amigable
- Visión de conjunto de los temas
- Criterios para la selección del enfoque metodológico
- Enfoques preferidos
- Enfoques alternativos
- Resumen

PREVIOUS PAGE BLANK

ENFOQUE METODOLÓGICO: VALORACIÓN DEL IMPACTO

Los capítulos I, II y III de este documento tratan temas y conceptos que tienden a ser conocidos, al menos parcialmente, para la mayoría de los lectores. El material fue presentado paso a paso para aclarar aún más el proceso de evaluación de programas y lograr que las herramientas del seguimiento de programas sea más fácilmente comprensible para todos los lectores, independientemente de su experiencia en estadística e investigación.

En contraste, el capítulo IV considera temas que son metodológicamente más complejos. El objetivo del capítulo es promover una comprensión general de (1) las ventajas y desventajas de los varios enfoques y métodos para medir el impacto de programas, (2) las razones por las que el Proyecto EVALUATION recurre a algunos métodos y no a otros y (3) los requisitos generales para usar estos métodos. Los autores han intentado presentar el material en un lenguaje que comunique las ideas metodológicas básicas a lectores con escasa experiencia en estadística e investigación, al tiempo que ofrece suficiente información técnica a los lectores más versados, de manera que las ventajas técnicas relativas de los enfoques alternativos puedan ser apreciadas. Sin embargo, en este capítulo no se hace una presentación paso a paso como la de los primeros tres capítulos y se excluyen los detalles de cálculo. Las referencias a otras publicaciones cuentan con los detalles técnicos necesarios para la aplicación de los métodos discutidos en este capítulo.

VISIÓN DE CONJUNTO DE LOS TEMAS

Cada vez más, la atención se centra en el efecto final de los esfuerzos de los programas de planificación familiar y de salud; o sea, en el impacto de los programas y las intervenciones. Como se señaló en el capítulo II, la medición del impacto comprende algo más que el simple seguimiento de los cambios en los indica-

dores de resultados; requiere una evidencia empírica de que un cambio observado en los indicadores de resultados es atribuible a un determinado programa o intervención. En otras palabras, requiere la evidencia de que el programa o la intervención evaluada ha causado el cambio.

Para medir el impacto de programas están disponibles varios métodos o enfoques distintos. Estos difieren, por ejemplo, en términos de los tipos de medidas de resultados utilizadas, el número y tipos de supuestos requeridos, la fuerza de las conclusiones respecto al impacto de programa alcanzado, los requerimientos operativos y los requerimientos de datos. Este capítulo ofrece una visión de conjunto de lo que el Proyecto EVALUATION considera como los métodos más robustos disponibles para medir el impacto de programas de planificación familiar. Nuestro objetivo es ofrecer una guía para facilitar la selección entre las opciones disponibles.

El material de este capítulo tiene en mente al funcionario del programa (o el investigador) a cargo de la evaluación, quien tuvo la responsabilidad de preparar el plan de evaluación para un nuevo programa o ciclo de programas. Las decisiones sobre cómo debe medirse el impacto del programa es mejor tomarlas en la etapa de diseño o de planeamiento de los programas o nuevos ciclos de programas, porque solo en esta etapa se tiene disponible una amplia gama de opciones. Una vez que un programa o ciclo de programas está en proceso, la oportunidad para usar algunos de los diseños y métodos de investigación más fuertes se pierde y se hace necesario recurrir a opciones más débiles.

El contraste con el seguimiento de tendencias en las estadísticas de servicio o la realización de estudios descriptivos de la marcha del programa (por ejemplo, encuestas de satisfacción de clientes), la valoración del impacto, en el sentido estricto de la palabra, requiere métodos estadísticos y destrezas analíticas que carecen muchos administradores de progra-

ma, personal de agencias donantes y especialistas en evaluación de países anfitriones. En este capítulo, presentamos una descripción de los distintos métodos destinada a lectores con escasos conocimientos estadísticos, pero que les permite juzgar las ventajas de usar una técnica en su propio escenario (aunque el análisis propiamente sea realizado por otras personas). Después de estas descripciones generales se ofrece una discusión de temas metodológicos clave y un resumen de las ventajas y desventajas, destinado a proveer a estadísticos y demógrafos con más información sobre las implicaciones metodológicas de las distintas selecciones de métodos.

CRITERIOS PARA GUIAR LA SELECCIÓN DEL ENFOQUE METODOLÓGICO

Puesto que el propósito principal de este capítulo es ayudar a los usuarios a seleccionar entre los distintos métodos y enfoques alternativos para medir el impacto de programas, es útil empezar por especificar algunos criterios para hacer esta selección. Para valorar los distintos métodos descritos en este capítulo, el evaluador debe considerar los criterios siguientes:

- Posibles amenazas a la validez¹⁰

El criterio más importante para valorar un método es la validez de sus estimaciones del impacto del programa. Aunque todos los métodos son vulnerables a fallas en la validez, éstas varían considerablemente en términos del número y tipo de falla (por ejemplo, riesgos de confusión) a las cuales están sujetos.

- Supuestos requeridos

Todos los métodos requieren supuestos. Algunos métodos requieren supuestos extremos que rara vez son válidos en la realidad, mientras otros requieren supuestos más débiles y con más probabilidad de ser válidos. En igualdad de condiciones, se prefieren los métodos que requieren menos supuestos o supuestos menos rigurosos.

- Habilidad para aislar los efectos del programa

Idealmente, las medidas del impacto del programa incluyen solo resultados directamente atribuibles a éste. En la mayoría de situaciones, factores como el desarrollo económico, múltiples programas sociales, una estructura demográfica cambiante y la

presencia de actividades de planificación familiar fuera del programa complican los intentos de medir el impacto de éste. Los métodos difieren en la medida en que el evaluador puede aislar los efectos del programa de las influencias de otros factores. Los métodos preferibles son aquellos más eficientes para aislar estos efectos.

- Costos

Aquí se incluyen los costos de la recolección y análisis de datos. En igualdad de condiciones, se prefieren los métodos de menor costo. En muchos casos, sin embargo, los métodos difieren en otros aspectos además del costo, por lo que deben tomarse decisiones sobre el costo-beneficio.

- Datos requeridos

Los métodos varían considerablemente en cuanto a los datos requeridos. Aparte de las diferencias en el volumen de datos requeridos, algunos métodos necesitan datos que son muy difíciles de recolectar o son más vulnerables a errores de medición que otros métodos. Esto puede aumentar el riesgo de que el error de medición oscurezca los efectos de los programas o exagere la magnitud del impacto logrado realmente.

- Indicios de las trayectorias causales

Los métodos varían mucho en relación a la información que proporcionan sobre la forma como los insumos son transformados a productos y resultados. Aunque tal información no se requiere para la medición del impacto, provee útiles indicios de cómo mejorar los programas en ciclos de programa subsiguientes.

- Tipos de indicadores de resultados utilizados

Algunos métodos han sido diseñados específicamente para la medición de ciertos tipos de resultados. Por ejemplo, algunos métodos que han sido desarrollados específicamente para medir el impacto en la fecundidad de los programas de planificación familiar no son fácilmente adaptables a la medición del impacto en otros aspectos

¹⁰ El término "validez" como se usa aquí significa el grado en el que las medidas de impacto de programa de un determinado diseño constituyen medidas insesgadas y libres de distorsión del verdadero impacto de programa. En términos legos la validez se refiere al hecho de que uno esté realmente midiendo el fenómeno

(por ejemplo, resultados de salud). Otros métodos son más versátiles y pueden usarse para medir diversos tipos de resultados de programas, así como resultados en diferentes niveles. Por lo tanto, el tipo de resultado de interés en una evaluación particular determina parcialmente la selección del método.

- Grado de control del programa que se requiere

Ciertos diseños, como el experimental aleatorizado y, en menor grado, el cuasi-experimental, al mismo tiempo que proveen la evidencia más sólida de impacto del programa, requieren condiciones altamente controladas en la forma en que se evalúa el programa y si otras intervenciones se llevan a cabo simultáneamente. Los estudios "no experimentales", en cambio, no requieren que los programas se ejecuten de cierta manera para proveer mediciones válidas de su impacto, pero generalmente requieren mayor cantidad de datos y análisis más complejos para producir resultados válidos. El grado en que es posible ejercer control sobre la forma de ejecución del programa para facilitar la medición del impacto es, por ende, un factor importante en la elección del diseño de estudio.

- Recursos y destrezas técnico-estadísticas requeridos

Aunque todos los métodos considerados requieren de conocimientos y destrezas básicas en las áreas de investigación y estadística, algunos métodos y enfoques requieren destrezas relativamente avanzadas y en algunos casos software especializado.

Son, por tanto, numerosos los factores a considerar en la elección de un enfoque determinado para medir el impacto del programa. Para facilitar la selección del método apropiado, hemos clasificado los enfoques de valoración del impacto en dos categorías, que reflejan lo que los autores percibimos como su fortaleza global, sobre la base de los criterios siguientes:

- Enfoques preferidos

Los métodos que pertenecen a esta categoría son aquellos considerados como los diseños más robustos (por las razones que se indican a continuación) y son recomendados como primera elección siempre que sea posible.

- Enfoques alternativos

En los casos en que el uso de los métodos preferidos no es posible, están disponibles varios enfoques alternativos. Estos métodos se basan en diseños menos rigurosos y generalmente producen resultados menos convincentes que los métodos preferidos, pero bajo ciertas circunstancias son capaces de producir estimaciones apropiadas del impacto de programas.

Los métodos específicos considerados en este capítulo y su clasificación en los dos grupos definidos anteriormente se muestran en la Gráfica IV-1.

Obsérvese que los métodos presentados en la Gráfica IV-1 excluyen ciertos métodos que

Gráfica IV-1

Clasificación de los métodos más importantes para valoración del impacto según su validez general

Métodos preferidos

- Experimentos aleatorizados
- Cuasi-experimentos
- Métodos de regresión multinivel

Métodos alternativos

- Descomposición (modelo de determinantes próximos)
 - Método de prevalencia
-

aparecen en la literatura (United Nations, 1979, 1982, 1985; Chandrasekaran y Hermlin, 1985; Lloyd y Ross, 1989; Buckner et al., 1995). Este manual no los considera por varias razones. Algunos de estos métodos (por ejemplo, los años de protección de pareja estándar - APPS, el análisis del proceso reproductivo y los métodos de proyección por componentes) se basan en datos a nivel de establecimientos y están, por lo tanto, limitados en su capacidad para medir el impacto de programas en la población¹¹. Los APPS y los métodos de análisis del proceso reproductivo requieren datos que están disponibles solo ocasionalmente para todo el país. Otros mé-

¹¹ No se debe confundir APPS con los años-pareja de protección convencionales (APP), que es ampliamente usado para rastrear los insumos de programas.

descomposición genérica y el análisis de proyección y tendencia de la fecundidad) son métodos relativamente crudos que a menudo conducen a conclusiones ambiguas respecto al impacto del programa¹². Finalmente, la simulación se visualiza como un método genérico para el análisis que tiene sentido en el planeamiento de programas y en el planeamiento estratégico y como complemento de métodos más robustos, pero no es muy útil como método a usar en solitario para evaluar impacto.

De este modo, la atención de este manual se centra en los métodos y enfoques que el Proyecto EVALUATION considera que tienen las mayores posibilidades de producir medidas relativamente "limpias" de impacto de programas.

ENFOQUES PREFERIDOS

Experimentos aleatorizados

Descripción

Los investigadores dedicados a la evaluación consideran que el experimento aleatorizado o "verdadero" es el "estándar de oro" para medir lo que ha ocurrido como resultado de un pro-

grama o intervención. La idea fundamental de un experimento aleatorizado es muy simple. En un experimento aleatorizado, los sujetos o grupos de estudio son asignados al azar a los grupos de "tratamiento" o de "control", o sea, una vez que los sujetos o grupos de sujetos que van a ser estudiados son seleccionados, algunos son asignados al grupo de tratamiento y otros al grupo control (o de comparación) en forma aleatoria. Con una muestra suficientemente grande, la aleatoriedad aumenta las posibilidades de aislar sin ambigüedad los efectos de un programa o intervención, distribuyendo los factores extraños por igual a todos los grupos de comparación. Esto equivale a asegurar que los grupos de tratamiento y de control son equivalentes respecto a todos los factores, excepto la exposición al programa evaluado. Este método tiene grandes ventajas sobre otros métodos de medición de impacto del programa. Por esta razón, Rossi y Freeman (1993) se refieren al experimento aleatorizado como "la nave insignia" de la evaluación.

Diseño y Análisis

Dos de los experimentos aleatorizados más usados aparecen en la Gráfica IV-2. En el primero, el "diseño de grupo control de posttest solamente", se supone que la aleatoriedad ha producido grupos de tratamiento y de control equivalentes, y que entonces solo se necesita comparar las medidas de los resultados para los grupos de tratamiento y de control después de que el programa ha estado operando por un período de tiempo suficientemente largo para valorar el impacto del programa o

Gráfica IV-2

Diagrama de diseño de 2 experimentos aleatorios comunes

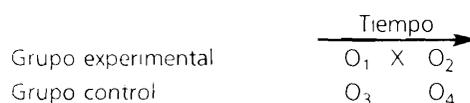
Posttest-solamente con grupo control

Asignación aleatoria:



Pretest-posttest con grupo control

Asignación aleatoria:



Donde

X = El programa o intervención

O = observaciones o mediciones

Fuente: Campbell and Stanley, 1963

¹² Debe observarse, sin embargo, que algunos de los métodos excluidos pueden ser útiles cuando se usan junto con los métodos más robustos. La estandarización, por ejemplo, se usa frecuentemente como primer paso en una evaluación de impacto para determinar la porción de cambio en la fecundidad que es atribuible a los cambios en la estructura demográfica, puesto que esta porción de cambio en la fecundidad claramente no puede ser atribuible a las intervenciones del programa de planificación familiar. De igual manera, las técnicas de simulación pueden usarse para complementar la información obtenida de ciertos métodos (ver, por ejemplo, la discusión de los métodos de regresión multinivel en la sección IV) indicando la magnitud del cambio en las variables de resultado que pueden ser esperadas de los cambios especificados en los insumos o productos de programas.

intervención que está siendo evaluada. En el segundo diseño, el "diseño de grupo control de pretest y postest", las mediciones se toman para los grupos de tratamiento y control antes de la implementación del programa y de nuevo después del tiempo considerado como suficiente para que el programa haya tenido el impacto buscado. La medición "antes y después" permite al investigador hacer ajustes para corregir la posibilidad de que la aleatoriedad no haya producido grupos completamente equivalentes¹³.

En ambos diseños, las medidas para el grupo control ofrecen una estimación de lo que se habría observado en el grupo experimental si el programa bajo estudio no se hubiera implementado. En el diseño postest-solamente, una estimación del impacto de programa lo da la diferencia entre los resultados del grupo experimental y el grupo control, más (o menos) un componente de error que es considerado como parte del análisis estadístico, o sea:

Impacto = $(O_1 - O_2) \pm \text{error}$
 donde: O_1 = medidas de resultado del grupo control; y
 O_2 = medida de resultado del grupo control; y
 error = errores de diseño y de medición¹⁴.

En el diseño pretest-postest, el impacto de programa se mide por la diferencia entre el cambio observado en las medidas de resultados para el grupo experimental menos las del grupo control, más/menos error:

Impacto = $(O_2 - O_1) - (O_4 - O_3) \pm \text{error}$
 donde: O_1 y O_2 = medidas de resultado pretest y postest, respectivamente, para el grupo experimental;
 O_3 y O_4 = medidas de resultado pretest y postest para el grupo "control"; y
 error = errores de diseño y de medición.

¹³ Esta es una preocupación particular en los estudios pequeños. En la medida que el tamaño muestral para el estudio aumenta, la probabilidad de que la aleatoriedad produzca grupos experimentales equivalentes también aumenta.

Gráfica IV-3

Ejemplo de un experimento aleatorio para la valoración del impacto de un programa, Taiwan

El experimento Taichung fue diseñado para valorar el impacto de una intervención para aumentar el conocimiento y uso de anticonceptivos en la ciudad de Taichung, Taiwán, a principios de los sesenta. Áreas locales o "lins" fueron asignadas aleatoriamente (después de una estratificación geográfica según densidad) a uno de cuatro grupos experimentales: (1) Paquete completo, esposo y esposa: los hogares de este grupo recibían visitas de trabajadores en salud, correo con información y reuniones vecinales; (2) Paquete completo, sólo esposa: la misma intervención del primer grupo, excluyendo la visita al esposo en el hogar; (3) solamente correo y (4) ninguna intervención fuera de afiches sobre planificación familiar distribuidos por la ciudad (o sea, grupo control). Las lins fueron asignadas a los grupos experimentales de la siguiente forma: (1) n=427, (2) n=427, (3) n=768, y (4) n=767. Los niveles de anticoncepción pre-intervención se supusieron iguales en los grupos aleatorios y, por eso, se usó el diseño de grupo control post-intervención solamente.

Las tasas de aceptación de anticonceptivos (por 100 mujeres casadas de 20-39 años de edad) post-intervención para los cuatro grupos experimentales y para periodos seleccionados fueron las siguientes:

Grupo experimental	Aceptan anticoncepción	
	13+ meses	29+ meses
1. Paquete completo: esposo y esposa	17	25
2. Paquete completo: sólo esposa	17	26
3. Sólo correo	8	16
4. Grupo control	8	18
Total	11	20

Interpretación: 29 meses luego de la intervención, un aumento en la tasa de aceptación de anticonceptivos del 7% puede atribuirse a la intervención "paquete completo, esposo y esposa" (i.e., tasa de este grupo experimental, 25%, menos tasa del grupo control, 18%). Incluir a los esposos en las visitas a los hogares no tuvo efecto sobre las tasas de aceptación de anticonceptivos, como se puede inferir de la similitud en las tasas de aceptación de anticonceptivos de los grupos experimentales "esposo y esposa" y "sólo esposa". Aparentemente, tampoco lo tuvo la información por correo.

Fuente: Freedman y Takeshita, 1969.

¹⁴ El componente de error consiste tanto de efectos de diseño como de errores estocásticos o aleatorios. Los efectos de diseño se refieren a los sesgos introducidos por factores en los que los grupos experimentales difieren a pesar de la aleatoriedad (se supone que son menos importantes en un experimento aleatorio, con un tamaño muestral suficiente). Los errores estocásticos se supone que son equivalentes entre los grupos experimentales.

El estudio experimental Taichung descrito en la Gráfica IV-3 ilustra cómo el experimento aleatorio puede usarse en una situación real.

Varios puntos adicionales sobre los experimentos aleatorizados deben ser mencionados. Primero, debe notarse que mientras la asignación aleatoria de sujetos de estudio individuales es muy común en experimentos clínicos y estudios más pequeños que involucran componentes de programa individuales (por ejemplo, los efectos de una mejor orientación en la continuación anticonceptiva), la asignación individual es más difícil y generalmente no es factible en estudios grandes a nivel de población. En estos estudios, la aleatorización se lleva a cabo por grupos, pueblos, municipalidades, distritos, etc. La asignación aleatoria de grupos de sujetos de estudio en los grupos experimentales se ilustra en la Gráfica IV-3¹⁵.

Segundo, es posible evaluar simultáneamente varias actividades o intervenciones en un experimento aleatorizado mediante la inclusión de varios grupos experimentales en el diseño, uno para cada tipo o variedad de "tratamiento". Este diseño "factorial" también se ilustra en el ejemplo de la Gráfica IV-3.

Tercero, no es necesario privar al grupo control de todo tratamiento o beneficio para usar los experimentos aleatorizados. Esto es importante, puesto que en los estudios poblacionales de impacto de programas de planificación familiar, es realmente difícil encontrar una población sin acceso a algún servicio de planificación familiar (o sea, un grupo control "puro"). Lo que se va a medir en tales casos, sin embargo, es el impacto diferencial o incremental del programa; o sea, la diferencia entre la actividad identificable que constituye la intervención o el "programa" y otros programas que puedan estar operando simultáneamente. Aunque esto puede parecer como un escenario poco deseable para algunos, el hecho de que un gran porcentaje de la población nacional de los países en desarrollo tengan acceso a algunos servicios de planificación familiar significa que el impacto de los nuevos programas será incremental respecto a aquellos programas o servicios que ya existen. En este sentido, el impacto incremental es una medida apropiada de lo que un programa ha logrado.

Finalmente, debe observarse que los experimentos aleatorizados son diseños de investigación genéricos. Caso contrario son los métodos desarrollados específicamente para medir un tipo de resultado particular de progra-

mas de planificación familiar (por ejemplo, los métodos diseñados específicamente para medir el impacto en la fecundidad). Por lo tanto, pueden ser aplicados para valorar los resultados de programas en los distintos niveles; a saber:

- en los resultados a nivel de población (por ejemplo, en términos de prevalencia anticonceptiva o fecundidad actual);
- en los productos a nivel de programa (por ejemplo, mejoramientos en la accesibilidad o calidad del servicio, aumento en el número de aceptantes); y
- a nivel de las áreas funcionales de la entrega de servicio (por ejemplo, los efectos en la prestación de servicio de los nuevos programas de capacitación de personal, los sistemas de supervisión o los procedimientos operativos clínicos).

Los resultados de programa pueden también medirse en relación a los costos; por ejemplo, a nivel de programa, el aumento en el número de aceptantes nuevos puede estar relacionado con el aumento en los costos de las diferentes intervenciones o paquetes. Para las áreas funcionales, los efectos de dos programas de capacitación pueden estar relacionados con sus costos.

Ventajas

Las principales ventajas de los experimentos aleatorizados pueden resumirse así:

- Versatilidad

Los experimentos aleatorizados pueden usarse para valorar los resultados de actividades del programa a diferentes niveles, además del impacto global.

- Alta validez interna

Este diseño es superior a todos los otros para la medición de resultados de programa, en términos de minimizar la amenazas a la validez interna.

- Pocos supuestos requeridos

Los principales supuestos requeridos son que: (a) la aleatoriedad haya producido grupos de tratamiento y de control equivalentes, (b) las influencias ajenas al estudio afecten ambos grupos por igual, (c) todos los

¹⁵ Obsérvese que si la aleatoriedad se realiza a nivel de grupo, la configuración ideal es tener tantos grupos pequeños como sea posible.

grupos de tratamiento (o miembros de grupos) reciban el tratamiento con la misma "intensidad", y (d) la asignación al grupo experimental no altere el comportamiento de los proveedores de servicio o de los sujetos de estudio respecto a los resultados estudiados.

- Análisis relativamente simple

Si los supuestos señalados en el punto anterior son válidos, en el análisis solo se requieren pruebas estadísticas simples¹⁶.

Limitaciones y consideraciones prácticas

A pesar de sus atractivos teóricos, la aplicación de los experimentos aleatorizados ha sido limitada en la evaluación de programas de planificación familiar, especialmente en la medición de resultados en la población. Varias razones se han citado para ello (éstas han sido revisadas sistemáticamente en Bauman et al., 1994), algunas son bien fundadas, pero otras son discutibles.

- Sensibilidades políticas o éticas

Un factor frecuentemente citado para no usar los experimentos aleatorizados es la sensibilidad política o ética de negar a algunos segmentos de la población un programa deseable. Aunque esto puede plantear un dilema en el corto plazo, en muchos casos puede no ser, ni posible, ni prudente, implementar un nuevo programa o intervención para cubrir de una vez a toda la población. Probar los programas en forma limitada y por etapas es una práctica muy común en la conducción de experimentos¹⁷. Es por ello que, no necesariamente, un experimento aleatorizado niega beneficios en un grado mayor que si el estudio experimental no se realizara.

- Tiempo y costo

Otro argumento en contra del uso de los experimentos aleatorizados es el tiempo y el costo. En relación al tiempo, debe reconocerse que los esfuerzos de un programa pueden tardar en madurar y, por lo tanto, si el objetivo primario de la evaluación es medir resultados a nivel poblacional a mediano o largo plazo, no hay más alternativa que esperar el tiempo necesario para que los cambios se den a nivel de la población. En este sentido, los experimentos aleatorizados no consumen más tiempo que otras formas de medir resultados de programas.

En términos de costo, a pesar de que los experimentos aleatorizados se consideran caros, su costo no es necesariamente mayor que el de una encuesta poblacional a gran escala, como las DHS (dependiendo, por supuesto, de la magnitud del estudio experimental y, sobre todo, del número de observaciones postest realizadas).

- Posibilidades de generalización

A no ser que el experimento se lleve a cabo a nivel nacional, la posibilidad de generalización de los resultados de los estudios experimentales en subpoblaciones suele ser incierta.

- Amenazas a la validez

Aunque menos vulnerables que otros diseños, los experimentos aleatorizados están potencialmente sujetos a varias amenazas serias a su validez. Entre ellas, las principales son:

- √ Contaminación

Cuando una parte o todo el grupo control está expuesto a la intervención que se estudia a través de la comunicación o migración entre los grupos. Esta amenaza está expuesta en el experimento Tai-chung, donde el aumento en las tasas de aceptación de anticonceptivos en áreas control tiende a ser resultado de la difusión de información sobre planificación familiar a los residentes de esas áreas.

- √ Influencias externas distorsionantes

Uno de los supuestos principales de un experimento aleatorio es que los grupos de tratamiento y control están expuestos a las mismas influencias externas a través de la vida del estudio. En los escenarios de un país en desarrollo, sin embargo, la presencia de varios donantes puede introducir nuevos programas o iniciativas durante un determinado período de tiempo;

¹⁶ Sin embargo, debido a los limitados de los tamaños muestrales y frecuentes violaciones de supuestos, los experimentos aleatorizados suelen ser tratados como cuasi-experimentos para propósitos analíticos y de esta forma la ventaja de la simplicidad analítica se pierde, al menos parcialmente.

¹⁷ En un sentido, la incapacidad de alcanzar toda la población objetivo de un programa a la vez provee una razón adicional para realizar experimentos aleatorizados.

por ejemplo, un ciclo programático quinquenal. En tales situaciones, es posible que el grupo control de una intervención pueda ser considerado como que está siendo mal atendido y se convierta en objetivo para otra intervención. También es posible que los grupos de tratamiento para una intervención puedan usarse como grupo de tratamiento para otra intervención, distorsionando los intentos de medición del impacto independiente de las dos intervenciones. Las influencias externas que afectan los grupos experimentales de manera diferente debilitan la validez de los experimentos aleatorizados. Entonces, para que los experimentos aleatorizados sean una opción realista para medir el impacto de programas en la población, se requiere mantener el control sobre la introducción de nuevas intervenciones mientras dure el estudio experimental.

Variaciones en el tratamiento

Especialmente en los programas a gran escala, los programas pueden ser implementados de manera diferente a través de las áreas geográficas y/o proveedores de servicios. Por ejemplo, uno o más elementos del programa pueden estar modificados para satisfacer las condiciones locales, o el programa prescrito sencillamente puede estar implementado con distintos niveles de intensidad en las diferentes áreas. Entonces, la medición de impacto del programa reflejará el impacto promedio de todas las modificaciones del programa y sus variaciones en intensidad, en vez de la del programa tal como fue diseñado. Si las evaluaciones de procesos no se realizan al mismo tiempo para medir y comprender tal variabilidad en la implementación, esta variabilidad no medida podría conducir a inferencias equivocadas sobre la magnitud del impacto del programa.

Aparte de las amenazas a la validez, hay al menos otra consideración práctica que puede limitar la utilidad de los experimentos aleatorizados para medir el impacto de programas: puede carecer de sentido desde el punto de vista programático, ubicar programas o intervenciones en forma aleatoria. De hecho, los programas suelen ser destinados a áreas geográficas o subgrupos poblacionales por dos razones diametralmente opuestas: (1) porque se considera que están mal atendidas o (2) porque son especialmente receptivos al pro-

grama. Aunque es posible realizar experimentos aleatorizados dentro de semejantes poblaciones especiales, el resultado probable sería diluir el impacto del programa a corto o mediano plazo. En esta situación, la necesidad de demostrar impacto puede competir con la capacidad de generarla.

Para concluir, recomendamos con vehemencia el uso de experimentos aleatorizados, pero reconocemos que las realidades prácticas suelen limitar su uso, particularmente en los estudios de impacto a nivel nacional. Hay bases, sin embargo, para la creencia de que los experimentos aleatorizados han sido subutilizados en investigación de operaciones y en estudios que involucran los resultados a nivel de programa y deberían ser el método seleccionado para trabajos de ese tipo. Debido a las limitaciones de los experimentos aleatorizados para medir resultados de programas a nivel nacional, es necesario considerar los enfoques alternativos.

Cuasi-experimentos

Descripción

El término "cuasi-experimento" se refiere a un grupo de diseños de investigación experimentales en los cuales los sujetos o grupos de estudio no son asignados aleatoriamente. Los diseños cuasi-experimentales más usados, "diseños de control contruados", siguen la misma lógica e involucran la comparación de los sujetos del grupo de tratamiento con el grupo control de la misma forma que en los experimentos aleatorizados. En otros diseños, llamados "diseños de control reflejos", los sujetos del grupo de tratamiento sirven como sus propios controles y se usan métodos de series de tiempo para medir el impacto neto del programa (Rossi y Freeman, 1993). Aunque más vulnerables a las amenazas a la validez que los experimentos aleatorizados, los cuasi-experimentos no requieren la asignación aleatoria a los grupos experimentales y, por lo tanto, tienden a ser más factibles que los experimentos aleatorizados.

Los diversos diseños de investigación cuasi-experimentales se discuten con detalle en otras fuentes (Campbell y Stanley, 1963; Cook y Campbell, 1979; Rossi y Freeman, 1993; Fisher et al., 1991). En este documento, centramos nuestra atención en un diseño que tiene la más amplia aplicabilidad en la valoración de impacto de programas de planificación familiar: el diseño de grupo control pretest-postest no equivalente¹⁸.

Diseño y Análisis

El esquema básico para el diseño de grupo control no equivalente y pretest-postest es idéntico al del experimento aleatorio pretest-postest que aparece en la Gráfica IV-2, excepto que la aleatoriedad no se usa para asignar sujetos de estudio en los grupos de tratamiento. Más bien, uno o más grupos control (o de comparación) son identificados, de manera que se asemejen al grupo de tratamiento en tantos factores como sea posible. En muchas aplicaciones, los grupos de tratamiento y de comparación son pareados respecto a las características que se consideran asociadas con el resultado bajo estudio (características diferentes, por supuesto, al programa o intervención que se evalúa). Por ejemplo, los sujetos o subgrupos de población que son lo más semejantes posible al grupo de tratamiento respecto a condición económica, ubicación geográfica, etnicidad y otras características podrían ser seleccionados a propósito, para servir como grupo de comparación.

Como en los experimentos aleatorizados, el impacto del programa se mide en este diseño por la diferencia entre el cambio en las medidas de resultados para el grupo de tratamiento y las del grupo de testigo, más o menos el error aleatorio:

$$\text{Impacto} = (O_2 - O_1) - (O_4 - O_3) \pm \text{error}$$

donde: O_1 y O_2 = medidas pre- y postest, respectivamente, para el grupo de tratamiento;

O_3 y O_4 = medidas pre- y postest para el grupo testigo; y

error = efectos aleatorios de medición.

En un cuasi-experimento, es fundamental compensar las diferencias entre el grupo de tratamiento y el grupo control mediante la aplicación de métodos estadísticos multivariados. Aún en estudios pareados, generalmente es necesario introducir controles estadísticos para controlar las diferencias en los factores que no es posible parrear. Las diferencias en los grupos experimentales que no están adecuadamente controladas se reflejarán

en el componente de error de diseño señalado anteriormente e influirá directamente en la magnitud del impacto estimado. Esta es la principal desventaja de los cuasi-experimentos, en comparación con los experimentos aleatorizados en los que los efectos de diseño se han minimizado mediante la asignación aleatoria a los grupos experimentales. La validez de los estudios cuasi-experimentales depende, entonces, de la efectividad con la cual se minimizan los efectos de diseño mediante el pareo y el análisis multivariable.

En las Gráficas IV-4 y IV-5 se ofrecen algunos ejemplos de aplicaciones del diseño de grupo control no equivalente. La Gráfica IV-4 muestra los resultados de una variante relativamente robusta del diseño considerado que incluye múltiples observaciones del indicador de resultados antes y después del programa. La Gráfica IV-5 muestra la situación más típica en la que se hacen mediciones de la intervención "pre" y "post" una sola vez.

Ventajas

Las principales ventajas del diseño de grupo control no equivalente son:

- Ofrece una aproximación al experimento aleatorio cuando la aleatoriedad no es posible.
- Es versátil. Como los experimentos aleatorizados, los cuasi-experimentos pueden

¹⁸ Varios diseños cuasi-experimentales potencialmente poderosos han sido excluidos de la discusión porque es poco probable que puedan ser aplicables en la evaluación de programas de planificación familiar. Por ejemplo, el diseño de series de tiempo es relativamente fuerte donde existen largas series temporales. Pero esto ocurre en muy pocos lugares, principalmente donde se han implementado sistemas de vigilancia epidemiológica basados en la población (por ejemplo, en Matlab, Bangladesh, y en Cebu, Filipinas). Un segundo diseño, el de la discontinuidad en la regresión, es tal vez el más poderoso de los diseños cuasi-experimentales, pero debido a que el nivel de tamizaje del programa requerido para una aplicación significativa del diseño (por ejemplo, el uso de ingreso u otro criterio de elegibilidad para seleccionar los participantes del programa) es poco probable en programas de planificación familiar (que se basan en el concepto de libre elección del consumidor), es difícil prever circunstancias bajo las cuales este diseño sería aplicable. Mayor detalle sobre estos diseños se ofrece en Rossi y Freeman (1993) y Cook y Campbell (1979).

para medir resultados en la población o en los programas.

- Con diseño, control y análisis adecuados, los cuasi-experimentos pueden proveer una evidencia del impacto del programa casi tan robusta como los experimentos aleatorizados y más robusta que la mayoría de los estudios no experimentales.

Limitaciones y consideraciones prácticas

El diseño de grupo control no equivalente está sujeto a los mismos supuestos y limitaciones generales que los experimentos aleatorizados indicados anteriormente (además de los relacionados con la falta de aleatoriedad).

Además:

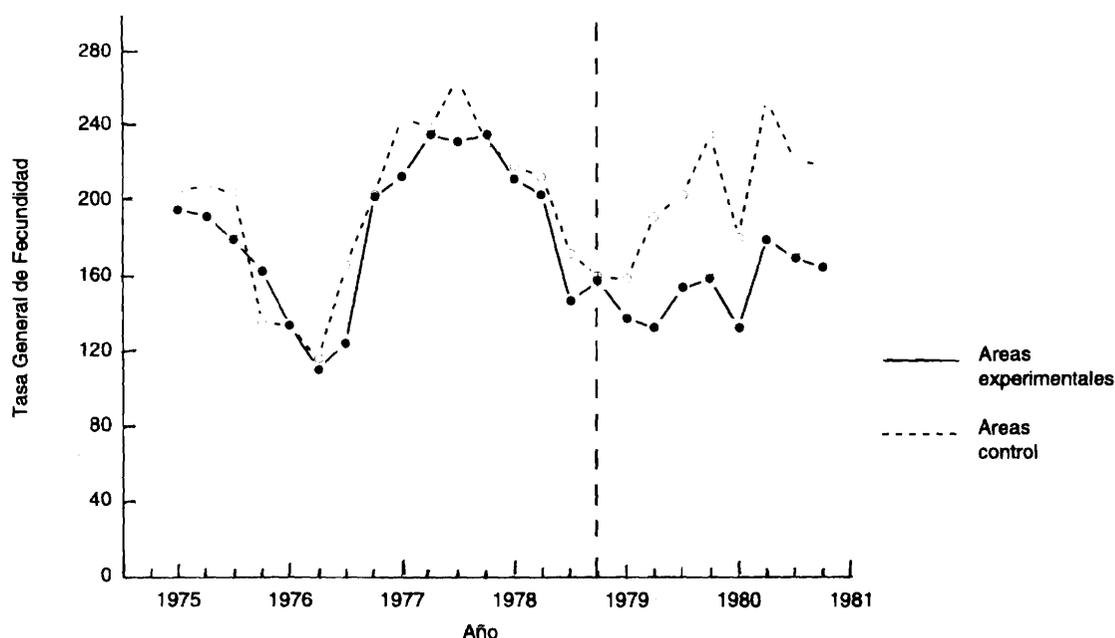
- El diseño es más vulnerable al sesgo de selección que los experimentos aleatorizados; o sea, que las diferencias en las características de los grupos experimentales estarán correlacionadas con los resultados que se estudian,

distorsionando, por lo tanto, los hallazgos de impacto.

- Depende mucho de los métodos estadísticos multivariados y, por lo tanto, es sensible al uso de modelos estadísticos apropiados y al tratamiento correcto de los problemas de estimación estadística.

En la práctica, en los estudios cuasi-experimentales suele ser posible compensar las diferencias en los grupos experimentales respecto a las características clave mediante el pareo y el análisis multivariable. Una preocupación persistente es, sin embargo, si los grupos experimentales difieren en factores no observados que influyen en los resultados estudiados. En contraste con los efectos distorsionantes de las diferencias en los factores que son observables/medibles y que pueden ser tomadas en cuenta mediante el pareo y la introducción de variables de control en los modelos estadísticos multivariados, los factores que no son observables (por

Gráfica IV-4
Ejemplo de aplicación del diseño cuasi-experimental con Grupos control no equivalente y múltiples observaciones antes y después, Matlab, Bangladesh, 1974-1980



Fuente: Philips, J. et al., 1982: 131-140

disposición o motivación diferencial) no pueden ser compensados de esta forma y pueden conducir a estimaciones equívocas y/o sesgadas del impacto del programa. Este factor de "heterogeneidad no observada" es, de hecho, una preocupación en todo diseño que no sea un experimento aleatorio.

Se ha invertido esfuerzos considerables en el desarrollo de métodos estadísticos para medir y controlar los efectos de los factores no observados. Estos desarrollos han sido incorporados en los métodos basados en regresión para medir el impacto de programas discutidos en el próximo capítulo. Aunque estos desarrollos no han sido usados muy frecuentemente junto con los diseños de investigación cuasi-experimentales hasta la fecha, no habría razón para que no fueran usados en investigaciones futuras.

Métodos de regresión multinivel

Visión de conjunto

Las valoraciones de impacto que se basan en los métodos de regresión multinivel pertenecen a la categoría general de los estudios no experimentales o de observación; o sea, estudios en los cuales no hay grupos experimentales o de control como tales. Debido a que se trata de un enfoque no experimental, los tratamientos varían de área a área como resultado de los procesos de toma de decisiones que están más allá del control del investigador a cargo de la evaluación. Los criterios subyacentes en las decisiones de ubicación y de asignación de recursos del programa pueden ser conocidos o desconocidos por el investigador¹⁹.

Los métodos de regresión multinivel son una extensión de los métodos de regresión multivariados por áreas (o ecológicos) que fueron muy usados hacia finales de la década

¹⁹ Aunque la mayoría de los programas tienen lineamientos o criterios para las decisiones de asignación de recursos, las variaciones en el esfuerzo del programa entre las áreas geográficas también están influenciadas por políticas anteriores, la presencia de muchos actores en la prestación de servicios de planificación familiar, programas de implementación no equitativos y factores políticos. Entonces, la relación entre las políticas actuales y la asignación de los recursos del programa entre las unidades geográficas pueden reflejar o no las prioridades actuales de asignación de recursos.

Gráfica IV-5

Ejemplo de diseño cuasi-experimental de grupo control no equivalente con una sola observación pre- y post-intervención, Guatemala, 1983-1984

Para valorar el impacto de tres estrategias de comunicaciones diseñadas para incrementar el conocimiento y la aceptación de la vasectomía en Guatemala, se hicieron mediciones "antes y después" en cuatro comunidades con características socioeconómicas similares. Las tres estrategias de comunicación fueron: (1) la radio, (2) un promotor masculino y (3) radio y promotor masculino. Una comunidad fue seleccionada para implementar cada estrategia que iba a ser probada, y una cuarta fue seleccionada como comunidad control. Se recolectaron con encuestas datos de bases y de seguimiento para n=400 hombres en edad reproductiva en cada una de las cuatro comunidades en junio de 1983 y de nuevo en julio de 1984, junto con estadísticas de servicio indicando el número (mensual) de operaciones realizadas en cada comunidad.

Se valoraron los efectos del programa de comunicaciones sobre el conocimiento y las actitudes comparando las mediciones pre- y post-intervención de ciertos indicadores. Se usó regresión logística para controlar las diferencias iniciales entre las cuatro comunidades y posibles efectos "históricos". El impacto fue valorado comparando las tasas de vasectomía pre- y post-intervención en las comunidades respectivas. Algunos hallazgos clave fueron los siguientes:

Grupo experimental

	Radio y promotor		Solo radio		Solo promotor		Control	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Tasas de vasectomía								
% de interesados en hacérsela	16.5	22.4	31.9	31.7	33.1	37.5	22.8	27.4
% interesados que se la han hecho	0.5	0.8	1.5	2.0	1.0	3.0	0.3	1.0

Interpretación: El programa no aumentó el interés en hacerse una vasectomía en el grupo objetivo (hombres que habían oído de la operación, no querían más hijos, todavía no estaban esterilizados, ni lo estaban sus esposas); o sea, los aumentos en las tres comunidades no diferían significativamente del aumento en la comunidad control. En términos de la verdadera prevalencia de la vasectomía, solamente en la comunidad de "sólo promotor" el cambio en la prevalencia excedió el observado en la comunidad control por un margen estadísticamente significativo.

Fuente: Bertrand et al., 1987

de los setenta y hasta mediados de la década de los ochenta (por ejemplo, Hermalin, 1975, 1979, y 1982; Poston y Chu, 1987; Chamrathirong et al., 1986). La idea principal de una regresión de áreas es tratar de demostrar una relación estadística entre las medidas de la actividad o esfuerzo del programa e indicadores de resultados (por ejemplo, prevalencia anticonceptiva, tasa global de fecundidad) usando áreas geográficas como unidad de análisis y manteniendo constante los efectos de los factores ajenos al programa, como la composición por edad y sexo, urbanización, educación femenina y participación en la fuerza laboral, etnicidad, etc. Los efectos de los factores ajenos al programa sobre los resultados de interés son controlados mediante la aplicación de los métodos de regresión. El enfoque trata de responder la pregunta "excluyendo los factores ajenos al programa de planificación familiar ¿Tienden las comunidades con mayor presencia del programa a tener tasas más altas de prevalencia anticonceptiva o tasas más bajas de fecundidad que aquellas con

menos presencia del programa?"

El enfoque multinivel extiende las ideas fundamentales de la regresión de áreas al caso en que se usan variables que han sido medidas en distintos niveles; por ejemplo, variables medidas para cada entrevistado, conglomerado muestral, distrito y/o mayores niveles de agregación. El uso de datos medidos en diferentes niveles permite una percepción más profunda de las trayectorias causales seguidas por los insumos del programa de planificación familiar para influir en el comportamiento de los individuos y de las formas como los programas interactúan con otras variables (por ejemplo, la provisión de servicios en pueblos con mayores niveles de educación pueden tener un mayor efecto sobre el uso de anticonceptivos que en pueblos con niveles de educación inferiores).

Debido a estas características deseables, así como la amplia disponibilidad de datos a nivel individual mediante encuestas como la Encuesta Mundial de Fecundidad y la Encuesta Demográfica y de Salud, el enfoque multinivel ha reemplazado bastante el enfoque de regresión de áreas en evaluaciones recientes. Todavía se usan los métodos de regresión de áreas, pero solamente en análisis de corte transversal (ver Bongaarts, 1993, para un ejemplo de aplicación reciente).

En esta sección describimos tres tipos de modelos de regresión multinivel que pueden usarse para medir el impacto de programas. La pregunta fundamental que cada modelo intenta responder aparece resumida en la Gráfica IV-6.

El modelo básico de corte transversal

El modelo transversal multinivel trata de valorar si existe una relación estadística entre las variables de planificación familiar (por ejemplo, presencia de una clínica de planificación familiar en la comunidad, el número de métodos anticonceptivos ofrecidos dentro de un área de 30 kilómetros de la comunidad) y los resultados de la planificación familiar, controlando los factores socioeconómicos y otros factores ajenos al programa. El modelo usa aleatoriamente comunidades seleccionadas u otros agregados geográficos (por ejemplo, municipalidades, distritos, provincias, etc.) y muestras individuales de mujeres o parejas para cada comunidad como unidades de análisis.

Gráfica IV-6

Preguntas básicas en los principales modelos de regresión para medir el impacto de programas

Modelo	Pregunta
Transversal básico	¿Existe una relación estadística entre las variables del programa y de resultados (por ejemplo, la prevalencia anticonceptiva, la fecundidad) si se controlan estadísticamente los efectos de otros factores observados que influyen en el resultado?
De efectos aleatorios	¿Existe una relación estadística entre las variables del programa y de resultados si se controlan estadísticamente los efectos de otros factores observados que influyen en el resultado y los de factores no observados que influyen conjuntamente en las variables del programa y en las variables de resultados?
Panel	¿En qué medida están asociados los cambios observados en las variables de resultados con los cambios en las variables del programa si se toman en cuenta los efectos de los niveles iniciales y los cambios en los otros factores (tanto observados como no observados)?

A manera de ilustración, consideremos un modelo de corte transversal relativamente simple que relaciona los insumos del programa con el comportamiento de la fecundidad. Un modelo típico podría incluir tres tipos de variables medidas en dos niveles (i.e., a nivel individual o del hogar y a nivel de comunidad):²⁰

- factores específicos para las mujeres y hogares (por ejemplo, edad, paridez, educación, demanda de hijos, recursos en el hogar, estructura familiar, etc.);
- factores específicos para las comunidades u otras agregaciones poblacionales, pero comunes a todos los hogares e individuos dentro de la misma comunidad (por ejemplo, condiciones del ambiente, infraestructura de la comunidad, condiciones del mercado laboral, etc.)²¹; y
- medidas a nivel de comunidad del vigor del programa de planificación familiar, que se suponen comunes a todos los hogares e individuos en la comunidad (por ejemplo, la presencia de métodos de planificación familiar disponibles en establecimientos ubicados a una distancia específica de la comunidad, calidad de los servicios).

El modelo básico puede ser expresado de la siguiente forma:

$$\text{Variable de resultado} = \text{Factores del programa} + \text{Factores individuales} + \text{Factores de la comunidad} + \text{Interacción entre factores}^{22} \pm \text{Error}^{23}$$

Gráfica IV-7

Ejemplo de aplicación del modelo multinivel transversal, Tailandia

Diseño Entwisle et al. (1984) usaron datos de la segunda ronda del Estudio de Prevalencia Anticonceptiva de Tailandia (CPS2) y métodos de regresión multinivel para valorar los efectos de la disponibilidad de PPS de planificación familiar sobre la probabilidad de uso de anticonceptivos en el área rural de Tailandia. En el análisis usaron datos recolectados entre 4,956 mujeres con edades de 15-44 años que estaban casadas o en unión al momento de la encuesta. Se clasificaron los caseríos muestrales en tres grupos sobre la base de su cercanía a distintos tipos de establecimientos que ofrecen servicios de planificación familiar: (1) caseríos ubicados cerca (i.e., dentro de 4 km.) de un centro de salud distrital, (2) caseríos ubicados cerca de un centro de salud municipal ("tambol") y (3) caseríos ubicados cerca de cualquier tipo de establecimiento. Las variables explicativas a nivel individual fueron la edad, la escolaridad y el deseo de más hijos.

Resultados Los resultados de la regresión mostraron fuertes efectos de la disponibilidad de servicios de planificación familiar y de las variables a nivel individual, así como fuertes interacciones entre la disponibilidad del servicio y el deseo de más hijos y la escolaridad. Los chances ajustados (calculados con los resultados de la regresión) de usar un método de anticoncepción moderno entre las mujeres que no deseaban más hijos (en comparación con las que deseaban más hijos) se muestran a continuación para las categorías de las variables edad y disponibilidad del servicio.

Disponibilidad del servicio	Edad		
	15-24	25-34	35-44
Cerca centro salud distrital	1.90	4.45	5.29
Cerca centro salud Tambol	2.07	4.06	7.40
> 4 km. de cualquiera	2.04	1.72	4.77

Interpretación Los efectos más fuertes de disponibilidad del servicio sobre el uso de anticonceptivos fueron observados entre las mujeres de 25-34 años. Las mujeres que vivían cerca de los centros de salud distritales o "tambol" tenían un chance más de dos veces mayor que el de las mujeres que vivían a más de 4 km. de cualquiera de ellos de estar usando un método anticonceptivo moderno en el momento de la encuesta, con los otros factores controlados estadísticamente.

Fuente: Entwisle et al., 1984

²⁰ Las variables a nivel individual y del hogar usualmente son tratadas como si pertenecieran al mismo nivel. Modelos con dos o más niveles son posibles, pero el software disponible actualmente es limitado en varias formas (por ejemplo, el paquete de software MLn usa aproximaciones en modelos de 3 y 4 niveles y puede manejar solo variables de resultados dicotómicas y continuas). El lector interesado en más detalles sobre el paquete MLn puede consultar Woodhouse (1995).

²¹ Un cuarto tipo de variable es frecuentemente utilizada en los modelos multinivel: variables a nivel de comunidades que son derivadas agregando o haciendo un promedio de las observaciones individuales dentro de las comunidades muestrales. Algunos ejemplos son el ingreso medio por hogar o la proporción de viviendas con electricidad. Este tipo de variable sirve el mismo propósito que las variables a nivel de comunidad descritas en el texto y ha sido omitido de la discusión con el objeto de simplificar la presentación.

La magnitud y la significancia estadística de los parámetros de las variables del programa en el modelo y las interacciones entre variables del programa y otras variables son de interés primordial para la evaluación. Los parámetros de regresión para las variables del programa indican la intensidad de la asociación entre las medidas del programa y los resultados a nivel individual (por ejemplo, uso de anticonceptivos o fecundidad) cuando los efectos de otros factores incluidos en el modelo han sido controlados. Los términos de la interacción proveen información sobre si el programa tuvo un mayor impacto sobre ciertos subgrupos poblacionales que sobre otros. Por ejemplo, la inclusión de términos de interacción en el modelo podrían permitir al evaluador probar si la apertura de CBDs tuvo mayor influencia sobre la conducta anticonceptiva de las mujeres de menor ingreso (que podrían tener menos acceso a los establecimientos fijos) que la de las mujeres de ingresos altos.

En la Gráfica IV-7 se ofrece un ejemplo de aplicación del modelo de corte transversal a la medición de los efectos de algunas características selectas del entorno de oferta de planificación familiar sobre el uso de anticonceptivos.

Una limitación importante del modelo de corte transversal es la dificultad de escoger las relaciones causa-efecto entre las mediciones realizadas en un determinado momento. En el modelo de corte transversal, una relación positiva observada entre las variables de los programas de planificación familiar y las variables de resultados (por ejemplo, uso de anticonceptivos o fecundidad) puede ser interpretada como que el programa ha influido o causado el resultado observado, excluyendo las otras variables del modelo. Sin embargo, también es posible que la trayectoria causal realmente fue en dirección opuesta; o sea, que la demanda de servicios de planificación familiar predispuso una oferta mayor.

Los administradores de programas suelen ubicar justificadamente las clínicas u otros tipos de establecimiento de manera que se pueda satisfacer la demanda existente de servicios o seleccionan lugares con poblaciones aventajadas o con infraestructura superior (por ejemplo, carreteras, electricidad, etc.). Tales factores, sin duda han influido en las decisiones de ubicación de muchos programas de planificación familiar, particularmente en las primeras etapas del desarrollo del programa. Existe el peligro de que las evaluaciones de impacto basadas en datos de corte trans-

versal puedan sobrevalorar el verdadero nivel de impacto del programa, en la medida que las poblaciones servidas por estos establecimientos estén predispuestas a un mayor uso de anticonceptivos y una menor fecundidad. Obsérvese que la forma como se ubican los programas puede conducir también a una subestimación del impacto del programa con base en datos de corte transversal; por ejemplo, si los programas dan preferencia a áreas de alta fecundidad y mortalidad.

El asunto fundamental es si las variables del programa de planificación familiar deben ser consideradas como variables "endógenas" o "exógenas" en las ecuaciones de regresión²⁴. Si los programas son implementados de manera uniforme y aleatoria a través de las unidades sub-nacionales, entonces las variables

²² El término "interacción" se refiere a la dependencia estadística del efecto de un determinado factor o variable del valor de otras variables. Por ejemplo, los efectos de agregar puntos CBD podría depender del grado de acceso a clínicas fijas en un escenario determinado. Las interacciones entre los factores individuales y los factores programáticos, entre los factores de comunidades y los factores programáticos y entre los factores individuales y los factores de las comunidades son posibles.

²³ En modelos multinivel, tanto los errores a nivel de comunidades como a nivel individual están presentes. Para lectores más inclinados hacia la estadística, el modelo multinivel de corte transversal está expresado en forma de ecuación de regresión en el Apéndice A. El lector notará la especificación de los términos de error a nivel de comunidades y a nivel individual en las ecuaciones de regresión.

²⁴ Los factores o variables endógenas son variables independientes o predictoras que están determinadas por el mismo conjunto de factores o los mismos procesos de toma de decisiones que determinan la variable de resultado que está en estudio. Los factores exógenos, en cambio, son variables que no están determinadas por factores que también influyen en las variables programáticas. Por ejemplo, la participación en la fuerza laboral es endógena en una ecuación de regresión que predice el uso de anticonceptivos, puesto que tanto la participación en la fuerza laboral como el uso de anticonceptivos son variables de elección para cada mujer que están influenciadas por factores comunes (por ejemplo, educación, tamaño y estructura del hogar). El nivel de salarios de la comunidad para las mujeres sería un ejemplo de un factor exógeno en una regresión de este tipo, puesto que esta variable está determinada por factores que no están sujetos a elección individual.

del programa pueden ser consideradas como factores o variables exógenas sin riesgo de incurrir en sesgos de estimación. Sin embargo, si los programas son implementados de manera no uniforme o no aleatoria de acuerdo con algún tipo de proceso de toma de decisiones, entonces el problema de endogeneidad cuenta. Si es así, esto significaría una violación del supuesto estándar de regresión de que no existe correlación entre las variables de la ecuación y los términos de error. Una preocupación importante es si los factores "no observados" que están correlacionados con los resultados de interés pueden haber influenciado las decisiones de ubicación del programa. El punto crucial es este -- si las decisiones de "ubicación" del programa se hacen sobre la base de factores que no pueden ser medidos o controlados en un modelo estadístico, uno puede obtener estimaciones inconsistentes o sesgadas del impacto del programa. El lector que se interese por conocer mayores detalles de estos temas en el contexto de la evaluación del impacto de programas de planificación familiar puede consultar Bollen et al. (1992 & 1995).

Al valorar el impacto con modelos de regresión multinivel, es fundamental considerar el tema de la endogeneidad. Seguidamente se presentan dos enfoques para hacerlo: los modelos de efectos aleatorios de ecuaciones múltiples y los modelos de efectos fijos en un panel.^{25 26}

Los modelos de efectos aleatorios de ecuaciones múltiples

Se han desarrollado varios enfoques para lidiar con los problemas estadísticos ocasionados por las variables no observadas de los datos de corte transversal. El lector que desee conocer mayores detalles y considerar enfo-

ques alternativos puede consultar Bollen et al. (1992, 1995) y Mroz y Guilkey (1992). Aquí, centramos nuestra atención en un enfoque que el Proyecto EVALUATION encuentra particularmente prometedor: una adaptación del modelo de efectos aleatorios a los tipos de datos que generalmente están disponibles en las encuestas tipo DHS, complementado con información selecta de programas de planificación familiar.

Aunque los modelos de efectos aleatorios generalmente están asociados con datos longitudinales (contrario a los de corte transversal), los desarrollos metodológicos recientes han extendido las ideas fundamentales para usar datos de corte transversal cuando cierto tipo de información:

- ha sido recolectada retrospectivamente en los cuestionarios de las encuestas (por ejemplo, las historias de embarazos retrospectivos en las DHS);
- pueden obtenerse de la información en los cuestionarios de la DHS y del SAM mediante la definición de fechas en el pasado (ejemplo: si la mujer asistía a la escuela cuando tenía 15 años de edad o si esta clínica ya había sido abierta en 1990); o
- está disponible en otras fuentes y pueden usarse para complementar la información de las encuestas tipo DHS (Mroz y Guilkey, 1992; Newman, 1988).

El modelo de efectos aleatorios considera el problema de la endogeneidad mediante el control estadístico de los efectos de los factores no observados que influyen tanto en las decisiones de ubicación de los programas como en los resultados que están siendo estudiados. La idea fundamental es estimar la distribución de estos factores y controlarlos estadísticamente en el modelo de regresión. Sin embargo, los procedimientos estadísticos para lograr esto son muy complejos y están más allá del alcance de este documento. Para mayores detalles se refiere al lector interesado a Bollen et al. (1995), Mroz y Guilkey (1992), Guilkey y Cochrane (1994), y Newman (1988).

El modelo también usa un enfoque de ecuaciones estructurales, en el cual las influencias de los factores "observables" a nivel de comunidades sobre la ubicación de los recursos del

²⁵ Se observará que el modelo básico de corte transversal descrito anteriormente también puede ser estimado dentro de un esquema de efectos aleatorios dependiendo de los supuestos sobre los términos de error. La característica que distingue el enfoque de efectos aleatorios descrita en la sección siguiente es el uso de ecuaciones múltiples que son estimadas simultáneamente.

²⁶ Notese que otros enfoques también están disponibles pero son considerados por el Proyecto EVALUATION como meros prometedores que los enfoques presentados en esta publicación. El lector interesado en valoraciones críticas de estos puede consultar Mroz y Guilkey (1992) y Bollen et al. (1995).

programa en los conglomerados muestrales se miden en un conjunto de ecuaciones de regresión y los efectos de las variables de programa sobre los resultados de interés (por ejemplo, la fecundidad, el uso de anticonceptivos, etc.) en otra ecuación. Ambas ecuaciones son estimadas simultáneamente usando un procedimiento de estimación de máximo-verosimilitud con información completa (MVIC). El objetivo es deshacerse de las influencias de los factores no observados que pueden haber influenciado las decisiones de ubicación del programa en la ecuación que relaciona las variables del programa con los resultados²⁷. La magnitud y la significancia estadística de los coeficientes de las variables del programa en la ecuación de resultado son de interés primordial para la evaluación del programa.

El modelo puede expresarse de la siguiente forma:

Ecuación(es) de ubicación del programa:

Variables del programa =
Factores de comunidad
observados + Factores de
comunidad no observados \pm Error.

Ecuación de resultado:

Variables de resultado =
Variables del programa + Factores
individuales + Factores de comunidad no
observados²⁸ +/- Error²⁹.

En las Gráficas IV-8 y IV-9 se presentan dos ejemplos de aplicaciones del modelo de efectos aleatorios modificado. En el primer ejemplo, el modelo se usa para valorar el impacto de características selectas del entorno de la oferta de planificación familiar en el uso actual de anticonceptivos modernos en Zimbabwe (Guilkey y Cochrane, 1994). En el segundo ejemplo, se estima un modelo un poco más elaborado para la zona rural de Tanzania. Aprovecha datos retrospectivos de la historia de embarazos (disponibles en encuestas DHS estándar) y el período de tiempo que los métodos anticonceptivos han estado disponibles en los conglomerados muestrales visitados en SAM de la DHS para estimar el impacto del programa sobre la fecundidad durante las dos décadas anteriores a la DHS de Tanzania de 1991 (Angeles et al., 1995).

Conviene notar que en ambos ejemplos los resultados se presentan como "simulaciones de política". Las simulaciones ofrecen un me-

dio para expresar resultados multivariantes en una forma que es fácilmente comprensible para los administradores de programa y los encargados de la toma de decisiones; o sea, en forma de respuestas a preguntas como "¿qué pasa si...?". Por ejemplo, ¿Cuál habría sido la tasa global de fecundidad durante un determinado período si todos los conglomerados muestrales hubieran tenido todos los servicios de planificación familiar y los métodos disponibles? Los resultados de la simulación son derivados de los resultados de la regresión sustituyendo valores de variables selectas del programa en las ecuaciones de regresión para delinear diferentes escenarios programáticos (por ejemplo, cobertura total de pueblos por CBDs, agregar un proveedor de servicio de planificación familiar capacitado en cada centro de salud, etc.) y comparando los valores "esperados" que resultan de la variable de resultado bajo estudio con los resultados que se basan en los verdaderos valores de estas variables medidas en una encuesta³⁰.

El modelo panel de efectos fijos

La segunda alternativa para lidiar con el problema de los factores no observados que influyen en las decisiones de ubicación del programa, el modelo panel de efectos fijos,

²⁷ La idea fundamental aquí es tratar de incluir el máximo de factores que pueden haber influido en las decisiones de ubicación del programa como sea posible en la primera ecuación para controlar sus efectos.

²⁸ Notar que un término para los inobservables del nivel individual también puede agregarse a la ecuación y medirse usando los mismos métodos usados para estimar los efectos de inobservables de nivel comunitario. Con el objeto de simplificar la presentación no son considerados aquí.

²⁹ El modelo está expresado en formato de regresión en el Apéndice A.

³⁰ Obsérvese que en el ejemplo de la Gráfica IV-8, que se basa en datos de Zimbabwe, no se incluye una ecuación para la instalación del programa y por lo tanto es necesario suponer que el programa de planificación familiar es exógeno. Este tipo de modelo es útil para examinar las contribuciones de los programas al uso de anticonceptivos que resultan indirectamente de los efectos del programa sobre las intenciones de fecundidad.

Gráfica IV-8

Ejemplo de aplicación del modelo de efectos aleatorios, Zimbabwe

Diseño

El modelo modificado de efectos aleatorios fue aplicado recientemente a datos de la Encuesta Demográfica y de Salud Reproductiva de 1988/89 de Zimbabwe y la Encuesta de Disponibilidad de Servicio de 1989/90 de Zimbabwe para medir los efectos del acceso a los servicios de planificación familiar sobre el uso de anticonceptivos. Los datos de la encuesta de 2,050 mujeres casadas en esa fecha fueron usados en el análisis, junto con datos a nivel de comunidades y del servicio de planificación familiar para 167 comunidades (o sea, los conglomerados muestrales de la DHS). Los datos de servicio contenían medidas de acceso físico a los puntos de entrega de servicios, así como una serie de medidas de la idoneidad y funcionamiento del sistema de entrega de servicio (por ejemplo, la presencia de electricidad y agua entubada, el número de personal capacitado en planificación familiar, la disponibilidad de métodos y oferta de anticonceptivos y la cortesía del personal). El modelo estadístico empleado fue un sistema de cuatro ecuaciones

Las variables de resultado para las cuatro ecuaciones fueron: (1) la probabilidad de que una entrevistada haya tenido "r" hijos durante el transcurso de su vida reproductiva, (2) la probabilidad de que una entrevistada haya tenido "r" hijos fallecidos durante el transcurso de su vida reproductiva, (3) las intenciones de fecundidad y (4) el método anticonceptivo actual. Las variables independientes en las cuatro ecuaciones consideradas fueron variables exógenas a nivel individual y a nivel de hogares, características selectas de las comunidades y un conjunto de variables que miden varios aspectos del entorno de la oferta de los servicios de planificación familiar (medidos a nivel de comunidades). Cada ecuación también incluyó un parámetro que representa los factores no observados que hipotéticamente influyen en las cuatro variables de resultados. Las ecuaciones fueron estimadas simultáneamente usando un estimador de máximo-verosimilitud con información completa.

Resultados

Las siguientes variables a nivel individual mostraron efectos estadísticamente significativos sobre el uso de anticonceptivos modernos en la forma reducida de las ecuaciones: edad de la entrevistada, religión, años de escolaridad y si la entrevistada residía en una comuna. Dos variables a nivel de comunidad también surgieron como determinantes significativos del uso de anticonceptivos: las oportunidades educativas disponibles en la comunidad y la presencia de un punto CBD en la comunidad. La mayoría de los parámetros de heterogeneidad no observados resultaron estadísticamente significativos, indicando que su omisión del modelo habría resultado en estimaciones sesgadas de los efectos de las otras variables del modelo.

regresiones revelaron los siguientes efectos estimados de los puntos CBD:

Proporción que usa

Variable y condición	Ningún Método	Método Moderno	Método Tradicional
Observado	.47	.44	.09
Punto CBD en la comunidad	.44	.47	.09
Ningún CBD	.50	.40	.10

Las simulaciones basadas en los resultados de las

Interpretación

Los predictores más fuertes del uso de anticonceptivos modernos fueron las características a nivel individual. De las variables del programa de planificación familiar probadas, solo la presencia de un punto CBD en la comunidad tuvo un efecto significativo sobre el uso de anticonceptivos modernos. Las simulaciones indican que si los puntos CBD

fueran establecidos en cada comunidad, se esperaría que la prevalencia de métodos anticonceptivos modernos creciera en alrededor de un 7% sobre los niveles observados (de .44 a .47) y en un 17% del nivel que prevalecería si no hubieran puntos CBD (de .40 a .47), esto luego de controlar los efectos de los otros factores.

Fuente: Guilkey y Cochrane, 1994.

Gráfica IV-9

Ejemplo de aplicación del modelo de efectos aleatorios a la medición del impacto en la fecundidad del programa de IPF, Zona rural de Tanzania, 1969-1991

Diseño

El modelo transversal de efectos aleatorios se aplicó recientemente a los datos de la DHS de 1991 de Tanzania y su Módulo de Disponibilidad del Servicio. El análisis se centró en la estimación del impacto en la fecundidad de la PF en Tanzania durante el periodo 1969-1991 y se basó en los datos de la encuesta de hogares de 5,215 mujeres residentes en 242 conglomerados muestrales rurales que tenían menos de 35 años en 1991. La información sobre la ubicación de los establecimientos de PF, la disponibilidad y fecha de inicio de los servicios de PF y otras características de la entrega del servicio fue proveída por el SAM. Debido a la falta de disponibilidad de series de tiempo confiables en Tanzania, fue necesario estimar los niveles y tendencias de la fecundidad y mortalidad infantil de las historias de embarazo de la DHS y establecer las fechas de inicio de los servicios de PF retrocediendo en el tiempo los datos de la DHS y el SAM. Las cifras históricas sobre gastos de salud del gobierno estaban, sin embargo, disponibles y fueron usadas en el análisis.

La variable de resultado usada en el estudio fue la probabilidad de que una entrevistada haya tenido un hijo en cualquier año *i* durante el periodo de estudio. Las siguientes variables fueron incluidas como predictores: (1) edad de la entrevistada en el año *i*, (2) nivel de escolaridad, (3) si había un hospital, centro de salud y dispensario dentro de un radio de 30 km. del conglomerado muestral en el año *i*, (4) la tasa de mortalidad infantil a nivel de distrito para el año *i*, (5) si los servicios de PF esta-

ban disponibles en los hospitales, centros de salud y dispensarios ubicados en un radio de 30 km en el año *i*, (6) la "edad" de los servicios de PF en cada hospital, centro de salud y dispensario en el año *i*, y (7) si los servicios de PF estaban disponibles en los hospitales, centros de salud y dispensarios ubicados dentro de una distancia de 30 km cuando la entrevistada tenía 12 años.

Puesto que los dispensarios y, en menor medida, los centros de salud han sido establecidos deliberadamente en áreas con alta mortalidad infantil durante los últimos 10-15 años, fue crucial considerar en el análisis la ubicación no aleatoria de programas. Por ello, tres ecuaciones distintas fueron estimadas en las cuales la variable de resultado fue si había o no un hospital, centro de salud y dispensario (respectivamente) ubicado en un radio de 30 km del conglomerado en el año *i*. Las variables predictoras en estas ecuaciones eran: (1) la tasa de mortalidad infantil a nivel de distrito en el año *i*, (2) si fueron ofrecidos servicios de PF en otros establecimientos ubicados dentro de una distancia de 30 km. en el año *i*, (3) los gastos del gobierno en salud en el año *i*, y (4) la población del distrito en el año *i*. Estas ecuaciones fueron luego estimadas simultáneamente con la ecuación de resultado (o sea, anticoncepción) para estimar el impacto del programa de PF de Tanzania mientras se controlaban otros factores observados y no observados. En el análisis se usó un procedimiento de estimación de máximo-verosimilitud con información completa.

Resultados

El impacto estimado del esfuerzo de PF durante el periodo 1969-1991 se resume a continuación en forma de simulaciones de políticas.

	Tasa anual de fecundidad	Promedio de hijos tenidos	Porcentaje escenario 1
Valores reales (observados)	0.181	4.16	.95
Escenario I: Ningún servicio de PF	0.189	4.35	1.00
Escenario II: PF solo en los hospitales	0.158	3.63	0.83
Escenario III: PF solo en los centros de salud	0.165	3.78	0.87
Escenario IV: PF solo en los dispensarios	0.166	3.82	0.88
Escenario V: PF en todo establecimiento	0.120	2.76	0.63

Interpretación

Si la PF ha estado disponible continuamente en los tres tipos de establecimiento durante el período de estudio (con todos los otros factores constantes en los niveles observados), las probabilidades anuales

de nacimiento y el número medio de hijos por mujer hubiera sido un 37% menos que el observado para este período (0.120 frente a 0.181).

Fuente: Angeles et al., 1995.

ideas fundamentales del modelo de corte transversal al caso en que las observaciones son obtenidas para la misma muestra de individuos o comunidades en dos o más fechas. Los conjuntos de datos en que la misma muestra de individuos reciben seguimiento a través del tiempo son relativamente escasos. Sin embargo, si las comunidades o conglomerados se usan como unidades de análisis, muchos de los datos requeridos pueden obtenerse de DHS sucesivas que incluyen Módulos de Disponibilidad del Servicio realizadas en los mismos conglomerados muestrales³¹. Este diseño considera la pregunta: ¿Son las comunidades con los mayores cambios en el entorno de la oferta de planificación familiar entre dos fechas las que muestran también el mayor cambio en uso de anticonceptivos, controlando los cambios en otros factores?

El modelo básico puede expresarse de la siguiente forma:

Cambios en la variable de resultado =

Cambios en factores del programa +
Cambios en factores individuales +
Cambios en factores comunales +/-
Error

Una ventaja clave del diseño tipo panel es que permite el uso de un procedimiento de estimación particular referido en la literatura como estimador de "efectos fijos". En este enfoque, las variables o factores se dividen en dos categorías:

- aquellos que varían durante el transcurso de la evaluación, o factores variantes en el tiempo, y
- aquellos que no varían durante el transcurso de la evaluación, o factores persistentes

³¹ Por ejemplo, rondas sucesivas de la DHS han sido realizadas en los mismos conglomerados muestrales en Marruecos (1987, 1992, y 1995) y Tanzania (1991 y 1994), aunque en Marruecos en 1995 las mismas mujeres entrevistadas en 1992, lo que constituye un verdadero panel de mujeres.

³² El modelo está centrado en la medición de la respuesta de la población a los cambios en el programa durante el mismo período considerado en la evaluación. Las decisiones de asignación de recursos y las inversiones del programa hechas anteriormente se supone que no son causas de los cambios en los resultados durante el período de la evaluación excepto a través de respuestas diferidas o retardadas.

en el tiempo o fijos.

Los factores no observados que pueden haber influenciado la ubicación de los recursos del programa antes del estudio son tratados como "fijos" en el modelo y son "diferenciados" de las ecuaciones de regresión, reduciendo, así, el riesgo de incurrir en sesgos de estimación³². Este procedimiento de "diferenciación" se presenta en el Apéndice A, donde aparece el modelo panel de efectos fijos expresado en formato de regresión.

Gráfica IV-10

Ejemplo de diseño de evaluación de impacto con el modelo panel de efectos fijos, Tanzania

El diseño panel multi-nivel se usará para medir el impacto de los esfuerzos del Programa Nacional de Planificación Familiar (PPFN) durante el período 1991-96 en Tanzania. El período cubierto por la evaluación del impacto coincide aproximadamente con el período del Proyecto de Servicios de Apoyo a la Planificación Familiar (PF-SA) de la USAID en Tanzania (1990-97). Una Encuesta Demográfica y de Salud fue realizada en 1991/2 y otra está planeada para 1996. Una DHS más pequeña o de "medio período" (conocida como la Encuesta de Conocimiento, Actitudes y Prácticas de Tanzania o TKAP) fue llevada a cabo en 1994, en parte para compensar las escasas estadísticas de servicio disponibles. Las tres rondas de la encuesta (1991/2, 1994, y 1996) han sido o serán realizadas en los mismos conglomerados muestrales e incluirán Módulos de Disponibilidad del Servicio para proveer medidas del esfuerzo y la actividad del programa (y los cambios correspondientes) a nivel de conglomerado.

El análisis se centrará en la valoración de la naturaleza y la magnitud de los cambios en la entrega del servicio a nivel de la comunidad (o conglomerado) durante 1991-96 y la influencia de estos cambios sobre la conducta anticonceptiva, los niveles de fecundidad y otros resultados de interés, controlando los efectos de los cambios en otros factores.

La DHS 1991/92 estimó una necesidad no satisfecha de planificación familiar del 30% de las mujeres en unión. En los últimos años, la PF-SA ha invertido mucho en capacitación, así como en suministros y logística. Por eso, se pronostica que en las encuestas siguientes se observarán grandes efectos en las variables de "entorno de la oferta", tales como el "personal capacitado", el "número de agotamiento de existencias en los últimos seis meses," y el "número de métodos anticonceptivos ofrecidos en el establecimiento." Como hipótesis se plantea que los aumentos en el uso de anticonceptivos y las disminuciones en la necesidad no satisfecha de planificación familiar será mayor en áreas donde se han dado los principales avances en el entorno de la oferta de planificación familiar.

La lógica del diseño de tipo panel de niveles múltiples y los tipos de datos requeridos para su aplicación se presentan para el caso de Tanzania en la Gráfica IV-10.

Datos requeridos

Los datos requeridos para los modelos multinivel son bastante exigentes. Para construir sólidos modelos de corte transversal de resultados de programas de planificación familiar se requieren los datos siguientes:

- Datos a nivel de hogares e individuales (por ejemplo, de las encuestas tipo DHS) sobre:
 - ∨ características demográficas y económicas;
 - ∨ preferencias e intenciones de fecundidad; y
 - ∨ uso actual de anticonceptivos, fecundidad y otras medidas de "resultados".
- Información sobre determinantes de fecundidad y demanda de fecundidad en las comunidades (usualmente se obtienen de las encuestas a comunidades):
 - ∨ condiciones del mercado laboral y nivel de salarios;
 - ∨ infraestructura de las comunidades, e
 - ∨ indicadores demográficos agregados de años anteriores (ver abajo para mayores detalles).
- Información sobre el entorno de la oferta de servicios de planificación familiar, usualmente obtenida de encuestas o censos de establecimientos como el Módulo de Disponibilidad del Servicio de la DHS, estadísticas del programa y/o encuestas a comunidades³³:
 - ∨ número y tipos de establecimientos de salud y planificación familiar dentro de una distancia fija para cada conglomerado muestral (por ejemplo, 30 km);
 - ∨ servicios y métodos anticonceptivos disponibles;
 - ∨ período de tiempo durante el cual los servicios y los métodos han estado disponibles; y
 - ∨ medidas de calidad del servicio.

En los modelos de efectos aleatorios es especialmente importante la disponibilidad de información sobre características a nivel de comunidad que puedan haber sido determinantes importantes de decisiones anteriores

de ubicación-asignación de recursos del programa (por ejemplo, niveles de fecundidad y mortalidad, indicadores de estándar de vida, etc.).

Pocas (si alguna) encuestas recogen todos los datos requeridos para aplicaciones importantes de este enfoque. Sin embargo, las Encuestas Demográficas y de Salud que incluyen un Módulo de Disponibilidad del Servicio (SAM) generalmente contienen mucha de la información necesaria. Esta puede ser complementada a un costo marginal relativamente bajo con los datos económicos adicionales requeridos del programa y de las comunidades, para proveer una base de datos adecuada para la estimación de modelos multinivel. Otras encuestas con información comparable también pueden ser utilizadas.

Los datos requeridos para aplicar el modelo panel de efectos fijos son, en buena medida, los mismos que para el modelo de corte transversal, exceptuando que se necesitan datos en dos o más fechas sobre los mismos individuos o los mismos conglomerados. Idealmente, la recolección de datos de la DHS en un país estaría programada para coincidir con los ciclos de actividades de los programas de planificación familiar (por ejemplo, encuestas al principio y al final de un ciclo de 5 a 7 años). De esta forma, las encuestas de propósito múltiple como las DHS servirían como mecanismo básico para recoger la información requerida para medir el impacto durante un determinado ciclo del programa. Sin embargo, no es estrictamente necesario que la programación de las rondas de la DHS corresponda exactamente a los ciclos del programa. En la medida que se acumulan las rondas, el período cubierto por el panel de multinivel podría extenderse para comparar los efectos del programa en las diferentes etapas de su implementación. Además, los análisis multinivel de corte transversal realizados en distintas fechas pueden usarse para mostrar que los diferentes insumos son importantes para lograr impacto en las distintas etapas de desarrollo del programa (efectos que podrían permanecer ocultos al agregar las distintas fases del programa).

³³ La encuesta de análisis de situación del "Population Council" es otro ejemplo del tipo de encuesta de establecimientos que puede usarse para recoger una buena parte de la información requerida a nivel del programa.

Ventajas

El enfoque de regresión multinivel tiene varias ventajas:

- Puesto que el enfoque relaciona los insumos del programa con los resultados a nivel de comunidad, permite la medición del impacto del programa como se implementa en la realidad.
- No requiere diseño experimental.
- Ofrece información más detallada que cualquier otro enfoque sobre las trayectorias seguidas por los programas para influir en la conducta anticonceptiva.

Limitaciones y consideraciones prácticas

Existen varias limitaciones y ataduras importantes en el uso de modelos de regresión multinivel:

- El enfoque es exigente en términos de datos. Se requieren encuestas poblacionales a gran escala como las DHS. Esto es especialmente cierto cuando se deben estimar modelos de panel, pues la unidad de análisis para estos análisis son las comunidades o los conglomerados. A no ser que una muestra suficientemente grande de conglomerados esté disponible, el diseño puede carecer de suficiente poder estadístico para detectar efectos del programa de la magnitud que probablemente se darían en períodos relativamente cortos (por ejemplo, cinco años). Para el modelo de corte transversal, la disponibilidad de datos a nivel de comunidad sobre factores que pueden haber influenciado las decisiones anteriores de ubicación del programa es fundamental si se desean obtener estimaciones insesgadas de impacto del programa. También se requieren medidas de la actividad del programa, las cuales suelen ser derivadas de las encuestas a establecimientos.
- El método es sensible al uso de modelos estadísticos apropiados y al tratamiento correcto de los problemas de estimación estadística. Para realizar análisis multinivel también se necesita personal debidamente capacitado y software apropiado.
- El desarrollo y la estimación de modelos estadísticos requiere conocimiento del entorno. Debido a que las variables específicas

incluidas en los modelos variarán en cada caso, no hay un conjunto estándar de variables que se puedan aplicar. Se requiere conocer el escenario social y la evolución de los servicios de salud y de planificación familiar para orientar la selección de variables apropiadas a nivel de comunidad y de programa.

- Los modelos son sensibles a las fechas de las inversiones en el programa en comparación con las fechas de observación del estudio de impacto. Puesto que algunas especificaciones del modelo de efectos aleatorios relacionan los indicadores del programa de planificación familiar con medidas de resultado recientes (ver Gráfica IV-9 para una ilustración de este punto), es fundamental contar con información precisa sobre las fechas de los cambios ocurridos en el programa para estimar con precisión su impacto. De igual manera, puesto que el modelo de efectos fijos relaciona los cambios en las variables del programa a los cambios en los indicadores de resultado durante intervalos de tiempo específicos, la ausencia de cambio en las mediciones del programa durante un determinado período de estudio lógicamente resultará en estimaciones minimalistas del impacto del programa³⁴. El modelo de efectos fijos tampoco considera las inversiones del programa hechas antes del período en estudio. Por ejemplo, si un país ha hecho grandes inversiones en planificación familiar en los setenta y principios de los ochenta y ha mantenido un fuerte apoyo a estas actividades de ahí en adelante, el país podría mostrar niveles bastante altos de prevalencia anticonceptiva. Pero si uno fuera a aplicar el modelo de efectos fijos con datos de fines de los ochenta y principios de los noventa, el análisis podría mostrar relativamente poco impacto, debido a que la fuerte inversión de años anteriores no estaría reflejada en las medidas de cambio del programa. Este fenómeno podría explicar, en parte, el moderado impacto de programa encontrado en el análisis Gertler y Molyneaux (1994) del programa de planificación familiar de Indonesia de mediados de los ochenta. Por eso, el modelo de efec-

³⁴ Este es un tema genérico sobre diseños de evaluación; un problema semejante surge al tratar de medir el resultado de un experimento aleatorio demasiado pronto después de la implementación del programa. Se destaca aquí, en relación al modelo de efectos fijos, para enfatizar su importancia al interpretar los resultados de los estudios de impacto basados en este enfoque.

tos fijos debe ser aplicado con cuidado en países con programas relativamente avanzados, donde las inversiones recientes del programa pueden ser relativamente modestas³⁵.

de microcomputadoras poderosas y software apropiado y el número creciente de investigadores con las destrezas metodológicas requeridas para aplicar modelos multinivel.

A pesar de estas limitaciones, el enfoque de regresión de multinivel es una metodología factible y potencialmente poderosa para valorar el impacto de programas, dada la disponibilidad de datos tipo DHS en forma recurrente en muchos países, la creciente proliferación

³⁵ Una solución a este problema es incluir efectos retardados "lags" en el modelo de efectos fijos. Sin embargo, esto es factible solo si hay tres o más rondas de datos de la encuesta de hogares con mediciones de las variables del programa.

Gráfica IV-11

Resumen de los enfoques preferidos para medir el impacto de los programas de planificación familiar

Enfoque	Observaciones
Experimentos aleatorios	La "norma de oro", pero generalmente impráctico debido a las dificultades para establecer y mantener controladas las condiciones experimentales en los estudios a nivel nacional.
Cuasi-experimentos	Más prácticos puesto que no requieren la asignación aleatoria. Sin embargo, son más vulnerables a sesgos de selección y otras amenazas a la validez que los experimentos aleatorios. El uso de di-seños cuasi-experimentales junto con métodos de control para la heterogeneidad no observada es muy prometedor.
Métodos de regresión multinivel:	
Modelo transversal básico	El menos exigente entre los enfoques no experimentales preferidos. Comprende la valoración de la relación entre las variables que miden la presencia/actividades y los resultados de los programas familiares basados en datos de corte transversal de hogares y establecimientos, recogidos en las mismas áreas geográficas (por ejemplo, los conglomerados muestrales DHS). La limitación clave - el peligro de estimaciones sesgadas, si los programas no son implementados en forma uniforme o aleatoria en las áreas geográficas.
Efectos aleatorios multi-ecuacionales	Refinamiento del modelo de corte transversal básico que provee una forma de tomar en cuenta la posible endogeneidad de la ubicación del programa. Requiere datos adicionales sobre factores o variables que podrían estar correlacionadas con las decisiones de ubicación del programa. Los procedimientos estadísticos de estimación son complejos.
Modelo panel de efectos fijos	Si se cuenta con datos de hogares y de establecimientos de las mismas áreas geográficas en dos o más fechas, ofrece un método alternativo (y más fácil de calcular) para considerar el problema de la endogeneidad de la ubicación del programa. Debido a que los conglomerados muestrales son las unidades de análisis, generalmente tiene menos poder estadístico que otros modelos.

ENFOQUES ALTERNATIVOS

¿Cuáles opciones están disponibles cuando es imposible implementar alguno de los tres enfoques preferidos de la sección anterior? En esta sección, se presentan dos alternativas, que a pesar de ser menos preferidas, tienen cierta utilidad:

- una variante del método de descomposición y
- el método de prevalencia.

El enfoque de descomposición requiere la recolección de datos en dos o más fechas y el método de prevalencia es del tipo de corte transversal.

Descomposición (Modelo de determinantes próximos)

Descripción

Estrictamente hablando, la descomposición es una técnica genérica de análisis demográfico o epidemiológico, en contraste con un método diseñado específicamente para propósitos de evaluación del impacto de programas. Aquí, nos centramos en un método específico de descomposición, el enfoque propuesto por Bongaarts y Kirmeyer (1980), el cual cuenta con varias características que lo convierten en una herramienta más valiosa para valorar el impacto de programas que los métodos genéricos de descomposición.

Al aplicarse en la medición de impacto de programas de planificación familiar, el objetivo de la descomposición es determinar qué porción del cambio observado en las tasas de fecundidad (por ejemplo, la tasa global de fecundidad) entre dos fechas es atribuible a cambios en distintos factores. El método considerado, denominado de aquí en adelante método de descomposición de los determinantes próximos (pues se basa en el muy conocido modelo de determinantes próximos de la fecundidad - ver Bongaarts, 1978), descompone los cambios en la TGF en cambios en: (1) proporciones de casadas o en unión, (2) duración promedio de infertilidad postparto, (3) prevalencia anticonceptiva, y (4) tasas de aborto³⁶. El impacto del programa está indicado por una reducción en los niveles de fecundidad durante un período particular estudiado, una gran porción del cual es atribuible a los aumentos en la prevalencia anti-

conceptiva del "programa"³⁷. Aunque no se contempla en el modelo original, las estimaciones de anticoncepción del programa pueden obtenerse usando datos de encuestas sobre fuentes de oferta o estadísticas de servicio para estimar qué porción del uso total de anticonceptivos es abastecida por el programa en contraste con fuentes ajenas al programa.

Diseño y análisis

La fuente de datos usual para la aplicación del método es dos encuestas de población consecutivas en una determinada población³⁸. Aparte de la disponibilidad de datos de dos o más fechas y la información sobre la fuente de la anticoncepción, no hay requisitos de diseño especiales para usar el método.

La medición del impacto del programa proviene de la descomposición de la fecundidad en componentes, como proponen Bongaarts y Kirmeyer (1980) y Bongaarts y Potter (1983). El modelo elemental es un modelo multiplicativo, en el cual los factores considerados están expresados como índices de sus efectos inhibidores en la fecundidad; o sea, como medidas del grado en que cada factor contribuye a la diferencia entre la fertilidad total y la fecundidad total. El modelo se presenta gráficamente en la Gráfica IV-12.

En un corte transversal, el modelo puede expresarse como:

³⁶ Los efectos de variación en factores como efectividad de uso de anticonceptivos y esterilidad se pueden también incluir en el modelo, pero los datos específicos del país de estas variables rara vez están disponibles. Jolly y Gribble (1993) han propuesto que un factor adicional, la fecundidad que ocurre fuera de matrimonios o uniones, se incluya en el modelo para poblaciones de África sub-Sahariana.

³⁷ La prevalencia anticonceptiva del programa se refiere al uso de anticonceptivos obtenidos de servicios y fuentes de oferta del programa (contrario a las fuentes no programáticas). Las fuentes no programáticas son los sectores comerciales y privados y varía en importancia relativa de país a país.

³⁸ Los censos de población también proporcionan los datos requeridos sobre niveles de fecundidad, composición por edad y proporción en matrimonios, pero no sobre otros determinantes próximos de la fecundidad como la lactancia.

$$TGF = TF * C_m * C_c * C_i * C_a$$

Donde:

TGF = tasa global de fecundidad observada;

TF = tasa total de fertilidad (no observable);

C_m = índice de matrimonios o uniones;

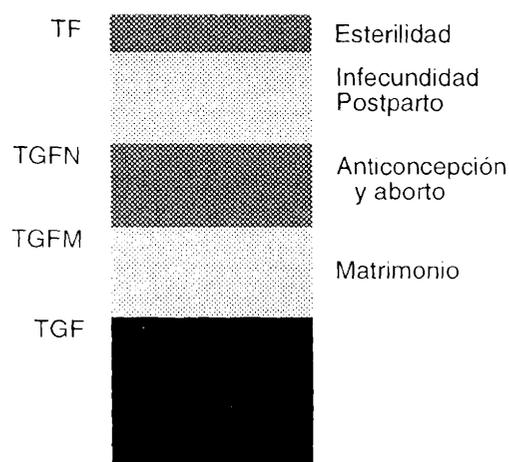
C_c = índice de anticoncepción;

C_i = índice de infertilidad post-parto; y

C_a = índice de aborto.

Gráfica IV-12

Efecto de los Determinantes Próximos sobre medidas seleccionadas fecundidad



Fuente: Bongaarts & Kirmeyer, 1980

Si se aplica a rondas sucesivas de encuestas, los cambios en cada índice se ligan al cambio observado en la TGF para valorar la contribución de cada determinante al cambio observado en la fecundidad. Por ejemplo, la contribución proporcional de cambios en la anticoncepción al cambio observado en la TGF se calcula de la siguiente forma:

$$C = (\ln C_{c,2} - \ln C_{c,1}) / (\ln TGF_2 - \ln TGF_1)$$

Donde:

\ln = logaritmo natural;

$C_{c,2}$ y $C_{c,1}$ = índices de anticoncepción en las rondas de encuestas 2 y 1, respectivamente; y

TGF_2 y TGF_1 = tasas globales de fecundidad observadas en las rondas de encuestas 2 y 1, respectivamente.

Como se observó anteriormente, la determinación del impacto de programas usando este método requiere tener una estimación de la porción del uso de anticonceptivos que se origina en fuentes programáticas, la cual puede obtenerse de datos de encuestas sobre la fuente del anticonceptivo usado o de estadísticas de servicio. En la Gráfica IV-13 se presentan ejemplos de resultados de la aplicación del método a encuestas sucesivas de las Filipinas.

Ventajas

- El método requiere datos fácilmente disponibles en las encuestas tipo DHS.
- Los cálculos son simples.
- Es efectivo para controlar los efectos de los cambios en los determinantes próximos de la fecundidad.

Limitaciones y consideraciones prácticas

- Debido a las fluctuaciones anuales en los niveles de fecundidad, el método es sensible a los años que cubre el estudio en particular; i.e., se pueden obtener estimaciones muy diferentes variando los años de cobertura.
- Está limitado a la medición del impacto sobre la fecundidad exclusivamente.
- No ofrece medidas directas de los efectos de los insumos del programa; los insumos del programa se infieren de los cambios en la prevalencia anticonceptiva (y de estimaciones de las contribuciones del programa a los cambios en la prevalencia).
- Si se carece de datos específicos del país sobre factores como efectividad del uso de anticonceptivos, el método se apoya en modelos que pueden no describir apropiadamente el país en cuestión.
- Un estudio (Reinis, 1992) sugiere que el

método puede no ser válido en países donde las mujeres usan anticonceptivos principalmente para finalizar la procreación una vez alcanzado el tamaño familiar deseado (en vez de para espaciar los embarazos), donde es común retardar el matrimonio y donde el uso de anticonceptivos es más prevalente a edades avanzadas. Este tema requiere más investigación.

- El método es sensible a: (1) la precisión de los datos de las encuestas sobre las fuentes de anticonceptivos y (2) las definiciones y respuestas (en las entrevistas de las encuestas) de la anticoncepción originada dentro y fuera del programa³⁹.
- El método ofrece solo una medida "bruta" de impacto; o sea, no considera los efectos de sustitución (los aumentos en la anticoncepción programática que se dan a costa de un menor uso del sector privado) y los efectos catalíticos del programa (los aumentos en la anticoncepción no programática que resultan de esfuerzos de promoción del programa).

Método de prevalencia

Descripción

El último método considerado, el método de prevalencia, provee una estimación burda del impacto de programas cuando la recolección de datos de dos o más fechas y la realización de experimentos aleatorizados "postest solamente" no son factibles. El método de prevalencia es un método de corte transversal diseñado para aprovechar la amplia disponibilidad de datos de encuestas sobre prevalencia anticonceptiva en países en desarrollo. Usando datos de encuestas sobre prevalencia anti-

ceptiva según fuente de abastecimiento (o sea, programática frente a no programática), tasas por edad de fecundidad actual y determinantes próximos de la fecundidad⁴⁰, el método estima la porción de la diferencia entre la fecundidad potencial⁴¹ y la fecundidad observada, que puede ser atribuida a la anticoncepción programática. A su vez, esto puede ser convertido a dos estimaciones de impacto de programas: (a) la reducción en las tasas de fecundidad y (b) el número de hijos evitados durante un intervalo de tiempo especificado (ge-

Gráfica IV-13

Ejemplo de una aplicación del método de descomposición de determinantes próximos a encuestas sucesivas en las Filipinas

Encuesta y fecha de referencia de las estimaciones	TGF	C _m	C _c	C _i	TF
Encuesta de fecundidad 1978 (1973-1977)	5.60	0.599	0.778	0.761	15.77
Encuesta demográfica 1982 (1978-1982)	5.28	0.599	0.713	0.778	15.90
Cambio en la TGF Cambio en puntos porcentuales de la TGF aportado por cambios en:					
Absoluto	Pct.	Nupcialidad	Anticoncepción*	Infertilidad	Residual
-0.32	-5.6	0.0	-8.5	+2.1	+0.8

Interpretación Durante el periodo cubierto por las encuestas, la fecundidad total bajó en un 5.6 por ciento. La caída en la TGF se explica completamente por un aumento en el uso de anticonceptivos. De hecho, si sólo el uso de anticonceptivos hubiera cambiado en este periodo, la TGF habría bajado 8.5 puntos porcentuales. Sin embargo, los niveles cambiantes de la infertilidad post-parto como resultado de una caída en la duración de la lactancia ejercieron una influencia contraria sobre la TGF de aumento en 2.1 puntos porcentuales. Los factores no especificados o residuales también contribuyeron a un aumento en la fecundidad total de 0.8 puntos. La contribución de los cambios en las tasas de aborto no pudo ser valorada en este estudio, debido a la falta de datos apropiados. Como los datos de la encuesta indican que el sector público proveyó aproximadamente el 50% de los servicios y suministros de anticonceptivos durante este periodo, puede atribuirse una reducción en la TGF de aproximadamente 4% al programa de PF del sector público.

Fuente: Casterline et al., 1988.

³⁹ Estos temas han sido unos de los menos confiables de las DHS. En las encuestas DHS-III, las preguntas sobre este tema se formulan de tal manera que se espera que obtengan información más confiable.

⁴⁰ Concretamente, datos sobre proporciones de mujeres casadas o en unión y la duración promedio del periodo de insusceptibilidad post-parto (a menudo estimada de la duración media de la lactancia materna). La información sobre tasas de aborto también puede usarse cuando está disponible.

⁴¹ La fecundidad potencial se define como el nivel de fecundidad que prevalecería en una población determinada en ausencia de anticoncepción.

neralmente un año) como resultado de la anticoncepción programática. El método se basa en el mismo modelo de la relación cuantitativa entre la fecundidad y sus determinantes próximos descrito anteriormente en el método de descomposición (ver Bongaarts, 1986, para una presentación completa del método).

Diseño y análisis

El método de prevalencia no requiere más datos que los normalmente recogidos como parte de las encuestas tipo DHS. Los datos básicos que se necesitan son:

- estimaciones de la prevalencia anticonceptiva para una determinada fecha, por grupos quinquenales y fuente de abastecimiento;
- tasas de fecundidad por edad para un determinado período de referencia;
- número de mujeres en edad reproductiva, en grupos de edad quinquenales; y
- tamaño total de la población.

Las estimaciones de efectividad de uso específicas del país, idealmente por método y fuente, y las tasas de fertilidad específicas por edad también son útiles, pero rara vez están disponibles. Si estos datos no están disponibles para un determinado país, se pueden usar tablas estándar (Bongaarts, 1986).

Con este método, el impacto de programas se mide por la porción de la diferencia entre la fecundidad potencial y la fecundidad real durante un período de referencia específico (por ejemplo, 1a 3 años antes de la encuesta) que es explicada por la anticoncepción programática. Este método produce dos medidas de impacto: (1) la reducción en la fecundidad (o sea, fecundidad potencial menos fecundidad observada) atribuible a la anticoncepción programática y no programática, respectivamente, y (2) los embarazos evitados por la anticoncepción programática y no programática.

Como ilustración, la Gráfica IV-14 muestra estimaciones para varios años a mediados de la década de los setenta en que se usó este método en varios países⁴² (otros ejemplos se ofrecen en United Nations, 1986). La Gráfica muestra los efectos sobre las tasas brutas de natalidad y los embarazos evitados debido a la anticoncepción programática y no progra-

mática.

Obsérvese que aunque el método tiene un diseño de corte transversal, las estimaciones para encuestas sucesivas proveen una serie de tiempo de estimaciones que podrían ser usadas para estimar las tendencias en el desempeño del programa a través del tiempo. Sin embargo, con datos de encuestas sucesivas, el método de descomposición de determinantes próximos es más informativo.

Ventajas

- No requiere que se lleven a cabo estudios especiales. Si se puede suponer que las tablas modelo de efectividad de uso y de fertilidad son aplicables a la población en estudio, el método requiere solo datos que están disponibles en las encuestas tipo DHS.
- Los cálculos son relativamente simples.

Limitaciones y consideraciones prácticas

- Si los datos sobre efectividad de uso específicos del país (para la anticoncepción programática y no programática) y las proporciones por edad de mujeres fértiles (normalmente no recolectadas en las encuestas tipo DHS) no están disponibles, el método requiere el supuesto de que las tablas modelo respectivas son aplicables a este país.
- No mide directamente los efectos de los insumos de programas; los insumos de programas son inferidos de los cambios en la prevalencia anticonceptiva (y estimaciones de las contribuciones del programa a los cambios en la prevalencia).
- El método es sensible a: (1) la precisión de los datos de las encuestas sobre la fuente de anticoncepción y (2) las definiciones y respuestas (en las entrevistas de las encuestas) de la anticoncepción programática y no programática.
- El método ofrece solo una medida "bruta" de impacto; sin considerar los efectos de

⁴² Los cálculos requeridos para aplicar el método son laboriosos aunque relativamente simples. La Figura IV-II ilustra el producto que resulta de usar el método para valorar la magnitud del impacto del programa. Para mayores detalles de cálculo se refiere al lector interesado a Bongaarts (1986).

Gráfica IV-14

Ejemplos de resultados de la aplicación del método de prevalencia, países selectos

País o área	Año de referencia	Población POP (millones)	Tasa de natalidad (TBN) (por mil)	Efecto crudo en la TBN		Nacimientos evitados (NE)	
				Anticoncepción programática PTBN-TBN	Anticoncepción no programática NTBN-PTBN	Anticoncepción programática NEP (miles)	Anticoncepción no programática NEN (miles)
África							
Gana	1978	11.37	46	1.7	0.0	19.5	0.0
Mauricio	1979	0.94	27	23.2	1.4	21.8	1.3
Tunisia	1977	6.01	35	6.5	1.1	39.2	6.9
Asia							
Hong Kong	1975	4.40	18	9.6	9.3	42.4	41.0
Indonesia	1979	148.09	33	9.6	2.0	1428.4	297.6
Malasia	1979	13.67	31	11.2	3.7	153.1	50.6
Filipinas	1979	47.68	32	6.9	9.1	329.4	432.3
Corea	1975	35.67	26	9.1	2.4	324.7	84.2
Singapur	1978	8.79	17	22.5	7.6	197.4	67.1
Tailandia	1975	42.42	34	7.7	3.2	325.8	137.2
Latinoamérica							
Rep. Dominicana	1976	5.14	38	5.7	4.8	29.1	24.7
El Salvador	1976	4.26	42	3.3	7.1	14.0	30.1
México	1978	64.09	36	9.6	10.6	616.5	681.4
Paraguay	1977	2.97	37	3.1	3.1	9.2	9.2
Oceania							
Fiji	1978	0.61	34	15.8	1.9	9.6	1.1

Interpretación para un país selecto - Indonesia: En ausencia del programa nacional de planificación familiar en Indonesia, la tasa bruta de natalidad (por mil personas) habría sido de 9.6 nacimientos por encima de la TBN de 33 registrada en 1979. La anticoncepción no programática redujo la TBN en otros 2 puntos. Se estima que más de 1,4 millones de nacimientos fueron evitados en 1979 debido a la anticoncepción programática y casi 300.000 más fueron evitados debido a la anticoncepción no programática.

Fuente: Bongaarts, J., 1986: Tabla 9.

sustitución (los aumentos en la anticoncepción programática que se dan a costa de un menor uso del sector privado) y los efectos catalíticos del programa (los aumentos en la anticoncepción no programática que resultan de esfuerzos de promoción del programa).

- El método se limita a medir el impacto en términos de la fecundidad.

RESUMEN

En este capítulo, los tres métodos preferidos y las dos "segundas alternativas" mejores para medir el impacto de los programas de planificación familiar fueron revisados y valorados críticamente. La Gráfica IV-15 resume las ventajas y limitaciones relativas de los enfoques considerados.

En una visión de conjunto, los métodos recomendados pertenecen a dos categorías amplias: los métodos experimentales y los métodos basados en encuestas⁴³. La confianza cada vez mayor en los métodos basados en encuestas para la evaluación de programas se debe a dos factores: (1) el hecho de que el impacto a nivel poblacional se mide mejor con datos a nivel poblacional, y (2) la disponibilidad relativamente grande de datos de encuestas grandes a nivel poblacional como la DHS. El valor de los datos de las encuestas para la evaluación de programas es mayor cuan-

do los esfuerzos de recolección de datos por encuestas están estratégicamente calendariados respecto a los ciclos de los programas y están complementados con datos del programa a nivel comunitario para las mismas unidades geográficas que los datos a nivel individual.

Al mismo tiempo, los experimentos (aleatorizados y cuasi-experimentos) tienen su rol en mejorar la validez de las conclusiones de la evaluación, particularmente en evaluaciones de áreas funcionales de los programas de planificación familiar (i.e., investigación de operaciones).

Idealmente, diferentes métodos deben usarse para contestar las distintas preguntas, como parte de las evaluaciones integrales de programas. Sin embargo, la incorporación de los diferentes componentes del diseño de evaluación requiere un planeamiento cuidadoso en la etapa de diseño de programas y nuevos ciclos de programas. Si hay un fuerte compromiso para medir el impacto, los administradores de programa pueden preferir modificar la forma de implementación de las actividades para ajustarse a una medición significativa del impacto del programa.

⁴³ Debe observarse que las encuestas también suelen ser utilizadas como medio de recolección de datos en los estudios experimentales.

Gráfica IV-15

Resumen de los métodos para medir impacto del programa

Característica ^a	Regresión Multinivel					Método prevalencia
	Experimento aleatorizado	Cuasi Experimentos	Encuesta única	Panel	Descomposición	
Amenazas a la validez	Bajo	Moderado	Moderado	Moderado	Moderado	Moderado
Supuestos requeridos	Poco	Moderado	Moderado	Moderado	Muchos	Muchos
Capacidad de aislar efecto de programa	Fuerte	Moderado	Moderado	Moderado	Moderado	Moderado
Volumen de datos requeridos	Bajo	Moderado	Alto	Alto	Moderado	Bajo
Informa mecanismos causales	Moderado	Moderado	Fuerte	Fuerte	Débil	Débil
Indicadores de resultado usados	Flexible	Flexible	Flexible	Flexible	Fecundidad solamente	Fecundidad solamente
Grado de control del programa	Alto	Alto	Ningún	Ningún	Ningún	Ningún
Factibilidad si el programa prefiere ciertas áreas	Bajo	Bajo	Moderado ^b	Moderado	Alto	Alto

a- Las metodologías no comparan los costos dado que estos varían mucho según las circunstancias

b- Factibilidad depende de si los factores subyacentes en las decisiones del programa pueden incorporarse a las ecuaciones de regresión como variables de control.

Capítulo V

Desarrollo de un plan de acción

- Definición de las instituciones y las personas responsables de la evaluación
- Establecimiento de un cronograma para actividades específicas
- Presupuesto de la evaluación

DESARROLLO DE UN PLAN DE ACCIÓN

Los capítulos anteriores de este manual hacen énfasis en los aspectos técnicos del diseño de una evaluación. Sin embargo, un plan de evaluación será de poco valor si no hay un plan claramente definido para su implementación. Este capítulo indica los aspectos que debe contemplar el desarrollo de un plan de acción para el seguimiento de programas o para la valoración del impacto.

DEFINICIÓN DE LAS INSTITUCIONES Y LAS PERSONAS RESPONSABLES DE LA EVALUACIÓN

Generalmente, una institución toma el liderazgo en el diseño e implementación de la evaluación de un programa nacional de planificación familiar, aunque suele ser necesario considerar la ayuda de otras organizaciones de entrega de servicios o de investigación. En algunos casos el ímpetu para una evaluación a gran escala surge del mismo país; en otros, es un requisito del donante. Tal vez el escenario más común es que los administradores y donantes tengan un interés común de saber si el programa está en el curso correcto y cómo puede ser fortalecido.

La organización líder tiende a ser el principal proveedor de servicio en el país, especialmente si es una institución gubernamental (por ejemplo, el Ministerio de Salud). Alternativamente, la asociación privada de planificación familiar puede liderar, sobre todo si juega un papel importante en la entrega de servicios o cuenta con una buena capacidad para investigar o evaluar. Es importante identificar quién tiene la responsabilidad principal de la evaluación e involucrar a otros grupos de interés.

Involucrar a los principales grupos de interés en el proceso de planificación

Para maximizar el beneficio y la utilidad de la evaluación de un programa nacional de planificación familiar, es importante incluir a los principales grupos de interés en el proceso desde el principio. Los grupos de interés

son todas las organizaciones o individuos que podrían estar interesados en cómo se realiza la evaluación, cuáles resultados muestra y cómo se puede usar la información posteriormente. También, la lista debería incluir cualquier institución que pueda contribuir con datos. Tales grupos de interés potenciales son los siguientes:

- Oficinas gubernamentales responsables de dar seguimiento a las cuestiones poblacionales, especialmente en países que han definido metas demográficas:
 - √ el Ministerio de Planificación
 - √ el Consejo Nacional de Población
 - √ Otras
- Organizaciones que ofrecen servicios de planificación familiar, incluyendo:
 - √ El Ministerio de Salud Pública
 - √ El afiliado a la IPPF
 - √ Otras ONG de envergadura
 - √ Programas subsidiados de mercadeo social de anticonceptivos
 - √ Firmas del sector privado que venden anticonceptivos
 - √ Asociaciones de proveedores privados (asociación de gineco-obstetras, asociación de parteras)
- Agencias donantes que apoyan el programa
- Grupos de salud de las mujeres y otros grupos de abogacía

En última instancia, si los grupos de interés no perciben que los datos y el análisis son útiles para las decisiones que deben tomar sobre el diseño e ejecución del programa, los resultados pueden no ser utilizados nunca según el propósito con que fueron creados y serían ignorados o desacreditados por aquéllos para quienes se hace la evaluación.

Definición de las necesidades técnicas e identificación de las fuentes disponibles dentro del país

Una evaluación a gran escala de un programa de planificación familiar, especialmente si incluye valoración del impacto, requiere pericia técnica en el diseño de estudio, preparación de instrumentos de información, supervisión de la recolección de datos, edición y procesamiento de la información, análisis de datos y preparación del informe.

Un número cada vez mayor de países tienen ahora personal calificado en estas áreas. Bajo circunstancias ideales el personal técnico apropiado está dentro de las organizaciones líderes o auxiliares y estos individuos pueden estar disponibles para trabajar en la evaluación. Sin embargo, puede ocurrir que los individuos adecuados estén ya comprometidos con otras actividades durante el período de la evaluación o que el nivel adecuado de pericia técnica no esté disponible dentro de las organizaciones auxiliares. En este último caso, es importante considerar fuentes alternativas de asistencia técnica en las áreas señaladas anteriormente (diseño de estudios, recolección de datos, análisis, etc.), tales como:

- firmas de investigación locales, especialmente las que se especializan en la investigación social;
- universidades locales (por ejemplo, departamentos de demografía, escuelas de salud pública); y
- consultores privados con experiencia en investigación y evaluación.

En países con poca pericia técnica local en el campo de la investigación social, puede ser necesario complementar las destrezas de los investigadores locales con asistencia técnica externa.

Establecimiento y mantenimiento de canales de comunicación efectivos

Si se quiere que todos los participantes compartan la propiedad del producto final, es importante para la organización líder mantener un diálogo honesto con los principales grupos de interés, no simplemente esperar que los otros "le pongan el sello" de sus sugerencias.

Este proceso de consulta no termina al completar el análisis de datos. Más bien, es útil mantener este mismo nivel de comunica-

ciones a través de las etapas de diseminación y utilización de resultados. Los esfuerzos de una institución por aplicar los resultados específicos de la evaluación (por ejemplo, intensificar la entrega de métodos de acción prolongada en áreas identificadas por la evaluación como áreas insuficientemente servidas) pueden ser un ejemplo para otras instituciones sobre cómo se puede aplicar la investigación académica para mejorar los programas. Además, el sentido de propósito común desarrollado a través del proceso de evaluación puede servir para reforzar la colaboración en el área de la entrega de servicios, aun entre grupos que no trabajan juntos normalmente.

ESTABLECIMIENTO DE UN CRONOGRAMA PARA ACTIVIDADES ESPECÍFICAS

El plan de evaluación indicará una serie de actividades de recolección de datos correspondiente a los objetivos de la evaluación, que se hará en forma escalonada durante la duración del proyecto (ver Gráfica III-6). Para cada tipo de dato, es esencial especificar la frecuencia con la que debe ser recolectado y reportado. (Por ejemplo, las estadísticas rutinarias de servicio podrían ser recolectadas mensualmente, pero reportadas trimestralmente.

La información para cada tipo de recolección de datos puede luego ser resumida para ofrecer una visión de conjunto de esta actividad durante la duración del proyecto. Este ejercicio tiene la ventaja dual de indicar (1) los posibles conflictos de agenda entre las actividades de evaluación, especialmente cuando el personal de recolección de datos es limitado y (2) posibles conflictos con otras actividades de la organización (por ejemplo, la preparación del reporte anual) o de la comunidad principal (por ejemplo, períodos de días feriados cuando baja el ritmo de trabajo).

La Gráfica V-1 muestra un ejemplo de cronograma sintético para una recolección múltiple de datos.

PRESUPUESTO DE LA EVALUACIÓN

Es fundamental estimar los costos de la evaluación propuesta, a no ser que los planes de uno sean frustrados por la realidad financiera. Los pasos para el desarrollo del presupuesto son:

- Identificar los recursos para la evaluación en la organización líder (y si procede, en

Gráfica V-1

Ejemplo de un cronograma para actividades de evaluación

	Año 1	Año 2	Año 3	Año 4	Año 5
Desarrollo de un plan de evaluación	x				
Pasos					
Aclarar los objetivos del programa	x				
Describir cómo debe funcionar el programa (marco conceptual)	x				
Establecer objetivos de la evaluación	x				
Identificar los componentes que deben recibir seguimiento	x				
Definir indicadores relevantes	x				
Identificar fuentes de datos	x				
Desarrollar un plan de recolección y procesamiento de datos	x				
Diseñar un formato para presentar resultados	x				
Diseñar/implementar las actividades de seguimiento					
Estadísticas de servicio					
Revisar y mejorar el SIG de estadísticas de servicio de rutina	x				
Recolección e informe de las estadísticas rutinarias de servicio (X grande significa reporte anual)	x x X	x x x X	x x x X	x x x X	x x x X
Encuestas a establecimientos (ilustrativo)					
Ejemplo: Análisis de situación (o SAM)	x x				x x
Valoraciones cualitativas periódicas de la entrega de servicios (ejemplo ilustrativo)					
COPE		x x			
Grupos focales en usuarios de los servicios a adolescentes				x x	
Estudios especiales de áreas funcionales^a (ilustrativo)					
Seguimiento de 12 meses al personal capacitado en NORPLANT		x x			
Seguimiento del alcance y los efectos de la campaña IEC		x x			
Investigación operativa del impacto de mejorar la consejería sobre las tasas de continuación				x x x	
Diseño de la valoración del impacto					
Pruebas aleatorias (ejemplos ilustrativos)					
Ejemplo: introducción al NORPLANT®		x x x x	x x x x	x x x x	
Modelo de regresión multilevel (ver Capítulo IV)					
Recolección de datos DHS y SAM	x x				
Segunda ronda de recolección de datos DHS/SAM					x x
Análisis para medir el impacto					x x
Utilizar resultados de la evaluación para modificar y mejorar el programa					
	x x x	x x x x	x x x x	x x x x	x x x x

^a Realizados a menudo para responder a problemas específicos que podrían ser desconocidos al inicio de los programas. Por lo tanto, el tiempo y los fondos para tales actividades deben ser asignados, pero las características y fecha exacta de los estudios debe ser determinada según las necesidades reales.

otras organizaciones participantes) que están ya cubiertos por otras fuentes y que estarán disponibles para la actividad sin costo adicional.⁴⁴

Ejemplos: Salarios del personal y cargas sociales, espacio de oficina, equipo de oficina como computadoras y fotocopiadoras, vehículos para el trabajo de campo, etc.

- (Si no están cubiertos en el punto anterior) Estimación del costo del personal y otros costos directos por la coordinación de los distintos elementos de la evaluación.

Ejemplo: El salario y las cargas sociales de una persona empleada para recolectar y sintetizar las estadísticas de servicio de diferentes ramas del programa o de diferentes agencias participantes.

- Estimación de los costos para cada actividad de recolección de datos que va a ser llevada a cabo para la evaluación.

Ejemplo: Suponiendo que la evaluación consiste de (1) un seguimiento de las estadísticas de servicio recolectadas de rutina, (2) un análisis de situación al principio y final del ciclo quinquenal del proyecto, (3) una DHS al principio y al final del ciclo quinquenal del proyecto⁴⁵, (4) aplicación de la metodología COPE en 30 PPS, y (5) análisis del costo por APP según tipo de mecanismo de entrega de servicio. Es importante presupuestar cada uno de estos 5 aspectos por aparte, teniendo cuidado de no duplicar el conteo de aspectos que aparecen en más de una categoría, tales como la compra de microcomputadoras para usos múltiples.

- Estimación de los costos del procesamiento y del análisis de datos.

Ejemplo: El personal local puede encargarse del procesamiento y reporte de las estadísticas de servicio recolectadas de rutina. En cambio, los costos del procesamiento del análisis de situación, de los datos DHS y de los análisis de multinivel para valorar el impacto suelen ser elevados.

- Estimación de los costos de la diseminación de los resultados.

Ejemplo: Como se describe en el Capítulo VI, existen múltiples canales para la diseminación. Dada la importancia de este paso final del proceso, es fundamental que los fondos correspondientes estén presupuestados desde el principio.

Confeccionar presupuestos es una destreza que puede ser desconocida por algunos evaluadores. Así como se busca ayuda especializada para el muestreo, el diseño del estudio, etc., también puede ser recomendable buscar asistencia de personas expertas en presupuestos en una determinada institución para que las estimaciones presupuestarias sean realistas y se asegure la inclusión de todos los aspectos.

⁴⁴ Algunas organizaciones pueden preferir presupuestar estos costos, aunque estén cubiertos por los fondos existentes, para poder rastrear y reportar los costos completos de la evaluación para su organización.

⁴⁵ En el caso del Análisis de situación y la DHS, los costos variarán mucho entre países, los presupuestos deben elaborarse junto con el personal de estos proyectos que cuentan con experiencia en procesos de presupuesto específicos para estas actividades.

Capítulo VI

Diseminación y utilización de los resultados

- Adaptación de la presentación de los resultados para la audiencia objetivo
- Identificación de resultados con implicaciones programáticas importantes
- Establecimiento de un foro para presentar los resultados que implican una acción
- Presencia permanente en el proceso

PREVIOUS PAGE BLANK

DISEMINACIÓN Y UTILIZACIÓN DE LOS RESULTADOS

ADAPTACIÓN DE LA PRESENTACIÓN DE LOS RESULTADOS PARA LA AUDIENCIA OBJETIVO

La preparación de la presentación de los resultados para una audiencia de formuladores de políticas de alto nivel puede requerir los "resultados clave" de una evaluación, por ejemplo:

- ¿ha aumentado la prevalencia anticonceptiva?
- ¿cuáles provincias tienen las tasas más altas de prevalencia anticonceptiva?
- ¿cuáles métodos anticonceptivos son más populares?
- ¿está el programa llegando al área rural?

Además, considerando el bombardeo de información que reciben diariamente los encargados de la toma de decisiones de alto nivel, el informe para esta audiencia debe ser rápidamente comprensible y estéticamente atractivo. Por eso, el medio preferido puede ser un llamativo folleto o panfleto pequeño, con un resumen ejecutivo de los resultados clave, ilustrado con gráficas a color.

Los administradores del programa, los investigadores de la planificación familiar y otras personas afines al programa querrán mucho más detalle que el contenido del resumen ejecutivo. Esta audiencia se beneficiará del informe completo, con una tabla de contenidos clara y detallada, que dirija al lector a los temas específicos. Las gráficas también deberían utilizarse en este tipo de reporte, aunque algunos lectores también querrán ver los "números que respaldan las gráficas". Los detalles de la metodología deberán presentarse con claridad, pero una parte de este material puede presentarse en apéndices, si van más allá del interés o la comprensión del lector promedio. Si se usan métodos multivariantes complejos, los resultados deben presentarse en un formato que pueda ser compren-

dido por personas con poco o ningún conocimiento estadístico.

Este reporte exhaustivo generalmente constituye la versión más completa de la evaluación y suele aparecer en el idioma principal del país.

Los resultados de la evaluación deberían ser retroalimentados en el sistema, para los administradores de programa y proveedores de servicio cuyo trabajo se refleja en la evaluación. Los médicos y enfermeros muy ocupados pueden desanimarse si se incluyen demasiados resultados clave, pero estarían muy receptivos a algunos resultados seleccionados que relacionen directamente la entrega del servicio (las tendencias en la prevalencia, los cambios en la mezcla de métodos, las características de los usuarios comparadas con las de los no usuarios, la concentración por condición socioeconómica y área geográfica de la demanda no satisfecha, etc.). Este grupo también se beneficiaría de los desplegados de fácil comprensión para los formuladores de políticas. Otra vía de comunicación con este grupo sería mediante la presentación directa de resultados (conferencias, seminarios) seguida de sesiones de preguntas y respuestas y discusiones de grupos pequeños.

Una cuarta audiencia posible son los investigadores, quienes pueden ser alcanzados mediante revistas y conferencias internacionales. En el caso de un artículo de revista, es importante concentrarse en un tema específico. Mucho del texto descriptivo que aparece en un informe nacional exhaustivo puede ser condensado en pocos párrafos que comuniquen la contribución de este estudio particular a la literatura científica. Aunque el reporte exhaustivo puede tratar de cubrir la mayoría de las variables del análisis, un artículo de revista o ponencia debe destacar solo los puntos clave.

⁴⁶ Los autores reconocen la contribución de Ms. Ann Laughrin de "Futures Group International" en la preparación de este capítulo.

IDENTIFICACIÓN DE LOS RESULTADOS QUE TIENEN IMPLICACIONES PROGRAMÁTICAS IMPORTANTES

Las primeras tres audiencias mencionadas antes — formuladores de políticas, administradores de programa y personal de servicio — tendrán una pregunta común: ¿Qué significado tiene esto para mí? Por lo tanto, es fundamental que el evaluador no solo presente los resultados, sino también interprete su relevancia para los programas vigentes. En muchos casos, las audiencias mencionadas captarán las implicaciones programáticas sin que estas estén claramente formuladas (por ejemplo, la prevalencia anticonceptiva es de 50% en zonas urbanas, comparada con 20% en zonas rurales). Pero, la mayoría de sus miembros se beneficiarán si los evaluadores plantean lo obvio, aunque sea solo para reforzar su propia interpretación de los resultados. Además, este enfoque de la presentación de datos la hace parecer más práctica y útil o “menos académica”.

La cuarta audiencia —los investigadores— también estarán interesados en las implicaciones programáticas, especialmente en revistas o conferencias de tipo más aplicado que puramente académico

ESTABLECIMIENTO DE UN FORO PARA PRESENTAR RESULTADOS QUE IMPLICAN UNA ACCIÓN

Cada vez hay más administradores de programa bien capacitados que pueden leer un reporte con estadísticas y derivar de inmediato el significado para su programa. Pero, generalmente las personas encargadas de actuar con base en los resultados se beneficiarán mucho si tienen la oportunidad de:

- discutir y comprender mejor los resultados;
- internalizar o digerir esta información;
- reconciliar esta “nueva información” con su comprensión del programa de acuerdo con su manera de operar en este momento;
- verbalizar las implicaciones de los resultados en sus propios términos;
- identificar acciones que consideren las situaciones no cubiertas por la evaluación; y
- llegar a un plan de acción para capitalizar en las áreas fuertes y mejorar las áreas débiles.

Lo ideal es un foro que promueva la discusión e interacción entre las personas clave que puedan influir en la dirección futura del programa. La propia dinámica del grupo hará que los presentes cuestionen los resultados, traten de comprender los factores subyacentes que explican los resultados y, a través de este proceso, concedan mayor importancia a los resultados. Pero puede haber dos excepciones que le resten valor a la interpretación colectiva de los resultados. Primero, si las personas se ponen a la defensiva en relación a los resultados, este tipo de sesión puede ser contraproducente. Segundo, si alguno de los aspectos de los datos es cuestionable en cuanto a su validez, la discusión puede centrarse más en los defectos de los datos que en los defectos del programa.

PRESENCIA PERMANENTE EN EL PROCESO

Con demasiada frecuencia los evaluadores que son convocados a diseñar y realizar una evaluación, presentan sus resultados, dan una lista de recomendaciones y se van. En el escenario ideal, el administrador del programa tratará de incorporar las recomendaciones en la reestructuración del programa. En muchos otros casos, se aplica el proverbio: ojos que no ven, corazón que no siente.

Existen tres razones por las cuales vale la pena que los evaluadores permanezcan en contacto con los administradores del programa durante el período de implementación de los cambios originados en los resultados de la evaluación.

- El evaluador está disponible para consultas sobre las recomendaciones. Por ejemplo, cuando los resultados de la regresión de multinivel identifican los cambios en los indicadores de impacto que surgirían de la implementación de ciertas acciones, el administrador del programa podría buscar más aclaración sobre las modificaciones en la prestación del servicio necesarias para que el cambio tenga lugar.
- El contacto regular entre el administrador y el evaluador del programa sirve como recordatorio fundamental de que la evaluación fue hecha originalmente con un propósito: identificar áreas en que se puede mejorar. Contrario a: “ojos que no ven, corazón que no siente”, la presencia del evaluador sirve para promover acción en las áreas identificadas por la evaluación.
- El evaluador puede ayudar a establecer diseños adicionales para determinar si los

cambios establecidos en la entrega del servicio realmente producen los cambios esperados en la utilización del servicio, la

prevalencia anticonceptiva, etc. Como tal, esto refuerza la evaluación como proceso permanente en vez de como evento único.

Capítulo VII

Adaptaciones a otras intervenciones de salud reproductiva

- Prevención ETS/VIH
- Embarazo seguro
- Lactancia materna
- Nutrición de la mujer
- Servicios de salud reproductiva para adolescentes

ADAPTACIONES A OTRAS INTERVENCIONES DE SALUD REPRODUCTIVA

A partir de la Conferencia sobre Población y Desarrollo de 1994 en el Cairo, la "planificación familiar" se está expandiendo rápidamente de una estrecha focalización en la anticoncepción, hacia una gama más amplia de servicios de salud reproductiva. En la medida que los administradores de programa y los proveedores de servicio se desplazan cada vez más hacia el dominio de la salud reproductiva, los evaluadores deben considerar la pregunta: ¿cómo se evalúan estas intervenciones?

Este capítulo se centra en las siguientes áreas de la salud reproductiva: la prevención de las ETS/VIH, el embarazo seguro, la lactancia, la nutrición de mujeres y los servicios de salud reproductiva para adolescentes. Los "servicios para adolescentes" no constituyen un área separada de la salud, sino más bien, una población objetivo específica. Sin embargo, con el creciente reconocimiento de la necesidad de servicios para este grupo, es útil considerar las implicaciones para evaluar estos programas.

Los lineamientos de los capítulos I-IV de este manual para la evaluación de programas de planificación familiar, generalmente son aplicables a otras intervenciones de salud reproductiva. En términos de las mediciones basadas en el programa, los indicadores pueden cambiar, pero el enfoque básico es similar en las diversas áreas de la salud reproductiva. El evaluador se interesa en dar seguimiento a:

- el entorno político-programático;
- el número y tipos de actividades realizadas en distintas áreas funcionales;
- el acceso a los servicios;
- la calidad de la atención; y
- el nivel de utilización del servicio (por ejemplo, el número de visitas, el número de clientes nuevos, el volumen de suministros distribuidos).

En cambio, hay diferencias muy marcadas

entre la planificación familiar y otras áreas de la salud reproductiva, en términos de población objetivo, medidas de resultado a nivel de población, y la factibilidad de la recolección de datos. Este capítulo está organizado para cubrir estos temas para cada área de la salud reproductiva. NO regresa a los temas de demostrar causalidad (excepto en el caso de la nutrición de las mujeres) discutido en capítulos anteriores de este manual, los cuales son igualmente aplicables a la planificación familiar que a otras áreas de la salud reproductiva.

Es importante reconocer que la planificación familiar es una de las intervenciones de la salud pública que ha sido evaluada con mayor rigor a nivel internacional. La literatura refleja más de 30 años de esfuerzo conjunto para encontrar la forma metodológicamente más sólida, pero práctica, de evaluar este tipo de programa. En cambio, los intentos por evaluar otras áreas de la salud reproductiva son más recientes; mucho del trabajo ha sido realizado a partir de mediados de la década de los ochenta. Con una mayor ejecución de intervenciones de salud reproductiva, los evaluadores obtendrán más conocimiento para adaptar las metodologías de evaluación existentes a las circunstancias especiales de estos otros programas.

PREVENCIÓN DE ETS/VIH

Población objetivo

Las intervenciones de salud reproductiva generalmente fijan como objetivo la población en riesgo. Para ciertas áreas de la salud reproductiva (por ejemplo, la planificación familiar, el embarazo seguro), incluye las mujeres del país de edades relevantes. En la práctica las intervenciones tienden a excluir grupos que, en virtud de su condición económica, están en capacidad de obtener servicios de salud reproductiva mediante fuentes privadas; sin embargo, el éxito en un determinado país se mide en términos de la población

total de mujeres -- o de mujeres casadas -- en el rango de edades reproductivas (por ejemplo, prevalencia anticonceptiva, porcentaje de mujeres que dan a luz bajo condiciones supervisadas).

En cambio, respecto a la infección ETS/VIH, no toda la población está en riesgo. Las intervenciones ETS fijan como objetivo (1) los "bolsones de difusores" (aquellos más responsables de mantener la epidemia ETS), (2) las personas sintomáticas que buscan alivio para sus síntomas, (3) las personas con alto riesgo de infección debido a su conducta o susceptibilidad biológica o (4) aquellos ya infectados (por ejemplo, hijos por nacer de mujeres embarazadas con sífilis). En la mayoría de los países, sería un desperdicio de recursos tratar de alcanzar la población en general y una forma muy poco efectiva de alcanzar los que están en riesgo. De hecho, el riesgo de la ETS/VIH varía notablemente dentro de un determinado país, región o, incluso, ciudad. El riesgo está relacionado con (1) el nivel de prevalencia ETS/VIH en una determinada área geográfica o población y (2) las normas y prácticas sexuales de las subpoblaciones. Las intervenciones ETS/VIH tienden a fijar como objetivo los grupos de personas con conductas de alto riesgo: trabajadoras comerciales del sexo, choferes de camiones, trabajadores migrantes y adolescentes. Por lo tanto, el programa debe evaluarse en relación a los efectos sobre la conducta de estas poblaciones.

En contraste con otras áreas de la salud reproductiva que fijan como objetivo a las mujeres en edad reproductiva, las intervenciones ETS/VIH deben fijar como objetivo, tanto a hombres como a mujeres. Algunos argumentarían que los hombres son, incluso, más importantes que las mujeres, dado su papel en la toma de decisiones en el ámbito sexual (y otros) en muchos países. Las encuestas tipo DHS están centradas principalmente en las mujeres en edad reproductiva, pero en años recientes se ha extendido a los hombres (como una muestra independiente de hombres seleccionados aleatoriamente o una muestra de esposos o compañeros). Dado que la conducta de riesgo ETS/VIH parece estar asociada con el estado conyugal, es preferible conseguir muestras independientes de hombres que muestras de esposos o compañeros para generar estimaciones más válidas de variables clave dependientes o independientes relacionadas con las ETS/VIH.

En resumen, no hay una población objetivo estándar o convencional para las intervencio-

nes ETS/VIH. Varía por programa y por país y debe ser valorado en cada situación.

Indicadores de resultados y factibilidad de medirlos

El objetivo principal de las intervenciones ETS/VIH es detener la propagación de la infección, particularmente la infección VIH. Lógicamente, parecería que la medida de éxito a largo plazo debería ser el nivel de VIH en un país determinado. De manera semejante sería útil rastrear los cambios en la incidencia de la infección (o sea, la tasa de nuevas infecciones).

Sin embargo, hasta la fecha ha sido prácticamente imposible obtener datos sobre prevalencia de VIH en una muestra seleccionada aleatoriamente de la población general de un país, por varias razones⁴⁷. Primero, la prueba VIH requiere muestras biológicas (o sea, sangre, orina o saliva) que son más difíciles de recoger que las respuestas verbales normales en el contexto de una encuesta nacional. Segundo, el costo de este tipo de encuesta es considerable. Tercero, y posiblemente lo más importante, las pruebas VIH plantean varios dilemas éticos: puede uno medir la condición VIH y no informar los resultados al entrevistado? ¿Es ético informar a personas que son seropositivas sin ofrecer orientación o tratamiento? ¿Cómo manejaría el programa el estigma que las personas seropositivas sufren en muchas sociedades?

Esporádicamente hay datos disponibles de los establecimientos que ofrecen servicios prenatales y/o obstétricos; de hecho, muchos de los datos usualmente citados sobre prevalencia VIH se basan en esta fuente de información. Aunque esta puede parecer una forma fácil de obtener datos sobre prevalencia VIH, tiene, en realidad, muchas limitaciones. Primero, las mujeres que usan los servicios no necesariamente representan la población general. Segundo, este tipo de prueba agregaría un procedimiento adicional para los establecimientos que no hacen rutinariamente pruebas VIH y no necesariamente están preparadas para hacerlas. Tercero, en la perspectiva de la parturienta, esta prueba significaría una molestia más adicional. Cuarto, dado que los datos basados en registros de hospitales

⁴⁷ Hay algunos ejemplos de muestras poblacionales para pruebas de VIH en Africa: en la región de Mwanza, Tanzania (Grosskurth et al., 1995) y en el distrito de Rakai en Uganda (Wawer et al., 1995). Sin embargo, estas son excepciones a la regla.

suelen ser de baja calidad, no hay razón para creer que los datos sobre condición VIH serían mejores, excepto en el contexto de un proyecto de investigación especial. En resumen, aunque los datos de sero-condición de mujeres en establecimientos terciarios son útiles para efectos de políticas (para reflejar la magnitud del problema en términos generales), estos datos no son muy satisfactorios para efectos de evaluación (Hassig, 1995).

Finalmente, aunque el objetivo a largo plazo puede ser determinar la prevalencia VIH, no es una medida útil para medir el impacto de programas, puesto que no cambia de inmediato como respuesta a los cambios en las prácticas o conductas deseadas. O sea, que aunque fuera posible detener la transmisión de la infección VIH, esto no resultaría en una caída inmediata de la prevalencia VIH. Para citar un ejemplo basado en el modelaje de una población hipotética de alta incidencia, Mertens et al. (1994) estiman que si la incidencia VIH disminuye en un 25% durante un período de cinco años debido a un programa de prevención exitoso, en el mejor de los casos se daría una disminución mínima en la prevalencia VIH observada en ese mismo período.

En resumen, las encuestas de sero-prevalencia no han sido ampliamente usadas como medio para dar seguimiento a la epidemia de SIDA o para evaluar las intervenciones. Aunque esos datos estuvieran disponibles, sería difícil asociar los cambios en la incidencia de la infección VIH con las intervenciones del programa (en ausencia de un experimento aleatorio).

Algunos han argumentado que la prevalencia ETS es un sustituto útil de la infección VIH. De hecho, en Tailandia, los datos sobre disminuciones en las tasas de ETS son citados como señal de esperanza para la disminución de la propagación del VIH (Hananberg et al, 1994). Debido a que existen tratamientos, aunque no curas, para la mayoría de las ETS, los dilemas éticos asociados con las pruebas de las ETS son menores que aquellos asociados con el VIH. Sin embargo, las dificultades técnico-operativas son realmente más desafiantes. Primero, los métodos disponibles para el tamizaje de la mayoría de las ETS (con la excepción de la sífilis) son onerosas o requieren un apoyo significativo de infraestructura para preservar y transportar las muestras, así como la preparación y realización de las pruebas de las muestras para determinar la presencia de infección (por ejemplo, LCR/PCR⁴⁸ o cultivos para herpes o gonorrea). Segundo, debido a que

varias ETS generalmente se encuentran en distintas proporciones en las poblaciones, la medición de una sola ETS puede ser inapropiada; los evaluadores necesitarían establecer un perfil ETS al inicio, si fueran a rastrear una sola ETS para efectos de evaluación. Tercero, algunas pruebas de tamizaje de ETS requieren pruebas confirmatorias para determinar la naturaleza (por ejemplo, sífilis nueva, vieja, tratada) de la infección, mientras que algunas ETS (por ejemplo, el herpes) son condiciones crónicas (como el VIH) y la sero-condición de una persona no va a cambiar, aunque ocurra un cambio de tratamiento o de conducta. Cuarto, aunque la existencia de opciones de terapia y curación para las ETS hace que las pruebas para ellas sean teóricamente más factibles, pocos países están adecuadamente provistos de medicinas suficientes para lidiar con el problema. Finalmente, solo unas pocas ETS pueden ser identificadas mediante procedimientos simples y poco invasivos como un pinchazo en un dedo o en una vena. La mayoría requiere frotis u otras muestras recolectadas genitualmente, que no son fáciles de obtener; sin embargo, las pruebas de orina o los frotis auto-administrados ofrecen formas potenciales para lograr que las pruebas a nivel de la población sean más factibles.

Soluciones recientes para estos problemas de medición

La importancia de incluir a los hombres en la evaluación de ETS/VIH está ampliamente reconocida. De hecho, las Encuestas de Salud Reproductiva DHS y CDC incluyen cada vez más entrevistados masculinos.

Sin embargo, la naturaleza y el alcance de la intervención ETS/VIH debe ser considerada cuidadosamente al tratar de interpretar cualquier dato de tales encuestas para propósitos de evaluación. Por ejemplo, un programa de prevención centrado en la población masculina militar y universitaria, trabajadores migrantes como choferes y poblaciones de mujeres de alto riesgo definidas geográficamente, pueden tener un impacto mayor sobre la dinámica de las epidemias gemelas de ETS y VIH, cambiando la conducta sexual dentro de estos grupos, pero en el mejor de los casos logran

⁴⁸ LCR se refiere a la reacción en cadena a la ligasa; PCR es una reacción en cadena la polimerasa. Ambas son técnicas de amplificación DNA

muy poca evidencia del cambio a nivel de la población⁴⁹. La mayoría de los programas de prevención de ETS/VIH no son ni financiados ni implementados con una intensidad y cobertura lo suficientemente grandes como para tener impacto nacional.

Para programas con objetivos fijados en una subpoblación específica, el enfoque del "sitio centinela" está surgiendo como una estrategia prometedora de evaluación. Estos esfuerzos piloto de recolección de datos han sido centrados en poblaciones no representativas, pero relevantes para el programa. En el ejemplo anterior, un programa podría establecer una "encuesta de vigilancia conductual" en las poblaciones universitarias y militares que permitiría la medición de conductas críticas (y posiblemente otros indicadores) en intervalos significativos. Para una descripción de cómo ha sido utilizado este enfoque exitosamente en Tailandia, ver AIDSCAP Evaluation Tools, Módulo 4 (AIDSCAP, 1995).

EMBARAZO SEGURO

Población objetivo

Embarazo seguro se refiere al mejoramiento del resultado del embarazo en una mujer y su recién nacido. La población objetivo para las iniciativas de embarazo seguro consiste en dos grupos: (1) las mujeres en edad reproductiva, y más específicamente, las mujeres que han tenido un embarazo durante un período de tiempo delimitado y (2) los recién nacidos durante el período de referencia. Una valoración integral de los resultados del programa requieren información de "casos" que incluya la terminación del embarazo y los eventos que ocurren durante el puerperio para los embarazos que resultan en nacidos vivos.

Las iniciativas de embarazo seguro involucran toda la población objetivo (como se definió anteriormente) en la medida en que la atención "de rutina" es promovida por las autoridades de salud para todas las mujeres embarazadas y todos los recién nacidos. Tales servicios pueden ofrecerse mediante muchos tipos de proveedores de servicios.

Una gran preocupación es la pericia y el desempeño de los sistemas de salud y de los proveedores de servicio para manejar las complicaciones del embarazo, la labor y el parto, así como aquéllas que surgen durante el período neonatal. Particularmente impor-

tantes son las complicaciones maternas que amenazan las vidas de las mujeres embarazadas (o sea, hemorragias, sepsis, labor prolongada/obstruida, aborto séptico, y desórdenes hipertensos) y las condiciones que representan las principales amenazas para los recién nacidos (o sea, asfixia, daños al nacer, sepsis, tétano y otras, debidas al bajo peso al nacer). Estas requieren una atención especializada que en muchos países no está disponible localmente, debiendo referirse a establecimientos de mayor nivel que implican logísticas más difíciles y costos más altos para las familias y para los proveedores de servicios. Por lo tanto, los casos de complicaciones constituyen un "objetivo" de especial interés para evaluar las iniciativas de programas de embarazo seguro.

Indicadores de resultados y factibilidad de medirlos

Los objetivos principales de las iniciativas de embarazo seguro son proteger la salud de las embarazadas y los recién nacidos mediante la prevención de resultados adversos del embarazo (o sea, abortos y nacidos muertos, defunciones perinatales, neonatales y maternas) mediante una oportuna y efectiva detección y manejo de las complicaciones que surgen durante el embarazo, el parto y el período neonatal. Varias clases o tipos de indicadores de resultados han sido identificados como útiles para evaluar las intervenciones de embarazo seguro.

Un grupo de indicadores consiste en lo que puede ser visualizado como medidas de resultado "final". A este grupo pertenecen varias tasas de mortalidad relacionadas con el embarazo. Por ejemplo:

- razón y tasa de mortalidad materna,
- tasa de mortalidad perinatal, y
- tasas de letalidad (para todas las complicaciones).

La razón de mortalidad materna, definida como el número de defunciones maternas por cada 100,000 nacidos vivos durante un período de referencia específico, ofrece una medida del riesgo obstétrico una vez que una mujer se embaraza. La tasa de mortalidad

⁴⁹ Tailandia es una excepción a esta regla. Parece que si la cobertura a los "bolsos difusores" es suficientemente alta, puede tener un impacto en la población general.

materna, el número de defunciones maternas durante un período de referencia por cada 100,000 mujeres en edad reproductiva, ofrece una medida más amplia de la posibilidad de quedar embarazada y morir durante el embarazo o el puerperio. Si los sistemas de registro civil son relativamente completos, tanto el numerador como el denominador de la razón de mortalidad materna, pueden ser derivados de esta fuente de datos. El denominador de la tasa normalmente se deriva de las proyecciones o estimaciones de población oficiales. Sin embargo, puesto que los sistemas de estadísticas vitales en la mayoría de los países en desarrollo tienden a estar incompletos, el dato del numerador de ambas medidas debe ser ampliado con datos de otras fuentes (por ejemplo, registros de defunciones de hospitales, registros de la morgue y de cementerios, etc.). Debe tenerse en cuenta que incluso en países desarrollados, las defunciones maternas han sido subestimadas hasta en un 50%.

El principal enfoque alternativo de medición, la estimación "directa" de la mortalidad materna en encuestas por muestreo, ha resultado difícil de implementar porque se requieren tamaños muestrales muy grandes para obtener estimaciones suficientemente precisas y por los sesgos asociados con los hogares que se disuelven al fallecer la mujer. Las estimaciones "indirectas" derivadas del "método de las hermanas fallecidas" son más prometedoras⁵⁰. Aunque este método indirecto es menos exigente en cuanto al tamaño muestral que las estimaciones directas, siempre se requieren muestras relativamente grandes, si se desean estimaciones desagregadas geográficamente. Otra desventaja es que las estimaciones del nivel de mortalidad materna que produce el método se refiere a una fecha en el pasado relativamente lejano, normalmente 10 o más años antes de la fecha de la encuesta.

Para estimar la tasa de mortalidad perinatal (la razón de defunciones fetales tardías más defunciones neonatales tempranas por 1,000 nacidos vivos durante el período designado) corresponden fuentes de datos y limitaciones semejantes. Sin embargo, debido al problema de recordar las fechas de las defunciones fetales tardías e infantiles tempranas y a un subregistro serio de defunciones que ocurren inmediatamente después del nacimiento, en muchos países se pone en duda que las encuestas puedan proveer una alternativa apropiada para las estadísticas vitales y/o registros hospitalarios.

La tasa de letalidad (o caso-fatalidad), defi-

nida como la razón del número de defunciones por complicaciones específicas del embarazo sobre el número de casos obstétricos complicados que se presentan en un establecimiento de salud durante un período designado, mide la posibilidad de que una mujer con complicaciones obstétricas sobreviva una vez que entra al establecimiento de salud para recibir tratamiento. A nivel agregado para los establecimientos, ofrece una medida directa de la capacidad del sistema de salud para lidiar con emergencias obstétricas. Sin embargo, para obtener estimaciones para la población total, se requiere una cobertura relativamente completa de todos los establecimientos que ofrecen servicios obstétricos de emergencia.

Un segundo grupo de indicadores pertenece a las características de las defunciones fetales tardías e infantiles. Por ejemplo:

- distribución porcentual de las defunciones perinatales por edad al morir;
- razón de fresco/macerado nacido muerto;
- tasa de mortalidad proporcional al peso al nacer; y
- tasa de mortalidad específica por peso al nacer.

La principal utilidad de estos indicadores es proveer información sobre las contribuciones relativas de los fracasos de los distintos componentes del proceso de entrega de servicio (o sea, atención prenatal, atención en el parto, atención al recién nacido, atención al niño) a la mortalidad perinatal y, en forma correspondiente, al componente que requiere ser reforzado.

Teóricamente tanto los datos de establecimientos como los datos de las encuestas podrían ser usados para medir estos indicadores. Es cuestionable, empero, que las entrevistadas en encuestas de muchos países en desarrollo puedan ofrecer información suficientemente precisa sobre la edad o duración de la gestación al morir y distinguir los óbito fetales frescos de los macerados. Si las encuestas no son confiables, es necesario depender

⁵⁰ El método de las hermanas fallecidas se basa en reportes de casos de hermanas de entrevistadas en encuestas que murieron después de los 15 años de edad por causas relacionadas con el embarazo. El método convierte esta información en una estimación de la razón de mortalidad materna. Los detalles sobre el método se encuentran en Graham et al., (1990).

así, hay problemas con la determinación de las edades exactas al morir. Puesto que en ciertos países en desarrollo la mayoría de los nacimientos ocurren en el hogar y el peso al nacer es difícil de medir en el hogar, los datos de establecimientos no necesariamente ofrecen la información necesaria para determinar los fracasos en los servicios de salud.

El tercer grupo de indicadores consiste en lo que pueden ser llamado indicadores de "conocimiento/cobertura". Este grupo incluye una serie de CAP típicas (por ejemplo, la proporción de adultos con conocimiento sobre las complicaciones del embarazo y el parto, el porcentaje con conocimiento sobre la ubicación de los servicios obstétricos esenciales, etc.) y los indicadores de "cobertura" (por ejemplo, la proporción de población que reside a una hora de viaje del establecimiento que ofrece atención obstétrica esencial [AOE], la proporción de mujeres atendidas al menos una vez durante el embarazo por personal médico capacitado, etc.). Además, un indicador que vincula directamente la oferta y la demanda de servicios obstétricos se sugiere como especialmente útil: la necesidad satisfecha de atención obstétrica de emergencia. Este indicador ha sido operacionalmente definido como la proporción de mujeres que se estima tienen complicaciones obstétricas directas, vistas en un establecimiento que ofrece atención obstétrica de emergencia durante un período de tiempo especificado⁵¹.

La mayoría de los indicadores de conocimiento/cobertura pueden ser medidos con precisión sobre la base de encuestas por muestreo convencionales. La medición del acceso a los establecimientos que proveen atención obstétrica de emergencia mejora mucho si los Módulos de Disponibilidad del Servicio se realizan junto con la DHS y/o los Estudios de Análisis de Situación, así como donde hay Sistemas de Información Geográfica (GIS). La medición relacionada con la demanda, aunque se interpreta sobre la base poblacional, requiere datos de los establecimientos sobre el número de complicaciones para los numeradores. También se cuestiona, en alguna medida, la validez del supuesto de que el 15% de todos los embarazos resulten en complicaciones⁵².

Soluciones recientes para estos problemas de medición

Las mejoras a los registros vitales y un refinamiento continuo del método de las hermanas son las promesas más importantes para la

medición de la mortalidad materna. Recientemente, se han propuesto los censos de áreas pequeñas y pequeñas encuestas "de cuasi cobertura" para ofrecer estimaciones más actuales y geográficamente desagregadas que las del método de las hermanas.

Uno de los temas clave bajo estudio es si algunos indicadores medidos actualmente sobre la base de los datos de establecimientos pueden ser medidos con razonable precisión en las encuestas de hogares. Se necesitan medios más precisos para obtener información válida sobre la edad o duración de la gestación al morir y pesos al nacer en las encuestas de hogares. La realización de "autopsias verbales" a gran escala, tal vez junto con encuestas de hogares grandes como la DHS, representan una opción prometedora para obtener datos a nivel verdaderamente poblacional sobre ciertos indicadores, pero requiere mayor experimentación en el contexto de la medición de resultados de embarazo seguro⁵³. El Grupo de Trabajo en indicadores de Salud Reproductiva del Subcomité del Proyecto EVALUATION también ha propuesto una iniciativa para revisar el sistema de registro y de expedientes en los establecimientos de salud sobre casos de complicaciones de embarazos para mejorar la calidad de la información disponible para el seguimiento y manejo de programas

LACTANCIA MATERNA

Población objetivo

La población objetivo para los programas de lactancia incluye todas las mujeres que están en riesgo de concebir en una población. Se da

⁵¹ Ver "Indicators for Safe Pregnancy," the EVALUATION Project, (1995) para definiciones precisas de los términos involucrados en estos indicadores y una discusión de sus usos.

⁵² Ver "Indicators for Safe Pregnancy," The EVALUATION Project (1995), y World Health Organization (1993), "Indicators to Monitor Maternal Health Goals", para una mayor discusión de este tema.

⁵³ Las "autopsias verbales" son instrumentos de encuestas usados para recoger información sobre signos y síntomas de enfermedades que preceden la muerte de un miembro del hogar reportado como muerto. Los detalles del método se encuentran en Gray et al. (1990) y Kalter et al. (1990).

énfasis a las mujeres embarazadas, en período perinatal o en lactancia. También hay subgrupos con necesidades especiales: madres por primera vez, madres trabajadoras y mujeres con "fracaso" de lactancia previo. Además de los grupos específicos, ciertas intervenciones educativas o promocionales también fijan como objetivo la población total para promover una cultura de apoyo a la lactancia materna en los hogares, lugares de trabajo y servicios de salud.

Los programas nacionales de lactancia materna generalmente son evaluados en relación a las prácticas de lactancia de las mujeres que han dado a luz recientemente. En cambio, las intervenciones destinadas a subgrupos específicos deberían ser evaluadas en relación a miembros de esos subgrupos. Por ejemplo, las intervenciones educativas destinadas a adolescentes y niños en edades escolares podrían ser valoradas en términos de los cambios en el conocimiento y las actitudes en el grupo.

Indicadores de resultados y factibilidad de medirlos

La política más reciente sobre la duración de la lactancia (UNICEF/UNESCO/WHO/UNFPA, 1994) establece que todos los niños deberían ser alimentados exclusivamente con leche materna desde que nacen hasta los 6 meses de edad. La Asamblea Mundial de Confirmación del código de mercadeo para sustitutos de la leche materna establece que los alimentos complementarios deberían ser introducidos aproximadamente a los seis meses.

Los programas de lactancia materna generalmente están diseñados para aumentar (1) la incidencia de la lactancia, (2) la prevalencia de la lactancia exclusiva durante los primeros 6 meses de vida, y (3) la duración de la lactancia. Las intervenciones específicas pueden estar destinadas a un obstáculo particular para la práctica óptima de la lactancia, como (1) la falta de conocimiento entre las madres y los proveedores de atención de la salud sobre los beneficios de la lactancia materna en la salud, la nutrición y el espaciamiento de los hijos, (2) procedimientos en el manejo de los recién nacidos en los hospitales que interfieren con la lactancia exitosa o (3) la ausencia de peritaje en la lactancia materna en la comunidad o los servicios de salud que ayuden a las madres a iniciar con éxito la lactancia desde el nacimiento de su hijo. Los programas de comunicación y educación hacen énfasis en evitar los sustitutos de la leche materna, en postergar

la introducción de líquidos y alimentos complementarios y en las formas de resolver problemas de lactancia en caso que ocurran.

Las medidas de resultado para la evaluación de programas de lactancia incluyen con frecuencia la iniciación de la lactancia en la primera hora de vida, la duración de la amenorrea, el uso de la amenorrea de la lactancia para la anticoncepción, las tasas de lactancia exclusiva, la duración media de la lactancia y la frecuencia de las tomas de alimento.

Las fuentes de datos preferidas para medir los resultados relacionados con la lactancia son las encuestas tipo DHS y versiones simplificadas de encuestas por conglomerados. En la mayoría de los países, es relativamente fácil hacer preguntas sobre la lactancia en una entrevista privada. El principal problema de medición es el sesgo para recordar. Puede resultar difícil a las mujeres recordar exactamente el número de tomas de alimento dadas al infante, incluso en las últimas 24 horas, o la edad exacta en meses en que se suspendió la lactancia. Además, la definición de lactancia exclusiva del investigador puede ser distinta de la de la madre, por lo que los indicadores deben ser operacionalizados cuidadosamente. Sin embargo, en comparación con otras áreas de la salud reproductiva, los indicadores de resultado pueden medirse de manera relativamente fácil y confiable usando encuestas por muestreo.

Si las encuestas DHS no están disponibles como fuente de datos para la valoración de los resultados de los programas nacionales, la alternativa preferida sería la encuesta a una muestra grande y representativa de mujeres que han tenido hijos recientemente (puesto que el recordatorio de la práctica de lactancia en el pasado es notablemente dudoso). Al diseñar estas encuestas, es importante realizar investigación cualitativa preliminar sobre el significado de los términos (por ejemplo, lactancia exclusiva) y probar con cuidado los instrumentos. Sin embargo, el costo de estas encuestas es elevado, lo cual excluye esta opción en la mayoría de los programas.

Aunque este manual está centrado en programas nacionales, muchos programas de lactancia son sub-nacionales, operando en puestos selectos de entrega de servicios o comunidades. La evaluación de tales programas no sería factible utilizando datos de encuestas a nivel nacional. Sin embargo, si una evaluación basada en establecimientos o en comunidades va a llevarse a cabo, es impor-

tante aclarar (1) cuáles son los resultados y objetivos específicos del programa (por ejemplo, el inicio dentro de la primera hora de vida, lactancia exclusiva por seis meses), (2) quién usará la información generada y (3) cuál es el propósito de la evaluación. Los usuarios típicos son los ejecutores de programas, los fundadores de programas y los encargados de la toma de decisiones.

Una fuente de datos para obtener información sobre las prácticas de lactancia neonatales (por ejemplo, inicio de la lactancia en la primera hora de vida) es la encuesta a nivel de hospitales o maternidades, o la entrevista de salida. Las entrevistas de salida pueden ser particularmente útiles si el grupo objetivo de la intervención son mujeres que dan a luz en los establecimientos de salud. También es una buena estrategia para valorar el grado en que los hospitales están realmente ejecutando iniciativas favorables al infante.

Soluciones recientes para estos problemas de medición

Para efectos de evaluación, a los administradores de programas y a las agencias patrocinadoras idealmente les gustaría saber el efecto de sus intervenciones sobre las prácticas de lactancia de la población objetivo. La base de datos en permanente expansión de las encuestas DHS ofrece una oportunidad prometedora para futuras evaluaciones empíricas de las prácticas de lactancia a nivel nacional y regional.

Los investigadores, a su vez, están interesados en demostrar el efecto de las prácticas de alimentación a los niños sobre la fecundidad, la nutrición infantil y resultados afines. Esto requiere grandes muestras y, preferiblemente, estudios longitudinales cuya realización y análisis representa un desafío (puesto que el impacto de las prácticas de alimentación a los niños también está afectado por la edad del niño, la salud de la madre y las condiciones socioeconómicas generales del hogar).

Hasta la fecha, la evaluación en esta área ha tratado estos dos temas en forma separada: (1) los efectos de las intervenciones de los programas sobre las prácticas de lactancia y (2) los efectos de la lactancia sobre la fecundidad y la mortalidad. El reto sigue siendo demostrar empíricamente que las intervenciones del programa no solo afectan la conducta a mediano plazo sino, también, influyen en la condición de la fecundidad y la salud a largo plazo.

Aunque demostrar impacto sigue siendo un

objetivo en la evaluación de intervenciones de lactancia materna, hay también mucho que aprender de los métodos menos rigurosos, los cuales son útiles para mejorar programas a nivel local. Por ejemplo, los resultados de salud neonatal y salud infantil rastreados antes y después de las intervenciones hospitalarias "pro-infantes" pueden proveer un medio útil para valorar su efectividad entre poblaciones selectas.

NUTRICION DE LA MUJER

Población objetivo

La definición de la población objetivo para las intervenciones de nutrición de mujeres en el contexto de la salud reproductiva es complicada por los efectos acumulativos de una mala nutrición. Una nutrición pobre durante los primeros años de la niñez puede y de hecho, causa, efectos físicos irreversibles. Los déficits de crecimiento causados parcialmente por una mala nutrición durante la niñez nunca son recuperados por muchas mujeres de los países en desarrollo, cuyos resultados son estatura pequeña, desproporción cefalopélvica, resultados de embarazo pobres y, posiblemente, otras consecuencias fisiológicas. Por lo tanto, las intervenciones destinadas a los problemas de nutrición de las mujeres pueden cubrir un amplio espectro de edad.

Las formas más comunes de una mala nutrición en los países en desarrollo son la malnutrición de proteína-energía y deficiencias de micronutrientes (vitamina A, yodo, hierro). Todas estas deficiencias están vinculadas a la pobreza y al analfabetismo. Sin embargo, la deficiencia de yodo puede estar más definida geográficamente. La deficiencia de vitamina A también está determinada parcialmente por factores ecológicos y hábitos dietéticos tradicionales. La deficiencia de hierro se extiende a través de los países desarrollados y en desarrollo.

La mayoría de los recursos dedicados a la nutrición de mujeres, sin embargo, se concentran en las mujeres embarazadas y en las que están en lactancia, por dos razones principales: (1) durante el embarazo y la lactancia, las necesidades nutricionales aumentan y (2) las deficiencias nutricionales maternas durante el embarazo pueden tener serias consecuencias en las mujeres y sus niños.

Las madres adolescentes representan una población objetivo especial porque deben satisfa-

cer sus propias necesidades nutricionales para el crecimiento y la de su hijo. Si los recursos lo permiten, los programas de nutrición de mujeres fijan como objetivo un rango de edad más amplio. Esto está siendo defendido, en parte, porque la condición nutricional antes del embarazo es un determinante importante de los resultados reproductivos y porque los asuntos de la salud de las mujeres se consideran ahora prioritarios por derecho propio.

Indicadores de resultados y factibilidad de medirlos

El objetivo a largo plazo de los programas nutricionales es reducir la incidencia y la prevalencia de las deficiencias de proteína-energía y de micronutrientes. Esta es una meta ambiciosa. Las causas inmediatas de las deficiencias nutricionales son el insumo dietético inadecuado y una absorción pobre de los nutrientes (a menudo resultado de infecciones y una deficiente atención de la salud). Las necesidades nutricionales de las mujeres en los países en desarrollo suelen acrecentarse por elevadas cargas físicas y frecuentes ciclos recurrentes de embarazo y lactancia.

La malnutrición de proteína-energía (PEM) se valora frecuentemente usando medidas antropométricas como el peso, la estatura y la circunferencia braquial, medidas que ahora se toman de rutina como parte de la DHS. Una baja estatura en las mujeres (<145 cm) refleja deficiencia nutricional durante la infancia y es poco probable que cambie durante el transcurso de la mayoría de los programas nutricionales (excepto, tal vez, en el caso de las adolescentes). Las medidas como peso, peso en relación a estatura y circunferencia braquial reflejan la delgadez de la mujer. Estas medidas deberían ser sensibles a las intervenciones nutricionales destinadas al problema PEM. Las medidas antropométricas maternas son fáciles de recoger y los indicadores basados en ellas se consideran válidos y confiables.

La valoración de los micronutrientes es más difícil que la condición antropométrica y por eso no forma parte de la rutina de la DHS. Se requieren pruebas de laboratorio o procedimientos hasta cierto punto invasivos, debido a que la condición de micronutriente se basa en el análisis bioquímico de la sangre (hierro y vitamina A), leche materna (vitamina A) o la orina (yodo). A pesar de los avances tecnológicos, estas pruebas complican todavía en forma sustancial la logística del trabajo de

campo. Los avances tecnológicos están simplificando cada vez más el proceso y han reducido mucho los costos financieros de las pruebas. En algunos casos, los países han optado por incluir la valoración de micronutrientes en sus programas DHS.

Para intervenciones selectas de micronutrientes, como el complemento y el refuerzo de la vitamina A y yodo, los niveles de cobertura son medidas "trazadoras" razonables de resultados intermedios (conceptualmente similar a la cobertura de la inmunización). La evaluación de los programas de complemento de hierro y programas destinados al PEM no pueden basarse en esta clase de medidas aproximadas.

Un problema que enfrentan todas las intervenciones de salud reproductiva es demostrar que una intervención determinada es responsable del cambio en la población objetivo (de hecho, es un tema importante en este manual). En ninguna parte se hace más evidente este desafío que en los programas de nutrición de la mujer (especialmente aquellos destinados a la PEM y a la anemia). Esto se debe, en parte, a las múltiples causas subyacentes de la malnutrición. Frecuentemente los programas no resultan en un cambio en la condición nutricional porque (1) no fijan como objetivo a las personas más necesitadas, (2) no fijan como objetivo las múltiples causas (dietéticas y no dietéticas) de la malnutrición o (3) lo hacen en forma inapropiada.

La seguridad y disponibilidad de alimentos en el hogar, las condiciones económicas favorables, el acceso a los servicios de salud y un ambiente saludable, son todas condiciones necesarias para un adecuado insumo dietético y control de enfermedades. Lo más frecuente es que la gente pobre sea la que está afectada adversamente por una disponibilidad inadecuada de alimentos. La naturaleza multicausal de la mala nutrición de las mujeres complica de manera particular la atribución causal.

Soluciones recientes para estos problemas de medición

El énfasis en el campo de la salud internacional en las deficiencias de micronutrientes es probable que resulte en una disponibilidad creciente de datos para identificar los indicadores trazadores de la condición de micronutrientes. También, los avances tecnológicos en las técnicas de valoración fisiológica y bioquímica aumentarán la factibilidad de la utilización de estos enfoques.

SERVICIOS DE SALUD REPRODUCTIVA PARA ADOLESCENTES

Los programas para adolescentes son relativamente nuevos y hasta la fecha ha habido poca evaluación sistemática de su efectividad en países en desarrollo (Senderowitz, 1995). La evidencia limitada de los países en desarrollo aún debe demostrar que los programas para adolescentes reducen las conductas sexuales riesgosas, el embarazo y la transmisión de ETS (Kirby, 1994). Aún en los EEUU, la evidencia empírica sobre la efectividad de los programas para adolescentes y su funcionamiento es limitada (Brown y Eisenberg, 1995).

El aumento reciente en los programas para adolescentes coincide con una acción focalizada renovada sobre el rendimiento de cuentas y los resultados en los círculos internacionales de salud reproductiva. En la medida que estos programas continúan desarrollándose, ofrecerán oportunidades para evaluar qué funciona y en cuáles contextos.

Población objetivo

Los "servicios de salud reproductiva para los adolescentes" se traslapan con otros temas presentados en este capítulo, puesto que pueden incluir una o más áreas (ETS/HIV, embarazo seguro, lactancia y nutrición) además de otros servicios sociales (orientación, prevención de drogas, capacitación laboral) y actividades recreativas. La característica común y determinante de este grupo de intervenciones multifacéticas es la edad de la población objetivo; aunque el criterio de edad no es estándar para todos los programas.

La adolescencia comprende etapas de transición física y emocional de la niñez a la edad adulta. Fisiológicamente, la adolescencia es un período de rápido crecimiento y desarrollo de características sexuales secundarias. También es un período de turbulencia emocional durante el cual el adolescente lucha por alcanzar independencia de sus padres o tutores. Aunque estas etapas son universales, pueden ocurrir a edades muy distintas en las diferentes culturas. Por esta razón, no hay un solo criterio de edad generalizable para usar en los distintos contextos.

Algunos programas, especialmente aquellos interesados en la anticoncepción, usan el rango de 15-19 años, el cual tiene la ventaja de coincidir con las encuestas tipo DHS. Sin embargo, muchos programas tratan de llegar a personas más jóvenes, al menos con infor-

mación y aclaración de valores, mucho antes de que lleguen a estar activos sexualmente. Por eso, el rango de edad para ciertas intervenciones puede bajar hasta incluir los 10-12 años. Debido a que los problemas de la adolescencia suelen continuar hasta los 20 y más años, ciertos programas extienden sus servicios para los jóvenes de estas edades (Stewart y Eckert, 1995).

La Organización Mundial de la Salud (OMS) ha definido la adolescencia en dos etapas. Todas las personas de 10 a 19 años se definen como adolescentes. El grupo de 10 a 14 se clasifica como "adolescencia temprana" y el de 15 a 19 como "adolescencia tardía". La segunda categoría puede subdividirse en dos rangos de 15 a 17 y de 18 a 19, si resulta apropiado para el programa. La OMS sugiere también que el término "juventud" puede usarse para referirse a personas de 15 a 24 años y "personas jóvenes" para todo el grupo de 10 a 24 (WHO, 1989).

Una segunda dimensión posible usada para definir la población objetivo de los programas para adolescentes es la condición marital. En muchos países los programas para adolescentes han evolucionado para satisfacer las necesidades de salud reproductiva de las mujeres jóvenes no casadas que no se sienten cómodas usando el mismo servicio de las mujeres mayores (la mayoría) casadas. Esta acción focalizada es particularmente apropiada si los adolescentes llegan a estar sexualmente activos en edades relativamente tempranas y el matrimonio se pospone para estudiar o por otras razones. Sin embargo, en partes de Surasia y el Mediano Oriente, hay necesidad de centrar los servicios en las parejas jóvenes casadas, especialmente si existen programas que ignoran las necesidades de las mujeres nulíparas o incluso definen el ser madre como condición para obtener servicios de planificación familiar (Mensch, 1995).

En resumen, no hay una definición estándar de adolescencia. La población objetivo usada para evaluar las intervenciones de salud reproductiva para adolescentes deberían ser consistentes con los criterios establecidos por el programa.

Indicadores de resultados y factibilidad de medirlos

Objetivos

Los programas de salud reproductiva para adolescentes tienen múltiples objetivos que difieren de un programa a otro. Por ejemplo, algunos están diseñados principalmente como intervenciones educativas para aumentar conocimiento, crear conciencia y formar actitudes positivas; no tienen un componente de servicio. Otros, tales como los servicios de salud integrales, pueden ofrecer servicios múltiples (por ejemplo, planificación familiar, tratamiento de ETS, orientación nutricional, protección de situaciones de abuso, orientación sobre drogas, etc.). Incluso otros programas pueden ofrecer uno o más servicios que no son de salud, como actividades de generación de ingreso, servicios legales, orientación laboral, actividades recreativas y otras. La evidencia anecdótica sugiere que los programas con una amplia gama de actividades pueden ser menos controversiales que aquellos que ofrecen solo servicios de salud reproductiva.

En resumen, no hay un objetivo único o grupo de objetivos que sirvan para evaluar sistemáticamente los servicios de salud reproductiva para adolescentes. Más bien, los objetivos deben ser definidos en relación a los programas específicos. Además, en la medida que los servicios de salud reproductiva se ofrezcan junto con otras actividades, la evaluación de estos programas debería también rastrear resultados ajenos a la salud.

Medidas a nivel de programa frente a medidas a nivel de población

Si los servicios para adolescentes fueran tan bien desarrollados y de largo alcance como son los servicios anticonceptivos para las mujeres adultas en muchos países del mundo, uno esperaría cambios en estos indicadores a nivel de población como resultado de las intervenciones de programas. Sin embargo, la realidad es que la mayoría de los programas para adolescentes (si existen) todavía están en pañales. Llegan a un segmento relativamente pequeño de la población y suelen estar limitados a la(s) principal(es) área(s) urbana(s) del país. Aunque estos programas pueden tener un efecto considerable sobre las personas que atienden, por ahora la cobertura es extremadamente limitada⁵⁴. Por eso, aunque uno esperara cambios en el conocimiento, actitudes y conducta de los clientes o participantes

de estos programas, es improbable que los programas para adolescentes, en su forma actual, tengan la magnitud o intensidad suficiente para afectar las medidas de la conducta a nivel de la población.

Conceptualmente, los cambios a nivel de la población son la meta a largo plazo de la mayoría de los programas para adolescentes; sin embargo, la evaluación de tales programas, en términos de cambios en la población, serían de utilidad cuestionable. De hecho, podría, incluso, ir en detrimento del apoyo sostenido a estos esfuerzos, si quienes carecen de una buena comprensión de los aspectos técnicos, concluyeran que los programas para adolescentes "no producen resultados". En muchos casos sería más productivo evaluar los servicios de salud reproductiva para adolescentes en términos de los cambios en conocimiento, destrezas y conductas entre los clientes o participantes en estos programas. (Stewart y Eckert, 1995).

Factibilidad de la recolección de datos

Hasta la fecha, la mayoría de esfuerzos de recolección de datos ha incluido entrevistas con adolescentes (en las escuelas, en los hogares, en las clínicas, etc.), cuestionarios autoadministrados o "tests" (por ejemplo, pretest-posttest para medir el conocimiento ganado con una intervención educativa), grupos focales, observación en las clínicas, análisis de registros clínicos y otros métodos estándar. Se debe poner atención especial a la formulación de las preguntas (para asegurarse que los conceptos clave se presentan en el lenguaje informal de los adolescentes, más que en la terminología de la comunidad médica). También puede requerirse una autorización especial, por ejemplo, para administrar un cuestionario a los adolescentes en la escuela o entrevistar un adolescente no casado en su casa.

Los resultados intermedios o las medidas de conducta claves en los programas para adolescentes (por ejemplo, edad a la primera relación sexual, uso de anticonceptivos en la primera y última relación sexual, embarazo no planeado, autoreporte de ETS, autorepor-

⁵⁴ Una excepción notable es el vídeo musical con Tatiana y Johnnie que recorrió Latinoamérica a fines de los ochenta; este podría ser llamado apropiadamente una intervención, pero no constituyó un programa en sí.

te de uso de drogas, etc.) puede obtenerse mediante autoreporte en encuestas. Aunque existen algunas interrogantes relacionadas con la validez de las respuestas de adolescentes a preguntas sensibles⁵⁵, esta información, puede recogerse mediante entrevista directa o cuestionario autoadministrado.

La factibilidad de medir resultados de largo plazo varía según el área de la salud reproductiva, como se refleja en las secciones anteriores. Por ejemplo, es relativamente fácil medir tasas de fecundidad específicas por edad o tasas de fecundidad deseada entre adolescentes (si se definen con el rango de 15 a 19 años de edad) basándose en el auto-reporte de un estudio tipo DHS. En cambio, es muy problemático medir tasas de mortalidad materna o prevalencia de VIH en la población adolescente. La condición nutricional es fácil de medir, pero es difícil atribuir el cambio a las intervenciones específicas de los programas.

Soluciones recientes para estos problemas de medición

Un primer paso importante para evaluar programas de salud reproductiva para adolescentes es desarrollar un inventario de los distintos objetivos comunes a todos estos programas y definir los resultados intermedios correspondientes. Aunque puede ser posible medir el cambio a nivel de la población en relativamente pocos casos, este inventario representaría un menú útil para aquellos involucrados en la evaluación de tales programas.

Dependiendo de los recursos disponibles, los programas pueden tomar una (o ambas) de dos direcciones: (1) obtener datos descriptivos -a menudo cualitativos- sobre cómo ha funcionado el programa y dónde deben hacerse las mejoras y (2) realizar experimentos o cuasi-experimentos para medir el grado en que una intervención afecta la conducta (ya sea entre los participantes del programa o en circunstancias especiales a nivel de la población). El primero será importante para proveer retroalimentación a los administradores sobre los cambios que pueden hacer a corto plazo y a nivel local para aumentar la aceptación de las intervenciones entre los clientes, los clientes potenciales y la comunidad en su conjunto. La segunda será extremadamente importante para demostrar a la comunidad de donantes y a otras partes interesadas el tipo y la magnitud del cambio que puede ser esperado de los distintos tipos de intervenciones.

Para finalizar, la evaluación de las intervenciones de salud reproductiva para adolescentes todavía está en una etapa muy temprana como para identificar "soluciones recientes". Más bien, lo que es prometedor es el interés creciente por evaluar sistemáticamente las intervenciones en esta área.

⁵⁵ La evidencia anecdótica sugiere que los adolescentes parecen estar dispuestos a discutir estos temas con franqueza si el entrevistador es del mismo sexo y es también joven (menor de 30) (Morris, 1995).

CAPITULO VIII

RESUMEN: LISTA DE PASOS EN EL DISEÑO DE UN PLAN DE EVALUACION

- Definición del alcance de la evaluación (capítulo II)
- Definición del enfoque metodológico: seguimiento de programas (capítulo III)
- Definición del enfoque metodológico: valoración del impacto (capítulo IV)
- Desarrollo de un plan de implementación (capítulo V)
- Diseminación y utilización de los resultados (capítulo VI)

PREVIOUS PAGE BLANK

RESUMEN: LISTA DE PASOS EN EL DISEÑO DE UN PLAN DE EVALUACIÓN

Este manual está diseñado para asistir a los profesionales de la comunidad internacional de planificación familiar y salud reproductiva para:

- Diferenciar entre los tipos principales de evaluación de programas: el seguimiento de programas y la valoración del impacto;
- Evaluar críticamente las ventajas y desventajas de los métodos alternativos de valoración del impacto;
- Valorar y seleccionar el o los tipos de evaluación más apropiada para una situación determinada;
- Identificar los indicadores y las fuentes de datos apropiadas para la evaluación;
- Diseñar un plan de evaluación que indique el o los diseños de estudios, los indicadores y las fuentes de datos que sirven como plan de acción para una implementación posterior.

El manual revisa los pasos a seguir en el diseño de un plan de evaluación. Supone que la evaluación incluirá algún tipo de seguimiento de programas. Si el objetivo es también valorar el impacto, muchos de estos mismos pasos se aplican, pero el diseño de estudio apro-

piado debe identificarse e implementarse, como se indica en el Capítulo IV.

Aunque el mandato original del Proyecto EVALUATION fue evaluar el impacto de los programas de planificación familiar sobre la fecundidad, el ámbito de muchos programas de planificación familiar se ha extendido últimamente para incluir otras áreas de la salud reproductiva. Por esta razón, este manual revisa una serie de asuntos metodológicos (por ejemplo, población objetivo, medición de indicadores de resultados, factibilidad de la recolección de datos) para evaluar programas en otras áreas de la salud reproductiva (Capítulo VII), a saber:

- Prevención de ETS/VIH
- Embarazo seguro
- Lactancia
- Nutrición de mujeres
- Servicios de salud reproductiva para adolescentes

Los pasos básicos para el seguimiento de programas y la valoración del impacto pueden ser resumidos en forma de lista de chequeo, como la de la siguiente página.

LISTA DE CHEQUEO DE LOS PASOS EN LA EVALUACIÓN DE UN PROGRAMA

Definir el alcance de la evaluación (Capítulo II)

- Determinar las metas y objetivos del programa
- Describir cómo "debería" funcionar el programa (modelo conceptual)
- Establecer los objetivos de la evaluación
- Indicar el alcance de la evaluación en el plan de evaluación

Definir el enfoque metodológico: Seguimiento del programa (Capítulo III)

- Aclarar el propósito principal del seguimiento
- Identificar los elementos que deben recibir seguimiento
- Definir los indicadores relevantes
- Identificar las fuentes de datos
- Diseñar un formato para la presentación de los resultados
- Resumir el enfoque metodológico

Definir el enfoque metodológico: valoración del impacto (Capítulo IV)

- Revisar los requisitos metodológicos de los enfoques "preferidos" para valorar el impacto de los programas de planificación familiar:
 - √ Experimentos
 - √ Métodos de regresión multinivel
 - √ Cuasi-experimentos

- Valorar la factibilidad de usar uno de los enfoques preferidos
- (Si uno de los enfoques preferidos es factible) Identificar y negociar las necesidades especiales de datos para la situación específica de un país
- (Si los enfoques preferidos no son factibles) Revisar los enfoques alternativos para medir el impacto:
 - √ Descomposición (Modelo de Determinantes Próximos)
 - √ Método de Prevalencia
- Determinar e implementar el diseño óptimo para las circunstancias específicas del país.

Desarrollar un plan de implementación (Capítulo V)

- Definir las instituciones e individuos responsables de la evaluación
- Establecer un cronograma para actividades específicas
- Presupuesto para la evaluación

Diseminar y utilizar los resultados (Capítulo VI)

- Preparar la presentación de los resultados para la audiencia correspondiente
- Destacar los resultados que tienen implicaciones programáticas importantes
- Establecer un foro para presentar los resultados
- Permanecer involucrado en el proceso

**REFERENCIAS
Y
APENDICE**

Referencias

- AIDSCAP, 1995. "Application of a Behavioral Surveillance Survey Tool," AIDSCAP Evaluation Tools, Module 4. Family Health International, Arlington, VA.
- Angeles, G., T.A. Mroz, & D.K. Guilkey. 1995. "Purposive Program Placement and the Estimation of Program Effects: The Impact of Family Planning Programs in Tunisia." Paper presented at the Annual Meeting of the Population Association of America, San Francisco, CA.
- AVSC International. 1995. *COPE: Client-Oriented Provider Efficient Services. A Process and Tools for Quality Improvement in Family Planning and Other Reproductive Health Services*. New York, New York
- Bauman, K., C. Viadro, & A. Tsui. 1994. "Use of True Experimental Designs for Family Planning Program Evaluation: Merits, Problems, and Solutions." *International Family Planning Perspectives* 20(3): 108-113.
- Bertrand, J.T., R.J. Magnani, & J.C. Knowles. 1994. *Handbook of Indicators for Family Planning Program Evaluation*. Chapel Hill, NC. The EVALUATION Project.
- Bertrand, J.T., R. Santiso, S.H. Linder, and M.A. Pineda. 1987. "Evaluation of a Communications Program to Increase Adoption of Vasectomy in Guatemala." *Studies in Family Planning* 18(6): 361-370.
- Bertrand, J.T. and A. Tsui (eds.), 1995. *Indicators for Reproductive Health Program Evaluation*. Chapel Hill, North Carolina: The EVALUATION Project, Carolina Population Center, University of North Carolina at Chapel Hill.
- Bogue, D.J. 1970. *Family Planning Improvement through Evaluation: A Manual of Basic Principles*. Community and Family Study Center, University of Chicago, Manual No. 1.
- Bollen, K.A., D.K. Guilkey, & T.A. Mroz. 1995. "Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia." *Demography* 32(1):111-31.
- Bollen, K.A., D. Guilkey, & T.A. Mroz. 1992. "Methods for Evaluating the Impact of Family Planning Programs in Structural Models." Chapel Hill, NC: Carolina Population Center, The EVALUATION Project.
- Bongaarts, J. 1993. "The fertility impact of family planning programs," New York: Population Council. Working Paper no. 47.
- Bongaarts, J. 1978. "A Framework for Analyzing the Proximate Determinants of Fertility." *Population and Development Review*, 4(1): 105-132
- Bongaarts, J. 1986. "The Prevalence Method." in United Nations *Manual IX Addendum: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. New York: Department of International Economic and Social Affairs, pp. 9-14.
- Bongaarts, J. and S. Kirmeyer. 1980. "Estimating the Impact of Contraceptive Prevalence on Fertility: Aggregate and Age-Specific Versions of a Model." in Hermaulin, A. and B. Entwisle (eds.) *The Role of Surveys in the Analysis of Family Planning Programs*. Liege, Belgium: Ordina Editions, pp. 381-408.

- Bongaarts, J. and R. Potter. 1983. *Fertility, Biology, and Behavior: An Analysis of the Proximate Determinants*. New York: Academic Press.
- Brown, S.S. and L. Eisenberg (eds). 1995. *The Best Intentions: Unintended Pregnancy and the Well-being of Children and Families*. Washington: National Academy Press.
- Buckner, B., A.O. Tsui, K. McKaig, & A.I. Hermalin. 1995. *A Guide to Methods of Family Planning Evaluation: 1965-1990*. Chapel Hill, North Carolina: Carolina Population Center, The EVALUATION Project.
- Campbell, D. and J. Stanley. 1963. *Experimental and Quasi-Experimental Designs in Research*. Boston, MA: Houghton Mifflin Co.
- Casterline, J., L. Domingo, & Z. Zablan. 1988. "Trends in Fertility in the Philippines: An Integrated Analysis of Four National Surveys." Manila, Philippines: University of the Philippines Population Institute.
- Chamrathirong, A., P. Prasartkul, & A. Bennett. 1986. "Multivariate Areal Analysis of the Efficiency of the Family Planning Programme and Its Impact on Fertility in Thailand." Bangkok: Economic and Social Commission for Asia and the Pacific, Asia Population Studies Series No. 68.
- Chandrasekaran, C. and A. I. Hermalin. 1985. *Measuring the Effects of Family Planning Programs on Fertility*. Dolhain, Belgium: Ordina Editions.
- Cook, T. and D. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin Co.
- Entwisle, B., A.I. Hermalin, P. Kamnuansilpa, & A. Chamrathirong. 1984. "A Multilevel Model of Family Planning Availability and Contraceptive Use in Rural Thailand." *Demography* 21(4): 559-74.
- Fisher, A., J. Laing, J. Stoeckel, & J. Townsend. 1991. *Handbook for Family Planning Operations Research*. New York: Population Council.
- Fisher, A., B. Mensch, R. Miller, S. Askew, A. Jain, C. Ndeti, L. Ndhlovu, & P. Tapsoba. 1992. *Guidelines and instruments for a family planning situation analysis study*. New York: The Population Council.
- Freedman, R. and J.Y. Takeshita. 1969. *Family Planning in Taiwan*. Princeton, NJ: Princeton University Press.
- Garate, M.R., et. al., "Comparison between two payment models to physicians in two private family planning agencies in Peru: Final report," The Population Council, Lima, Dec. 1993.
- Garcia-Nunez, Jos,. 1992. *Improving Family Planning Evaluation: a Step-by-step Guide for Managers and Evaluators*. West Hartford, Connecticut: Kumarian Press, Inc.
- Gertler, P. and J. Molyneaux. 1994. "How Economic Development and Family Planning Programs Combined to Reduce Indonesian Fertility." *Demography*, 31(1): 33-63.
- Graham, W., W. Brass, & R.W. Snow. 1989. "Estimating Maternal Mortality." *International Family Planning Perspectives*, 20(3):125-35.

- Gray, R.H., H.D. Kalter, & P. Barass. 1990. *The Use of Verbal Autopsy Methods to Determine Selected Causes of Death in Children*. Occasional Paper No. 10. Baltimore, MD: Institute for International Programs, Johns Hopkins University.
- Grosskurth, H., F. Mosha, J. Todd, et al. 1995. "Impact of Improved Treatment of Sexually Transmitted Diseases on HIV Infection in Rural Tanzania: Randomized Control Trial." *Lancet*: 346:530-536.
- Guilkey, D. and S. Cochrane. 1994. "Zimbabwe: Determinants of Contraceptive Use at the Leading Edge of Fertility Transition in Sub-Saharan Africa." Chapel Hill, NC: Carolina Population Center.
- Hanenberg, R.S., W. Rojanapithayakorn, P. Kunasol, & D.C. Sokal. 1994. "Impact of Thailand's HIV-control Program as Indicated by the Decline of Sexually Transmitted Diseases," *Lancet*, 344:243-245.
- Hassig, S.E. 1995. Personal communication.
- Hermalin, A. 1979. "Multivariate Areal Analysis." in *United Nations Manual IX: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility*. New York: Department of International Economic and Social Affairs, pp. 97-111.
- Hermalin, A. 1975. "Regression Analysis of Areal Data." in Chandrasekaran, C. and A. Hermalin (eds.), *Measuring the Effect of Family Planning Programs on Fertility*. Dolhain, Belgium: Ordina Editions, pp. 245-99.
- Hermalin, A. 1982. "Some Cautions in the Use and Interpretation of Regression Analysis for the Evaluation of Family Planning Programs." in *United Nations Evaluation of the Impact of Family Planning Programmes on Fertility: Sources of Variance*. New York: Department of International Economic and Social Affairs, pp. 265-67.
- Jain, A. and J. Bruce, 1994. "A Reproductive Health Approach to the Objectives and Assessment of Family Planning Programs," in Sen, G. et al. (eds.), *Population Policies Reconsidered: Health, Empowerment, and Rights*, Boston, MA: Harvard University Press. pp. 103-209.
- Janowitz, B. and J. H. Bratt. 1992. "Costs of Family Planning Services: A Critique of the Literature." *International Family Planning Perspectives*, 18: 137-144.
- Janowitz, B. and J.H. Bratt. 1994. *Methods for Costing Family Planning Services*, New York: UNFPA, and Research Triangle Park, NC, Family Health International.
- Jolly, C. and J. Gribble. 1993. "The Proximate Determinants of Fertility." in Foote, K., K. Hill, & L. Martin (eds.) *Demographic Change in Sub-Saharan Africa*. Washington, DC: National Academy Press.
- Kalter, H.D., R.H. Gray, R.E. Black, & S.A. Gultiano. 1990. "Validation of Postmortem Interviews to Ascertain Selected Causes of Death in Children." *International Journal of Epidemiology*, 19:380-6.
- Koblinsky, M., K. McLaurin, P. Russell-Brown, P. Gorbach (eds.), *Final Report of the Subcommittee on Safe Pregnancy*. 1995. *Indicators for Reproductive Health Program Evaluation*. Chapel Hill, North Carolina: The EVALUATION Project, Carolina Population Center, University of North Carolina at Chapel Hill.

- Lloyd, C. and J. Ross. 1989. *Methods for Measuring the Fertility Impact of Family Planning Programs: The Experience of the Last Decade*. Research Division Working Papers, No. 7. NY: The Population Council.
- McInerney, M., & C. de la Quintana, "A comparative study of three strategies to improve the sustainability of a Bolivian family planning provider," The Population Council, La Paz, 1994.
- Mensch, B. 1995. Personal communication.
- Mensch, B., A. Jain, et al. 1994. "Assessing the Impact of Family Planning Services on Contraceptive Use in Peru: A Case Study Linking Situation Analysis Data to the DHS." Paper presented at the 1994 Annual Meeting of the Population Association of America, Miami, FL.
- Mertens, T., M. Carael, P. Sato, J. Cleland, H. Ward, & G.D. Smith. 1994. "Prevention Indicators for Evaluating the Progress of National AIDS Programmes." *AIDS* 8:1359-1369.
- Miller, R., K. Miller, L. Ndhlovu, J. Solo, & O. Achola. 1996. "A Comparison of the 1995 and 1989 Kenya Situation Analysis Study Findings," New York: The Population Council (unpublished manuscript).
- Morris, Leo. Personal communication 1995.
- Mroz, T.A., & D.K. Guilkey. 1992. "Discrete Factor Approximations for Use in Simultaneous Equation Models with Both Continuous and Discrete Endogenous Variables." Chapel Hill, NC: Carolina Population Center.
- Newman, J. 1988. "A Stochastic Dynamic Model of Fertility." In Schultz, T.P. *Research in Population Economics*. New York: JAI Press, Inc., pp. 41-68.
- Ojeda, G., R. Murad, F. Leon, "Testing pricing/payment systems to improve access and cost-recovery from NORPLANT[®]: Final Report," The Population Council, Lima, May, 1994.
- Phillips, J.F., W.S. Stinson, S. Bhatia, M. Rahman, & J. Chakraborty. 1982. "The Demographic Impact of the Family Planning-Health Services Project in Matlab, Bangladesh." *Studies in Family Planning* 13(5):131-140.
- Poston, D., & B. Chu. 1987. "Socioeconomic Development, Family Planning, and Fertility in China." *Demography*, 24(4): 531-51.
- Reinis, K. 1992. "The Impact of Proximate Determinants of Fertility: Evaluating Bongaarts' and Hobcraft and Little's Methods of Estimation." *Population Studies*, 46(2): 309-326.
- Reynolds, J. 1993. *Cost Analysis, Primary Health Care Management Advancement Programme, Module 8, Users Guide*. Washington, DC: Aga Khan Foundation and University Research Corporation.
- Reynolds, J. 1970. *A Framework for the Design of Family Planning Program Evaluation*. International Institute for the Study of Human Reproduction: New York.
- Reynolds, J. and K. Celeste Gaspari. 1985. *Cost-effective Analysis*. Chevy Chase: Primary Health Care Operations Research Project (PRICOR).

- Robey, B., S.O. Rutstein, L. Morris, & R. Blackburn. 1992. "The reproductive revolution: new survey findings." *Population Reports, Series M. No.11.*
- Rossi, P.H., & H. Freeman 1993. *Evaluation: A Systematic Approach.* Newbury Park, CA: Sage Publications.
- Rossi, P.H., J.D. Wright, & A.B. Anderson. 1983. *Handbook of Survey Research.* New York: Academic Press.
- Senderowitz, Judith, 1995. *Adolescent Health: Reassessing the Passage of Adulthood.* Bank Discussion Paper. World Bank, Washington, D.C.
- Sherris, Jacqueline D., Kathy A. London, Sidney H. Moore, John H. Pile and Walter B. Watson, 1985. "The Impact of Family Planning Programs on Fertility." *Population Reports, XIII, 1:J733-J771.*
- Stewart L. and Eckert E. (eds.), 1995, "Indicators for Adolescent Reproductive Health Services," in Tsui, A. and J. Bertrand (eds.), *Indicators for Evaluating Reproductive Health Interventions, The EVALUATION Project, Carolina Population Center, Chapel Hill, NC.*
- Suarez, E., & C. Brambila, "Cost analysis of family planning services in private family planning programs, FEMAP, Mexico: Final Report," *The Population Council, Mexico City, June 1994.*
- Tsui, A.O. and P.D. Gorbach, forthcoming in 1996. *Framing Family Planning Program Evaluation: Cause, Logic and Action.* The EVALUATION Project, University of North Carolina at Chapel Hill.
- UNICEF/UNESCO/WHO/UNFPA. *Facts for Life.* New York, NY.
- United Nations. 1982. *Evaluation of the Impact of Family Planning Programmes on Fertility: Sources of Variance.* New York: Department of International Economic and Social Affairs.
- United Nations. 1986. *Manual IX Addendum: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility.* New York: Department of International Economic and Social Affairs.
- United Nations. 1979. *Manual IX: The Methodology of Measuring the Impact of Family Planning Programmes on Fertility.* New York: Department of International Economic and Social Affairs.
- United Nations. 1985. *Studies to Enhance the Evaluation of Family Planning Programmes.* New York: Department of International Economic and Social Affairs.
- USAID (United States Agency for International Development). 1995. "The Agency's Strategic Framework and Indicators 1995/96," *Performance Measurement and Evaluation Division, Center for Development Information and Evaluation, Bureau for Policy and Program Coordination.*
- Veney, J.E. and P. Gorbach, 1993. "Definitions for Program Evaluation Terms." Chapel Hill, N.C.: EVALUATION Project (Working Paper Series No. WP-TR-01).

- Vian, T. 1993. "Analyzing Costs for Management Decisions," *Family Planning Manager* 2(2):1-18.
- Wawer, M.J., R.H. Gray, T.C. Quinn, N.K. Sewankambo, F. Wabwire-Mangen, D. Serwadda, L. Paxton, 1995. "Design and Feasibility of Population-based Mass STD Treatment, Rural Rakai District, Uganda." Paper presented at the 1995 Annual Meeting of the International Society for STD Research, New Orleans, LA, August 1995.
- Wishik, S.M., & K.H. Chen. 1973. Couple-year of Protection: a Measure of Family Planning Program Output. International Institute for the Study of Human Reproduction: New York.
- Woodhouse, G. 1995. A Guide to MLn for New Users. London: University of London, Institute of Education.
- World Health Organization, 1989. Contribution to the Working Paper of UNESCO. Compiled for the World Youth Congress, Barcelona.
- World Health Organization. 1993. Indicators to Monitor Maternal Health Goals. Report of a Technical Working Group, Geneva, Nov. 8-12: 1993.

EL MODELO MULTINIVEL EN FORMATO DE REGRESIÓN

Modelo básico de corte transversal:

El modelo básico de regresión multinivel de corte transversal se puede representar así:

$$Y_{ij} = a + b P_i + G Z_i + g X_{ij} + d P_i X_{ij} + z Z_i X_{ij} + m_i + e_{ij}$$

Donde:

Y_{ij} = la variable de resultado que interesa, medida a nivel del individuo (i.e., para el individuo j en la comunidad i). Ejemplos: probabilidad de concepción en los últimos tres años, uso actual de anticonceptivos;

P_i = variables que miden el vigor del programa en la comunidad i;

Z_i = otros determinantes a nivel de la comunidad;

X_{ij} = determinantes a nivel del individuo u hogar;

m_i = factores no observados a nivel de la comunidad (conocida también como "heterogeneidad no observada")

e_{ij} = factores no observados a nivel del individuo; y

a, b, G, g, d, z = parámetros a estimar.

Para fines de evaluación, los parámetros claves de la regresión son los coeficientes de las variables del programa b y d. Los coeficientes b miden la magnitud de los efectos directos del programa, mientras que d mide la importancia relativa de las interacciones entre las variables del programa (a nivel comunitario) y las variables del individuo u hogar.

Modelo multi-ecuacional de efectos aleatorios:

$$P_i = a + b Z_i + l m_i + e_{ij} \quad (\text{Ecuación de ubicación del programa})$$

$$Y_{ij} = a + h P_i + G Z_i + g X_{ij} + q m_i + e_{ij} \quad (\text{Ecuación de resultado})$$

donde:

P_i = variables que miden el vigor del programa en la comunidad i;

Z_i = otros determinantes a nivel de la comunidad;

Y_{ij} = la variable de resultado que interesa, medida a nivel del individuo (i.e., para el individuo j en la comunidad i). Ejemplos: probabilidad de concepción en los últimos tres años, uso actual de anticonceptivos;

X_{ij} = determinantes a nivel del individuo u hogar;

m_i = factores no observados a nivel de la comunidad (conocida también como "heterogeneidad no observada")

e_{ij} = factores no observados a nivel del individuo; y

a, b, G, h, q, g, l = parámetros a estimar.

[Nota: se han omitido de la ecuación todos los términos de interacción para simplificar la presentación]

Modelo de panel de efectos fijos:

Las ecuaciones de corte transversal para dos rondas de encuestas pueden escribirse de la siguiente manera (todos los términos de interacción para simplificar la presentación):

$$Y_{ij1} = \alpha + \beta P_{i1} + \Gamma Z_{i1} + \gamma X_{ij1} + \mu_{i1} + \varepsilon_{ij1}, \quad y$$

$$Y_{ij2} = \alpha + \beta P_{i2} + \Gamma Z_{i2} + \gamma X_{ij2} + \mu_{i2} + \varepsilon_{ij2}, \quad y$$

Donde:

Y_{ij} = la variable de resultado, cambiante en el tiempo;

P_i = variables del programa, cambiantes en el tiempo;

Z_i = características comunitarias, cambiantes en el tiempo;

X_{ij} = características del individuo, cambiantes en el tiempo;

μ_i = características comunitarias no observadas, fijas en el tiempo;

ε_{ij} = error aleatorio; y

los "1s" y "2s" se refieren a las rondas de la encuesta; y

$\alpha, \beta, \Gamma, \gamma$ = parámetros a estimar.

Por diferencia entre las dos ecuaciones obtenemos:

$$Y_{ij2} - Y_{ij1} = \alpha + \theta (P_{i2} - P_{i1}) + \xi (Z_{i2} - Z_{i1}) + \phi (X_{ij2} - X_{ij1}) + (\varepsilon_{ij2} - \varepsilon_{ij1})$$

En razón de que los parámetros fijos son invariantes en el periodo estudiado, se anulan mutuamente en la ecuación de diferencia. El parámetro "f" es el de mayor interés para propósitos de la evaluación. Este mide el aporte relativo de los cambios en las variables programáticas a los cambios observados en la variable de resultado durante el periodo estudiado⁵⁶.

⁵⁶ Nótese que cuando el panel multinivel se aplica a encuestas sucesivas en los mismos conglomerados muestrales, los cambios a nivel individual son parte de cambios agregados a nivel comunitario.

PROYECTO FINANCIADO POR USAID

Contrato Número: DPE-3060-C-00-1054-00

PROYECTO «EVALUACIÓN»

Carolina Population Center
University of North Carolina at Chapel Hill
CB 8120 University Square
Chapel Hill, NC 27516-3997 USA

INSTITUCIONES COLABORADORAS

The Futures Group International
1050 17th Street NW, Suite 1000
Washington, DC 20036 USA

Tulane University
Center for International Health and Development
School of Public Health and Tropical Medicine
1440 Canal Street, Suite 2200
New Orleans, LA 70112-2737 USA