

## **Parte I**

### **Visión General del *Compendio***

#### **A. Razón Fundamental**

1. Objetivos del *Compendio*
2. Audiencia Destinatarios
3. Organización del *Compendio*
4. Selección de Indicadores
5. Uso del *Compendio*

#### **B. El Uso de Indicadores en la Evaluación de Programas**

1. Definición de Evaluación de Programas
2. El Valor de un Marco Conceptual
3. Mediciones a Nivel del Programa frente a Mediciones a Nivel de la Población
4. Fuentes de Datos
5. Insumo, Proceso, Producto y Resultado (*Input, Process, Output, Outcome*)
6. Causalidad
7. Indicadores Cuantitativos frente a Indicadores Cualitativos
8. El Esquema de Resultados (*The Results Framework*)
9. Niveles de Presentación de Informes para las Agencias Cooperantes de USAID

## A. Razón Fundamental

La elaboración del presente *Compendio* fue impulsada por dos acontecimientos importantes de la década de los años noventa. El primero fue la Conferencia Internacional sobre Población y Desarrollo (CIPD) realizada en El Cairo, Egipto en 1994, que cambió el sentido del modelo conceptual de la salud de la mujer. Desplazando al énfasis original en planificación familiar y en objetivos demográficos (para algunos países), el nuevo paradigma enfatiza la calidad de vida, la equidad de género y los derechos humanos. La Conferencia del Cairo provocó entre la comunidad internacional la necesidad de replantear la definición de la salud reproductiva y de los servicios que estos programas deben ofrecer. Como resultado de esto, el campo de la salud reproductiva se expandió para incluir otras áreas, además de la planificación familiar, entre las que figuran: la maternidad segura, las ITS/VIH/SIDA, la nutrición de la mujer, la lactancia materna, la atención postaborto (APA), los servicios de salud reproductiva para adolescentes, la violencia contra la mujer,<sup>1</sup> la mutilación genital femenina (MGF)<sup>2</sup> y otros temas relacionados. A partir de la Conferencia del Cairo, muchos gobiernos y organizaciones no-gubernamentales (ONG) se lanzaron a afrontar el reto de ofrecer esta amplia gama de servicios, muchas veces sin tener los fondos suficientes para abordar de manera adecuada los diversos aspectos que esto implica. Las metas establecidas por la CIPD demandan la necesidad de monitorear los resultados de salud reproductiva en términos de los indicadores claramente definidos, así como de los resultados con base en la evidencia para demostrar el progreso alcanzado.

El segundo evento fue el mayor énfasis en la responsabilidad de rendir cuentas, tanto de parte de los programas nacionales como de las agencias donantes a nivel internacional. Muchos gobiernos están cansados de financiar programas que no llevan a ningún resultado, por lo que han aprovechado el concepto de la evaluación de los programas para definir cómo se pueden fortalecer estos programas con el fin de mejorar el nivel y las

condiciones de vida de sus poblaciones. En lo que concierne a las agencias donantes, tanto la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) como el Departamento Británico para el Desarrollo Internacional, el Fondo de Población de las Naciones Unidas (FNUAP), el Fondo de las Naciones Unidas para la Infancia (UNICEF) y la Organización Mundial de la Salud (OMS) figuran entre las agencias donantes que han puesto una atención particular en la evaluación de los programas, tanto para mejorarlos como para justificar los fondos que se invierten en ellos. Por otra parte, las organizaciones privadas, que abarcan desde las grandes fundaciones hasta una serie de ONG especializada, han buscado activamente los métodos adecuados para monitorear el progreso alcanzado y medir los resultados de sus esfuerzos de una forma más sistemática.

<sup>1</sup> La violencia con base en el género es un término “recogelotodo” que cubre cualquier daño inflingido a una persona contra su voluntad, y que pone en peligro la salud física y/o psicológica, el desarrollo y la identidad de la persona. La violencia es el resultado de las relaciones de poder entre los género, en las que el poder se define por el papel social que se asigna a los hombres y las mujeres y que impacta desproporcionadamente a las mujeres y a los niños en casi todas las culturas. La violencia puede ser física, sexual, psicológica, económica o sociocultural. Las categorías de perpetradores pueden incluir a los miembros de la familia, a los miembros de la comunidad y/o aquellas personas que actúan en nombre de instituciones culturales, religiosas o del Estado.

<sup>2</sup> La “mutilación genital femenina” es una traducción imprecisa de “*female genital cutting*,” que, traducido literalmente, significaría “el corte genital femenino.” Este último tiene la ventaja de ser libre de interpretación peyorativa, es decir, no conlleva prejuicios culturales, aunque, al parecer de los traductores, *corte* es menos asequible que *mutilación*. Otros posibles términos son *remoción* o *circuncisión*, según el acto al que se refiera.

Este interés en evaluar los programas desató una proliferación de iniciativas para desarrollar los indicadores para todas las áreas de la salud reproductiva. El presente volumen sigue con la tradición iniciada por el *Manual de Indicadores para Evaluar los Programas de Planificación Familiar* (Bertrand, Magnani y Knowles, 1994) y por la serie de manuales *Indicators for Reproductive Health Program Evaluation* (Bertrand y Tsui, 1995). La demanda que esos manuales han tenido ha sido grande, habiéndose distribuido hasta la fecha más de 4,500 ejemplares de cada uno. Los dos volúmenes de este *Compendio* incluyen muchos indicadores que ya han sido puestos a la prueba en el campo y están siendo ampliamente utilizados (por ejemplo, la tasa global de fecundidad, la prevalencia de anticonceptivos). Además, incluyen algunos indicadores de naturaleza experimental: las “mejores suposiciones” de los expertos que trabajan en estos temas. Durante los últimos cinco años, tanto los administradores como los evaluadores han venido experimentado con estos indicadores, particularmente en las áreas emergentes de la salud reproductiva (por ejemplo, la maternidad segura, la atención postaborto y los programas para los adolescentes). El presente volumen cosecha los frutos de dichos experimentos, ya que los mismos especialistas que han usado estos indicadores en sus labores cotidianas de evaluación son quienes han redactado o revisado los capítulos.<sup>3</sup> Como tal, el presente *Compendio de Indicadores* refleja mucha más experiencia en la práctica a nivel del campo que los volúmenes anteriores.

Pocas personas podrían afirmar que tienen experiencia en la gama completa de los diversos aspectos de la salud reproductiva, ya que este es un campo que tiene un alto grado de diversificación y de especialización. A pesar de que todos estos programas comparten el objetivo de incrementar la calidad de vida de las mujeres y la de los hombres al mejorar la condición de su salud reproductiva, cada tipo de programa tiene una manera distinta de medir el progreso alcanzado.

## 1. Objetivos del *Compendio*

El objetivo general de este *Compendio* es fomentar la evaluación de los programas y mejorar la calidad del trabajo realizado en esta área. Ofrece, para ello, un amplio listado de los indicadores más utilizados en la evaluación de los programas de salud reproductiva en los países en desarrollo. Los indicadores se han

organizado conforme a una versión revisada del marco conceptual originalmente elaborado por el Proyecto EVALUATION. Este marco describe las rutas o trayectorias por medio de las cuales los programas logran los resultados (véase la Gráfica I.1) y constituye un marco lógico para elaborar los planes de evaluación con los indicadores más adecuados. El marco originalmente fue diseñado para los programas de planificación familiar, pero puede adaptarse fácilmente a otras áreas de la salud reproductiva. Algunas secciones del *Compendio* incluyen marcos detallados que explican las rutas a seguir para lograr los efectos programáticos específicos al área en cuestión.

Algunos de los esfuerzos de evaluación en el pasado consideraban las operaciones de los programas de salud reproductiva como una “caja negra,” pero este marco conceptual indica específicamente la manera en que las personas que diseñan los programas esperan que estos trabajen para lograr los resultados, tanto a nivel del programa como a nivel de la población. Por otro lado, el esquema actual destaca los diferentes aspectos que tienen que marchar satisfactoriamente (ya sea las áreas operativas, el acceso a los servicios o la calidad de la atención) para que el programa pueda lograr el resultado final que desea alcanzar.

Los objetivos específicos de este *Compendio* son:

- Recopilar en un solo volumen un “menú” de indicadores de salud reproductiva considerados como los más útiles para evaluar los programas de salud reproductiva, tanto a nivel del programa como a nivel de la población;
- Definir los indicadores con el fin de fomentar el uso de terminología consistente entre los diversos programas, países y agencias donantes;
- Promover la evaluación de los programas, poniendo los indicadores al alcance de los evaluadores.

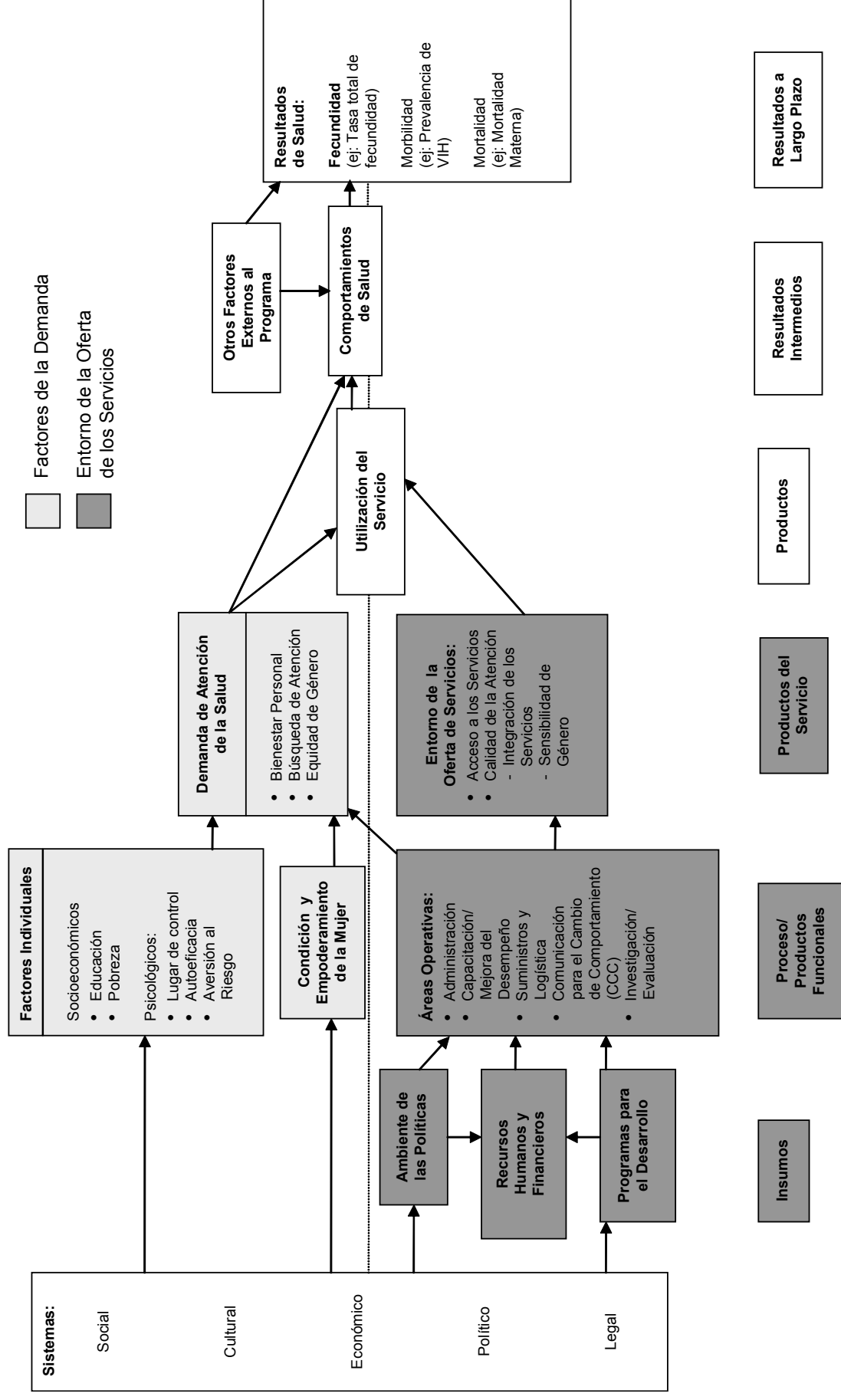
## 2. Audiencia Objetivo del *Compendio*

Son varias las audiencias que descubrirán que este *Compendio* es aplicable a sus propio trabajos, incluidos los siguientes:

---

<sup>3</sup> Esta aseveración no aplica a las Secciones III.L y III.M por ser áreas emergentes dentro de los programas de salud reproductiva.

Gráfica I.1. Marco Conceptual: Logro de Resultados en los Programas de Salud Reproductiva



- Los administradores, los gerentes de programa y los proveedores de servicios de salud reproductiva a nivel mundial;
- El personal de las agencias internacionales que trabajan en planificación familiar/salud reproductiva (PF/SR) y que tienen la responsabilidad de diseñar y evaluar los proyectos en colaboración con las contrapartes locales del país anfitrión;
- Los especialistas en evaluación en cada país que tienen la responsabilidad de monitorear el desempeño y evaluar la eficacia de los programas de SR en los ámbitos específicos;
- Las fundaciones privadas y otros donantes que apoyan los programas de SR;
- Los investigadores de las ciencias sociales (como los demógrafos, los sociólogos, los economistas), así como los epidemiólogos y otros profesionales de la salud; y
- Los estudiantes y los instructores.

### 3. Organización del *Compendio*

Este *Compendio* está organizado en tres secciones. La Parte I ofrece una visión general del documento, así como los conceptos básicos de la evaluación de los programas. El resto del volumen presenta un menú de indicadores para los distintos tipos de programas de salud reproductiva (incluidos su definición, los datos requeridos para medirlos, las fuentes de datos, el propósito y otros temas a considerar) de conformidad con el marco conceptual que se presenta en la Gráfica I.1.

La Parte II presenta los indicadores que entrecruzan las distintas áreas programáticas de la salud reproductiva, comenzando por **la condición y el empoderamiento de la mujer**, cuya influencia se difunde en todos los aspectos de la salud reproductiva. Además de los indicadores que se describen en esa sección, el papel del “género” se destaca a lo largo de todo el *Compendio* en una serie de recuadros que presentan la interpretación de otros indicadores desde la perspectiva de género. El tema del género aparece nuevamente al final de la Parte II, sección II.H.4, que presenta los indicadores para evaluar a las instituciones con relación a la equidad de género en sus estructuras administrativas y a la sensibilidad de género en su sistema de entrega de los servicios. Para finalizar, hay dos áreas programáticas de la Parte III del *Compendio* que son muy importantes para el factor

género: la participación del hombre y la violencia contra la mujer.

El resto de la Parte II aborda los factores del entorno de la oferta de los servicios de salud reproductiva. “El entorno de la oferta” se refiere a los distintos elementos que influyen en la provisión de los servicios de salud reproductiva, entre los que figuran:

- Los recursos y la infraestructura disponibles para los programas de desarrollo;
- El entorno de las políticas - la voluntad política, las políticas legales y las políticas regulativas;
- Las áreas funcionales (u operativas) que dan apoyo a la provisión de los servicios; y
- El entorno de la oferta de los servicios, que se caracteriza por el acceso a los servicios y la calidad de la atención.

La Parte II.B resume los indicadores para evaluar las iniciativas para influir en **el entorno político**. Las cinco secciones subsiguientes abordan las áreas operativas de los programas de SR (**administración, capacitación, suministros y logística, comunicación para el cambio de comportamiento e investigación/evaluación**<sup>4</sup>) que son las que, en conjunto, configuran e influyen en el entorno de la oferta de los servicios. El entorno de la oferta se puede evaluar directamente en términos del **acceso a los servicios** y de la **calidad de la atención**, junto con dos elementos asociados a la calidad: **la integración de los servicios y la equidad/sensibilidad de género**.

La Parte III presenta algunos indicadores específicos para las principales áreas programáticas de la salud reproductiva, comenzando por las tres áreas que han recibido la mayor atención de la comunidad internacional de salud reproductiva: **la planificación**

---

<sup>4</sup> El *Manual de Indicadores para la Evaluación de los Programas de Planificación Familiar* (1994) incluía indicadores para el monitoreo y la evaluación (M y E). Sin embargo, las autoras consideran que estos indicadores ya no se están utilizando en la práctica. En cambio, en los últimos dos años se han observado avances importantes en la elaboración de indicadores para una área específica de investigación aplicada: la investigación operativa (IO). Por lo tanto, este *Compendio* ha optado por incluir indicadores para la Investigación Operativa (IO) en lugar de los indicadores para M y E.

**familiar, las ITS/VIH/SIDA y la maternidad segura.** A esta última le sigue una área relacionada, pero relativamente nueva dentro de la evaluación de los programas: la salud del recién nacido. Luego vienen la **nutrición de la mujer y la lactancia materna.** A partir de la Conferencia del Cairo, hay otras áreas que incursionaron en el ámbito de la SR, entre ellas: **los programas para los adolescentes, la atención postaborto (APA), la participación del hombre, la violencia contra la mujer, la mutilación genital femenina y la salud reproductiva en situaciones de emergencia.**

Existen otras áreas que también son importantes, pero que han quedado excluidas del *Compendio* porque no han sido el eje principal de ningún programa grande a nivel nacional, salvo en casos aislados o en números muy limitados (por ejemplo, la infertilidad, la salud sexual y las pruebas de tamizaje para el cáncer cervical).

#### 4. Selección de los Indicadores

El proceso para seleccionar los indicadores suele desconcertar a las personas que tienen poca experiencia en evaluación. El apéndice B presenta algunas recomendaciones sobre los pasos para seleccionar los indicadores, entre las que figuran las siguientes:

- Exponer claramente los resultados deseados (es decir, los objetivos a alcanzar);
- Desarrollar una lista de posibles indicadores;
- Valorar cada posible indicador; y
- Seleccionar los “mejores” indicadores.

Para seleccionar los indicadores que se incluyen en este *Compendio* utilizamos los siguientes criterios:

- La validez (el indicador mide lo que pretende medir);
- La importancia (la medida logra captar algo “que hace una diferencia” para la eficacia del programa);
- La utilidad (los resultados señalan las áreas que se deben mejorar); y
- La factibilidad (los datos se pueden recopilar con un esfuerzo y a un costo razonables);

En la medida de lo posible, seleccionamos los indicadores que ya han sido puestos a la prueba en el campo, incluidos los que se miden en las Encuestas de

Demografía y Salud (DHS) y en las Encuestas de Salud Reproductiva (RHS).<sup>5</sup> Para algunas de las áreas emergentes (por ejemplo, la violencia contra la mujer, o la mutilación genital femenina), incluimos también algunos indicadores que consituyen las recomendaciones de los profesionales que trabajan en estos campos, con el fin de estimular el diálogo sobre la evaluación de estas áreas y proporcionar los posibles indicadores a ser puestos a la prueba a medida que estos programas se vayan volviendo más comunes.

Dos temas notables por su ausencia son: la capacidad técnica y la sustentabilidad. A pesar de que ambos temas son muy importantes en el campo de la salud reproductiva, no pudimos encontrar ningún conjunto de indicadores que ya hubieran sido puestos a la prueba y que estuvieran siendo utilizados en la práctica para monitorear la capacidad técnica y la sustentabilidad de los programas<sup>6</sup> (aunque los indicadores del entorno político, de la capacitación y de la administración abordan algunos aspectos del fortalecimiento de la capacidad institucional). Un grupo de trabajo que está analizando los indicadores para fortalecer la capacidad institucional bajo el auspicio del Proyecto MEASURE *Evaluation* concluyó que es indispensable que estos indicadores sean específicos para cada contexto. El grupo desarrolló una serie de guías que ayudan a las instituciones a estructurar el tema y a identificar los medios adecuados para evaluar el proceso del fortalecimiento de la capacidad institucional dentro de sus propias organizaciones (Lafond y Brown, 2003).

Este manual no aborda la evaluación especializada de los programas en lo que se refiere a los costos (por ejemplo, el análisis de los costos, el análisis del costo-efectivo). Los evaluadores pueden utilizar una serie de indicadores para establecer el costo por unidad para los productos de un servicio específico (particularmente los que sirven para medir los productos inmediatos, como el índice años de protección pareja, el número de mujeres embarazadas a las que se les hace una prueba y

<sup>5</sup> Este tipo de encuestas se discuten en mayor detalle en la Parte I.B.3.

<sup>6</sup> Varios de los colaboradores del *Compendio* citaron algunos trabajos que están en marcha o sugirieron indicadores para medir la sustentabilidad, pero se optó por no incluir los indicadores que están aún en fase de desarrollo, ya que nuestro enfoque es citar únicamente los indicadores que ya se están utilizando en la práctica.

reciben consejería para el VIH, el número de consultas en los servicios dirigidos a los hombres y otros más). Sin embargo, la metodología para el análisis de los costos y el análisis del costo-eficacia es lo suficientemente técnica, por lo que optamos por referir a los lectores interesados a otras fuentes especializadas (Janowitz y Bratt, 1994; Kumaranayake et. al., 1998). A pesar de que en este manual no abordamos este tema debido a la metodología especializada que requiere, recomendamos a los evaluadores hacer un análisis de los costos y del costo-efectivo en sus respectivos programas.

Por la manera en que están redactados muchos de los indicadores del *Compendio*, no se especifica el sexo de la persona de quien se habla; con el fin de simplificar, nos referimos a “audiencias, participantes, capacitadores, proveedores, adolescentes y otros”, sin especificar si estos son hombres o mujeres. A la vez, el texto que describe los indicadores casi siempre implica que la cliente es una mujer, lo que es frecuente pero no necesariamente siempre el caso. Hemos retenido la palabra “ella” para reflejar la realidad, en el sentido de que la gran mayoría de clientes para los servicios de SR en los países en desarrollo son mujeres, a pesar de los esfuerzos por incluir a los hombres en las iniciativas de SR. Sin embargo, al aplicar estos indicadores, se recomienda a los usuarios del *Compendio* especificar, en la medida que sea posible, el sexo de la persona de quien se está hablando.

## 5. Uso del *Compendio*

Este *Compendio* sirve de documento de consulta para ser utilizado a través de toda la comunidad que trabaja en salud reproductiva a nivel internacional. Aunque ha sido financiado por la USAID, se puede aplicar también en los programas de salud reproductiva que reciben el auspicio de cualquier agencia financiera, gobierno y ONG a nivel mundial. Concretamente, el *Compendio* proporciona un menú de indicadores para ser utilizados selectivamente conforme a las necesidades de evaluación de los programas a nivel nacional, a nivel regional o a nivel de los proyectos locales. Queremos enfatizar que ningún programa o proyecto debe pretender usar *todos* los indicadores descritos en el *Compendio*. A decir verdad, para los fines del monitoreo rutinario se recomienda seleccionar unos cuantos indicadores que sean relevantes para los objetivos del programa y que sean fáciles de recopilar y de interpretar. Si las instituciones necesitan mayor retroalimentación, pueden hacer estudios especiales para evaluar el desempeño de los programas en las áreas que

más interesen a su personal administrativo. En este caso, los estudios especiales deben ser de diversa naturaleza y distinta periodicidad para minimizar la carga de investigación en las instituciones.

Las descripciones de los indicadores se han diseñado con el fin de promover la estandarización de las definiciones y de los conceptos entre la comunidad internacional que trabaja en salud reproductiva. Sin embargo, aunque es útil estandarizar, se recomienda a las organizaciones adaptar los indicadores a sus circunstancias específicas. Este enfoque garantiza la aplicabilidad de los indicadores para la organización que los está utilizando y promueve un sentido de propiedad del proceso del monitoreo y de la evaluación. No obstante, las organizaciones que decidan adaptar (o modificar) los indicadores, deben expresar con claridad las nuevas definiciones y los métodos de la medición.

Al sentirse abrumados por el volumen de los indicadores que presenta este *Compendio*, muchos lectores se preguntarán: “**¿Cuáles son los diez o quince indicadores clave, que resultan esenciales para evaluar los programas de salud reproductiva a nivel mundial?**” Efectivamente, la Organización Mundial de la Salud (OMS) se planteó la misma pregunta y concluyó con una lista consensuada de 17 indicadores para medir los resultados, que han sido propuestos a partir de la CIPD del Cairo de 1994 para rastrear el progreso de los programas locales de salud reproductiva a nivel mundial (véase la Parte III.A). Estos indicadores son útiles para que los gobiernos y las agencias donantes internacionales puedan medir “el panorama general”.

Sin embargo, es probable que la gran mayoría de los gerentes de los programas consideren que este listado global no es práctico, por varias razones, para evaluar sus intervenciones específicas. Primera, porque los indicadores globales utilizan las encuestas demográficas, las cuales están fuera del campo de acción de las organizaciones que trabajan en una región específica o de las ONGs que trabajan con algunos subgrupos de población específicos (como los pilotos de camión y las trabajadoras del sexo para la prevención del VIH). Segunda, porque muchas organizaciones quieren que la evaluación cubra no sólo los resultados a largo plazo, sino que también el progreso que logran las distintas áreas operativas, como la capacitación, los suministros, la logística o la comunicación para el cambio de comportamiento (CCC). Los indicadores más apropiados para una



evaluación específica dependen directamente del propósito de la misma.

Por ejemplo, el Director de Suministros y Logística de un programa nacional de planificación familiar probablemente rastree los ocho indicadores cuantitativos de los suministros y logística, además del indicador para evaluar los sistemas logísticos (que se describen en la Parte II.E). Por otra parte, el Director del Programa Nacional de Planificación Familiar seguramente se concentrará en los resultados alcanzados en términos de la prevalencia anticonceptiva, de la calidad de la atención o de otros indicadores más generales que reflejan los logros de las áreas operativas en conjunto.

Este *Compendio* tiene el propósito de mejorar la evaluación de los programas para fortalecer la oferta de servicios de SR en los programas a nivel nacional (al igual que en los programas regionales o en los proyectos específicos de un país) y su fin no es generar una libreta de calificaciones que enfrente a un país contra otro al considerar el financiamiento futuro. Aunque se recomienda el uso consistente de los términos y de las definiciones a través de todos los países o de todos los programas, el contexto socioeconómico y cultural en el que cada uno opera es muy diferente de un país a otro o de una región a otra. Por ejemplo, una campaña en los medios de comunicación logra cierto nivel de cobertura con mucha más facilidad en un país pequeño con una población monolingüe, con un sistema de comunicación avanzado y con un entorno político favorable que en un país en el que estas condiciones no existen.

Aunque es inevitable comparar algunas variables entre los distintos países y hasta podría ser útil hacerlo en casos específicos, los resultados que estos indicadores producen para evaluar la SR podrían prestarse a ser mal utilizados si los evaluadores y los administradores de los programas no los interpretan en el contexto específico de cada país. En la mayoría de casos, resultará más productivo comparar los indicadores para un programa específico a lo largo del tiempo que utilizar los indicadores para hacer comparaciones entre los distintos países.

## **B. El Uso de Indicadores en la Evaluación de los Programas**

### **1. Definición de la Evaluación de los Programas**

El término “evaluación de los programas,” tal como se usa en el presente *Compendio*, abarca (1) el monitoreo rutinario y (2) la evaluación de los distintos componentes: los procesos, los resultados y el impacto. El término “evaluación,” tal como se utiliza en este manual, podría referirse a cualquier aspecto de la evaluación de los programas.

El monitoreo rutinario y la evaluación de los procesos miden qué tan bien está trabajando un programa (Adamchak et al., 2000). En cambio, la evaluación de los resultados y el impacto se relacionan con los cambios deseados (tal como se establece en los objetivos del programa). Para ser más específicos:

El **monitoreo** es el seguimiento o el rastreo rutinario de las actividades de un programa por medio de mediciones regulares y progresivas, para saber si se están realizando las actividades programadas. Los resultados revelan si las actividades se están implementando conforme al plan e indican el grado en que se utilizan los servicios del programa (Adamchak et al., 2000).

La **evaluación de los procesos** mide qué tan bien se están realizando las actividades del programa. Esta información se suele recopilar rutinariamente; por ejemplo, por medio de los informes del personal, pero también se puede recopilar periódicamente como parte de un esfuerzo más grande para evaluar los procesos (por ejemplo, por medio de los estudios especiales) que pueden incluir el uso de grupos focales u otros métodos de investigación cualitativa. La evaluación de los procesos se utiliza para medir la calidad de la implementación del programa y para valorar la cobertura del mismo; también puede servir para medir el grado en que la audiencia a la cual están dirigidos los servicios de un programa, los utiliza (Adamchak et al., 2000).

La **evaluación de los resultados** mide el grado en que ocurre un cambio en coincidencia con los objetivos del programa. Muchas evaluaciones centran su atención en cambios en los conocimientos, en las actitudes y en los comportamientos de (1) los clientes/los participantes



del programa o (2) de los miembros de una audiencia específica dentro de la población total (por ejemplo, las mujeres en edad reproductiva). Sin embargo, los “resultados” también se pueden referir a cambios en las políticas de salud reproductiva, cambios en los procedimientos administrativos, cambios en los sistemas de logística, cambios en la calidad de la atención de los servicios de salud y cambios en otros aspectos relacionados con el entorno de la oferta de los servicios.

La evaluación o **valoración del impacto** no solamente mide el cambio ocurrido, sino el grado en que dicho cambio es atribuible a la intervención del programa. Este tipo de evaluación está bastante restringida aún en los programas internacionales de salud reproductiva; pero, por lo general, abarca (1) los experimentos pequeños a nivel de las clínicas para determinar la eficacia de una estrategia específica o (2) los análisis con regresiones multivariable multinivel de encuestas a gran escala, basadas en los datos de las Encuestas de Demografía y Salud (DHS), para determinar la importancia relativa de distintos factores, incluida la intervención del programa.

## 2. El Valor de un Marco Conceptual

El marco conceptual que ilustra la Gráfica I.1 se adaptó de un modelo similar elaborado bajo el Proyecto EVALUATION para la planificación familiar. Este marco ilustra las rutas o trayectorias a través de las cuales los programas de salud reproductiva logran sus objetivos. La columna de la izquierda define el contexto en el cual opera un programa; es decir, el sistema social, cultural, económico, político y legal de una sociedad específica, incluidos sus programas de salud reproductiva. El lado superior izquierdo de la gráfica, sombreado en un tono claro, muestra el papel que juegan los factores de la **demand**a en la eficacia del programa. En los países en los que la población demanda los servicios (o sea, que tienen una “demanda alta”) decididamente basados las normas y las preferencias sociales, será mucho más fácil lograr los resultados que en aquellos en los que la población se muestra indiferente o negativa hacia el programa.

El lado inferior izquierdo del marco conceptual, sombreado en un tono más oscuro, describe los factores del **entorno de la oferta** de los servicios. Los países que cuentan con programas robustos para el desarrollo social y económico proporcionan un entorno más propicio para promover la salud reproductiva que aquellos que no cuentan con sistemas de apoyo para

este tipo de esfuerzos. El compromiso político (la “voluntad política”) hacia un programa también facilita su implementación, tal como lo demuestra el programa de planificación familiar de Bangladesh. Las agencias donantes y los gerentes de los programas originalmente consideraban el entorno político como una variable del contexto que podría tener cierta influencia en la implementación de un programa, pero hoy día participan activamente en el diseño de las intervenciones (por ejemplo, promoción y defensa pública), con el propósito de configurar el entorno político.

El entorno de la oferta abarca también las áreas operativas que brindan el apoyo necesario para la provisión de los servicios y el entorno mismo. Las áreas operativas, o áreas funcionales de un programa, proporcionan la estructura necesaria para llevar a cabo las intervenciones e incluyen la administración, la capacitación, los aspectos logísticos, la CCC y la investigación/evaluación. De hecho, tanto la USAID como otras agencias donantes financian programas completos que fortalecen las operaciones de las agencias que trabajan en estas áreas en los países en desarrollo. Estas áreas operativas contribuyen directamente al entorno de la oferta de los servicios, es decir, de los servicios que están disponibles para un cliente potencial en un país específico. La medición del entorno de la oferta centra su atención en el acceso a los servicios y en la calidad de la atención, así como en dos elementos relacionados con la calidad: la integración de los servicios y la equidad/sensibilidad de género.

Estos dos grupos de factores en conjunto—la oferta y la demanda—determinan el nivel de utilización de los servicios de un país. La utilización de los servicios no es esencial para la práctica de conductas específicas (por ejemplo, la lactancia materna exclusiva), pero suele jugar un papel importante para ayudar a un cliente a adoptar, por medio de la información, comportamientos saludables de la consejería, de los suministros (por ejemplo, los condones para la prevención del SIDA) o de los procedimientos clínicos (por ejemplo, la inserción de un DIU).

La casilla que se titula “comportamientos de salud” representa el objetivo de la mayor parte de los programas de salud reproductiva, es decir, los comportamientos que se recomiendan adoptar a los miembros de la audiencia objetivo del programa. Entre los ejemplos

se puede citar el uso de métodos anticonceptivos para la planificación familiar, el uso de condones o la reducción de las parejas sexuales para la prevención del SIDA, la atención del parto por un proveedor calificado y la lactancia materna. Es importante reconocer que hay factores ajenos al programa que también pueden jugar un papel a este nivel, influyendo tanto en los comportamientos de salud como en los resultados del programa. Por ejemplo, las mujeres están más expuestas que los hombres a contraer el VIH por una pareja infectada. La fecundidad se determina no sólo por la práctica de la anticoncepción, sino por la edad al casarse, el nivel de abortos inducidos, la infertilidad postparto y la esterilidad patológica. La cadena completa de eventos causales que se describen en la Gráfica I.1, los cuales conducen a la adopción de los comportamientos de salud específicos, tienen un efecto directo en el objetivo fundamental de los programas de salud reproductiva: mejorar los resultados de salud en términos de la fecundidad, la mortalidad y la morbilidad.

### **3. Mediciones con Datos a Nivel del Programa frente a Mediciones con Datos a Nivel de la Población**

Para los propósitos de la evaluación de los programas, es importante distinguir entre dos tipos de datos: los datos a nivel del programa y los datos a nivel de la población.

Los datos a nivel del programa consisten en la información disponible de las fuentes del programa (por ejemplo, de los expedientes administrativos, de los expedientes de los clientes, de las estadísticas del servicio) o en información que se puede obtener en el lugar donde se hace la evaluación (por ejemplo, la observación directa, la interacción entre el cliente y el proveedor, las entrevistas de salida con clientes o las encuestas con clientes o con observadores incógnitos), aunque la principal fuente de datos del programa la constituyen los sistemas de información rutinaria de salud. Así también, los estudios de seguimiento a los clientes que acuden a una clínica constituyen parte de la información del programa, ya que la información sobre los clientes proviene de los expedientes del programa. Algunos de los datos “a nivel del programa” corresponden a una pequeña red de establecimientos que ofrecen un servicio especializado, aunque el término “a nivel del programa” también se puede referir a los programas que tienen un alcance a nivel nacional.

La información al nivel del programa es muy importante para entender el desempeño de los programas y el tipo de resultados que estos logran (por ejemplo, el número de consultas mensuales en una clínica, el número de dosis de tétanos administradas a las mujeres embarazadas). Sin embargo, los datos a nivel del programa no reflejan el nivel de cobertura de dichos programas (a menos que se calcule un denominador para el área de captación que convierta las estadísticas del programa en la estimación de una tasa). Por otra parte, los datos de los participantes de un programa tienen un sesgo potencial de selectividad (no reflejan la situación de la población en general); es decir, las personas que deciden participar en un programa son muchas veces distintas a la población en general. Las ONG tienden a evaluar sus proyectos únicamente con datos a nivel del programa, ya que no pretenden una cobertura de la población en general a nivel nacional, ni siquiera en el área específica en la que trabajan.

En cambio, los programas gubernamentales que están diseñados para tener una cobertura a nivel nacional se evalúan en términos del efecto que tienen en el público en general. El término “a nivel de la población” se puede referir también a una región geográfica pequeña (por ejemplo, el área de captación de un proyecto de demostración o proyecto piloto, como un distrito regional), siempre que los datos se obtengan de una muestra representativa de la población. Las principales fuentes de datos para las actividades de salud reproductiva a nivel de la población son las DHS+ que conduce ORC Macro International y las RHS que conducen los Centros de Control y Prevención de Enfermedades (CDC). Además, algunos países tienen datos disponibles porque han realizado encuestas a nivel nacional, ya sean financiadas por el gobierno local o por otras agencias internacionales (por ejemplo, la Encuesta de Fecundidad de la India). Las encuestas DHS+ y RHS son particularmente útiles para medir los factores de la demanda de los servicios. Los módulos de la Evaluación de la Provisión de Servicios (EPS) de las Encuestas DHS+ miden los factores del entorno de la oferta de los servicios, pero no han sido aplicados con tanta frecuencia como las encuestas de los hogares. (Para mayores detalles sobre las encuestas a nivel de los establecimientos, véase la Parte II.H).

#### 4. Las Fuentes de Datos

Las descripciones de los indicadores que se presentan a continuación contienen una amplia variedad de fuentes de datos. En algunos casos, las distintas áreas de un mismo programa utilizan diferentes términos para referirse a conceptos similares; además, los términos pueden variar entre un país y otro. Por ejemplo, los programas de planificación familiar suelen utilizar el término “estadísticas del servicio” para referirse a los datos que el establecimiento mantiene con relación al número de clientes y al volumen de los anticonceptivos distribuidos. En cambio, las personas que trabajan en los programas de salud materna suelen utilizar los términos “registros del establecimiento,” “expedientes clínicos,” “datos de los servicios de salud” y otros términos relacionados.

Para ayudar a los lectores a comprender la gran diversidad de términos utilizados para describir las múltiples fuentes de datos disponibles, resumimos en el Cuadro I.1 las principales fuentes, así como las categorías específicas dentro de cada una de estas fuentes. Los términos “sistemas de información gerencial” (SIG) o “sistemas de información de la salud” (SIS) se refieren a distintos tipos de información, que, por lo general, consisten en ciertos subconjuntos de los datos que se incluyen como “datos a nivel del programa” en el Cuadro I.1.

##### **Cuadro I.1. Principales Fuentes de Datos Utilizadas para Evaluar los Programas**

###### **Datos a nivel del programa:**

Las estadísticas del servicio (también conocidas como estadísticas rutinarias de los servicios de salud o datos de los servicios de salud):

- Los expedientes de los ingresos (por diagnóstico)
- Las entrevistas con las personas que ingresan al programa
- Los expedientes de los clientes
- Los registros o la base de datos en los que se lleva el control del número de visitas, de los clientes, de los productos distribuidos
- Los expedientes clínicos y los expedientes de los servicios específicos (control prenatal, sala de emergencias, sala de operaciones, maternidad)
- Los resultados de las pruebas de laboratorio
- Los datos de la vigilancia de las enfermedades en las clínicas
- Los datos de la vigilancia centinela en las clínicas
- Los expedientes de las causas de defunción

Las encuestas en los establecimientos:

- Las auditorías a los inventarios del equipo y los suministros del establecimiento
- Las entrevistas de salida con los clientes, el nivel de satisfacción de los clientes
- La observación directa de la interacción entre el cliente y el proveedor, la observación a los capacitadores
- Las visitas al lugar donde se manejan los aspectos logísticos
- La revisión de los expedientes o de los registros del servicio

Las encuestas locales:

- Las encuestas de vigilancia conductual con relación a las ITS/VIH
- Los estudios prospectivos a nivel de la comunidad

Las encuestas y la opinión de los proveedores y de otros miembros del personal:

- La calificación del desempeño (por lo general basada en listas de verificación o de chequeo)
- La calificación *pre-* y *pos-test* para las actividades de capacitación
- Los exámenes para medir la competencia (utilizados en conjunto con los cursos de capacitación)
- La evaluación de los cursos (por ejemplo, de capacitación)
- Las evaluaciones por un evaluador externo
- La observación de la interacción entre el cliente y el proveedor
- Las entrevistas con el personal administrativo; las encuestas al personal
- Las herramientas de autoevaluación
- La cliente simulada

Los registros del programa (además de las estadísticas del servicio):

- Las nóminas del personal
- Los informes financieros
- Las descripciones de los puestos
- Las minutas oficiales de las reuniones
- El listado de los participantes a un evento (por ejemplo, de capacitación)
- Los datos de logística: niveles de existencias, proyecciones para los productos (necesarios para determinado programa), los planes para la adquisición de los productos, listados de los productos, calendarización programada para los envíos
- Los expedientes de la supervisión, los expedientes del personal administrativo
- Los documentos del proyecto (por ejemplo, los informes finales)
- Los protocolos para la provisión de los servicios, los estándares médicos y de enfermería
- Los registros del uso de o del acceso a Internet (generados por un servidor de la red)

### **Los datos a nivel de la población:**

Los censos (para definir la población del área de captación)

Los registros de las estadísticas vitales:

- Registros de los nacimientos
- Registros de las defunciones (por causa de la muerte)

Las encuestas con una muestra representativa a nivel nacional:

- Las Encuestas de Demografía y Salud (DHS) coordinadas por ORC Macro (incluidos los cuestionarios principales y los módulos opcionales)
- Las Encuestas de Salud Reproductiva (RHS) dirigidas por el Centro de Control y Prevención de Enfermedades (CDC)
- Las Encuestas de Conglomerados con Indicadores Múltiples apoyadas por la UNICEF
- Otras encuestas adicionales (por ejemplo, U.S. Bureau of the Census)
- Las encuestas nacionales realizadas independientemente de las DHS y las RHS

Las encuestas con muestras representativas de la población de interés:

- Los estudios de línea basal y los estudios de seguimiento para fines de evaluación
- Las encuestas de vigilancia de los comportamientos usadas con relación al SIDA
- Las encuestas de opinión

Los documentos del gobierno y los registros oficiales:

- Las políticas, los planes y las guías oficiales
- Las leyes, las regulaciones, los mandatos especiales, las resoluciones judiciales
- Los códigos de la Salud y los códigos penales
- Los presupuestos de los gastos nacionales, las cuentas, las partidas, las facturas
- Las minutas de las reuniones oficiales, las agendas, el listado de los participantes
- Los listados de membresía de las redes o de las coaliciones
- Los expedientes o los registros de las votaciones
- Los expedientes de la policía (por ejemplo, para la violencia de género)

#### **Estudios especiales:**<sup>7</sup>

- El Índice de Esfuerzo del Programa (por ejemplo, para la planificación familiar, el SIDA y la atención materna y neonatal)
- El análisis de tiempo y movimiento (*Time motion studies*)
- Los archivos de los medios de comunicación
- Las encuestas de opinión a los usuarios de Internet “en-línea”
- Las herramientas de autoevaluación
- El análisis de contenidos

#### **Métodos cualitativos (Por lo general se aplican a los clientes y a otra población no representativa)**

- Los grupos focales
- Las entrevistas en profundidad
- Los “*pile sorts*”
- La observación etnográfica

---

<sup>7</sup> Esta categoría no es mutuamente excluyente de las anteriores. Por ejemplo, cualquiera de las encuestas con clientes, proveedores, oficiales de gobierno u otros podría considerarse como un estudio especial. Sin embargo, los elementos que se describen bajo este encabezado no caen dentro de ninguna otra categoría de la lista.

## 5. Los Insumos, los Procesos, los Productos y los Resultados

En los primeros manuales desarrollados por el Proyecto EVALUATION, clasificamos los insumos, los procesos y los productos como información a **nivel del programa**; sin embargo, clasificamos los resultados como información a **nivel de la población**. Este enfoque es útil para evaluar los programas a nivel nacional, como en el caso del programa nacional de planificación familiar. Sin embargo, esta misma clasificación viene siendo menos útil (particularmente, el uso del término “resultados [*outcome*]”) para evaluar ciertas áreas operativas (como la capacitación, la logística y la CCC).

Por ejemplo, un objetivo común de los programas de capacitación suele ser el mejoramiento de la calidad de atención en el entorno de la oferta de los servicios. Los esfuerzos de todas las actividades de capacitación contribuirán, en conjunto, al logro de los resultados a nivel nacional (por ejemplo, mayor prevalencia de anticonceptivos, mayor número de mujeres con partos atendidos por un proveedor calificado), pero el efecto más directo y apto a ser medido es el mejoramiento de la calidad de los servicios. En este sentido, el resultado deseado para una serie de eventos de capacitación es la buena calidad de la atención en un sistema específico de establecimientos. Estos resultados no son a nivel de la población, pero representan un punto de conclusión apropiado para medir y evaluar los programas de capacitación. Por lo tanto, en el Cuadro I.2 definimos los insumos, los procesos, los productos (*outputs*) y los resultados (*outcomes*), advirtiendo que los “resultados deseados” para las áreas funcionales (administración, capacitación, logística y CCC) pueden medirse adecuadamente a nivel del programa.

El *Manual de Indicadores para Evaluar los Programas de Planificación Familiar* de 1994 utilizó una terminología ligeramente distinta, ya que se clasificaron a los resultados intermedios como “efectos” (por ejemplo, cambios en los comportamientos o en las prácticas) y a los resultados a largo plazo como “el impacto” (por ejemplo, cambios en las tasas de fecundidad o en las tasas de mortalidad). Sin embargo, el término “impacto” genera bastante confusión, ya que, para muchos evaluadores, “el análisis del impacto” implica la necesidad de demostrar la causalidad. Por esta razón, recomendamos el uso de los términos “resultados intermedios” (*intermediate outcomes*) y “resultados a largo plazo” (*long-term outcomes*), ya que estos no implican causalidad y (particularmente en el caso de los resultados a largo plazo) suelen estar sujetos a factores sociales y económicos que están más allá del control del programa.

## Cuadro I.2. Definiciones de Insumos, Proceso, Producto y Resultados

**Los Insumos (“inputs” en inglés)** son todos los recursos humanos y financieros, las instalaciones físicas, el equipo y las políticas operativas que hacen posible implementar las actividades del programa.

**El Proceso (“process”)** se refiere a las actividades múltiples que se realizan para lograr los objetivos del programa. Abarca no sólo lo que se hace, sino qué tan bien se hace.

En general, un nivel alto de insumos se refleja en una implementación satisfactoria del programa; sin embargo, teóricamente hablando, es posible tener un nivel alto de insumos, pero un programa insatisfactorio (por ejemplo, si un administrador de alto nivel que se opone a la planificación familiar logra bloquear la oferta de servicios en los establecimientos que están bajo su control). En cambio, existen numerosos ejemplos de la vida real en todo el mundo, en los que el personal administrativo de un programa trabaja con recursos muy limitados, pero se esfuerza por hacer el mejor trabajo posible bajo tales circunstancias.

**Los Productos (“outputs”)** son los resultados de estos esfuerzos a nivel del programa. A los gerentes de los programas a nivel del campo les interesan las tendencias nacionales que muestran los frutos de sus esfuerzos (como la prevalencia de anticonceptivos para la planificación familiar, la prevalencia de la lactancia materna para la promoción de la lactancia materna); sin embargo, tienden a limitar la evaluación de sus propias actividades a mediciones a nivel del programa, particularmente a medir los productos. Existen dos tipos de productos que aparecen en apartados separados en la Gráfica I.1. Estos son: los productos del servicio (que miden la idoneidad del sistema para proveer los servicios) y la utilización del servicio (que mide el nivel en que los clientes utilizan los servicios).

**Los Resultados (“outcomes”)**, por lo general, se refieren a los efectos del programa a nivel de la población.<sup>8</sup> La evaluación de los resultados mide el efecto que el programa ha tenido en la población de un área de captación específica (por ejemplo, todas las mujeres en edad reproductiva en determinado país). Es importante distinguir entre dos clases de resultados: los resultados intermedios y los resultados a largo plazo. Los resultados intermedios tienden a referirse a las conductas o las prácticas específicas por parte de la audiencia objetivo—tales como el uso de anticonceptivos, la lactancia materna, el uso de condones, el consumo de suplementos micronutrientes—que afectan el resultado deseado a largo plazo (por ejemplo, la reducción de la mortalidad, la morbilidad o la fecundidad). Los resultados a largo plazo se refieren a los efectos anticipados del programa a largo plazo (un cambio en las tasas de morbilidad, mortalidad o fecundidad). Sin embargo, el resultado a largo plazo casi siempre está sujeto a la influencia de factores externos al programa, incluidas las condiciones socioeconómicas y la condición de la mujer en un país específico.

---

<sup>8</sup>Nótese la excepción que se discute en el texto anterior, en el sentido de que los “resultados deseados” para las áreas operativas no necesariamente tienen que medirse a nivel de la población.



Los evaluadores suelen reportar los hallazgos de una evaluación a nivel del programa para la población como un todo, pero deberían desagregar los resultados por regiones o por subgrupos, cuando esto sea apropiado. Por ejemplo, un programa que logra sus resultados ofreciendo servicios a los residentes privilegiados de las áreas urbanas habrá logrado menos que uno que se orienta a las personas de escasos recursos que residen en las áreas rurales. De igual manera, cuando hay datos disponibles y el *n* es lo suficientemente grande, los evaluadores podrían considerar la posibilidad de desagregar los datos por factores étnicos, por factores geográficos o por factores de otra índole socioeconómica pertinente. Reducir las desigualdades entre ciertas regiones y entre ciertos subgrupos podría ser un objetivo programático que va más allá de mejorar el promedio de la población en general. La evaluación de este tipo de objetivos requiere la desagregación de los datos y hacer las mediciones por subgrupos.

Las intervenciones de salud reproductiva suelen girar alrededor de algunas prácticas que están profundamente arraigadas y que probablemente se logren cambiar a lo largo de un período de tiempo bastante prolongado. Por ejemplo, la inaudita baja en las tasas de fecundidad a finales del siglo XX a nivel mundial ocurrió a una velocidad récord y, sin embargo, tardó entre 20 y 30 años en llevarse a cabo. Desde la perspectiva del programa, resulta impráctico reportar anualmente sobre los indicadores para resultados a largo plazo, aunque estos “realmente importen” (por ejemplo, cambios en la mortalidad y en la fecundidad). En lugar de esto, la evaluación de los programas tiende a concentrarse en los resultados intermedios (que también se conocen como efectos), ya que están más estrechamente vinculados a los esfuerzos del programa y muestran cambios en períodos de tiempo más cortos. En el caso de la planificación familiar, el resultado intermedio que se reporta con mayor frecuencia es la prevalencia de anticonceptivos (es decir, el porcentaje de mujeres en edad reproductiva que utilizan actualmente algún método anticonceptivo). Los programas para prevenir el VIH suelen evaluar “el uso de condones durante la última relación sexual” entre los miembros de la población en riesgo. En el caso de los programas para eliminar la mutilación genital femenina, un resultado intermedio es el porcentaje de la población favorable a que dicha práctica continúe.

## 6. La causalidad

La mayoría de los programas tienen como objetivo lograr algún tipo de cambio, ya sea a nivel de la población (entre el público en general) o a nivel del programa (entre los clientes o los participantes de un programa). El propósito de muchas evaluaciones a nivel del programa es medir si dicho cambio ocurre. Sin embargo, relativamente pocas evaluaciones logran establecer el vínculo causal entre el programa y el cambio (por ejemplo, evaluar el impacto pudiéndolo atribuir a una intervención específica).

**Los indicadores *per se* no son los que definen si la evaluación mide el impacto o si no lo mide.** Más bien, es el diseño del estudio el que determina si se puede establecer el impacto (la causalidad) o no. Para ilustrar este punto utilizamos dos ejemplos. Primero, supongamos que el País X ha tenido un programa para promover la lactancia materna a nivel nacional durante los últimos cinco años. Las encuestas DHS realizadas en el Año 2 y en el Año 5 muestran un incremento evidente en la prevalencia de la lactancia materna exclusiva. Aunque asumimos que el programa contribuyó a lograr dicho incremento, no podemos atribuir el cambio definitivamente a nuestra intervención. Podría ser que otra ONG hubiera promovido la lactancia materna en las mismas áreas sin que lo supiera el Ministerio de Salud Pública (MSP). Estos factores de confusión pueden ser negativos en esencia pero pueden tener un efecto positivo en el resultado (por ejemplo, una depresión económica en el país hace que la leche sea muy cara para las madres). Este ejemplo ilustra una situación en la que el programa logró su objetivo (incrementar la lactancia materna), pero no se pueden descartar otros factores ajenos al programa (“factores de confusión”), que podrían ser los responsables del cambio. El rastreo de los cambios en una población específica (en ausencia de un grupo control), se conoce también como “monitoreo de los resultados” o “monitoreo de las tendencias.” Muchos de los encargados de formular las políticas están completamente satisfechos con este tipo de evaluación, particularmente si los resultados muestran el cambio deseado en la variable del resultado; a ellos no les interesan los otros factores de confusión que pudieran explicar el cambio. De hecho, muchas personas concluirían que “el programa ha tenido impacto” si el cambio deseado ocurre. En cambio, los

especialistas en evaluación reconocen que el simple rastreo o seguimiento del cambio no demuestra la causalidad.

En el segundo caso, supongamos que el gobierno está interesado en evaluar el impacto de su iniciativa para promover la lactancia materna. Para ello, asigna aleatoriamente las 20 provincias del país, dividiéndolas entre un área que recibe la intervención y un área de comparación (es decir, 10 provincias en cada área). Los resultados de un estudio de línea basal realizado en ambas áreas indican que las dos tienen características socioeconómicas y prácticas de lactancia materna similares. El gobierno hace, entonces, una promoción intensiva de la lactancia materna en las provincias experimentales, pero no en el área de comparación (aunque planea implementar las mismas actividades en esta última área al concluir la evaluación, si la intervención demuestra ser efectiva). Tres años después de la implementación, el gobierno conduce un estudio de seguimiento en ambas áreas. Los resultados indican que la lactancia materna exclusiva es mucho más alta en las provincias experimentales que en las provincias de comparación. Los resultados de este estudio cuasi experimental permiten medir qué pasó en ausencia de la promoción y, por lo tanto, atribuir el incremento de la lactancia materna a la intervención. En resumen, este segundo caso llena los requisitos para evaluar el impacto.<sup>9</sup>

Nota: Para estos dos ejemplos, los indicadores podrían haber sido los mismos. Lo que fue distinto—y que permitió valorar el impacto en el segundo caso—fue el diseño del estudio. (Para mayores detalles, véase Bertrand, Magnani, y Rutenberg, 1996, Capítulo IV).

Muchos de los indicadores de este manual se describen en términos de porcentajes porque requieren variables dicótomas con respuestas de “sí o no” (por ejemplo, el uso de anticoncepción, el uso de condones durante la última relación sexual, el uso de lactancia materna exclusiva, la presencia de un proveedor calificado durante el parto). Sin embargo, algunos de los porcentajes del *Compendio* se calculan dicotomizando algunas variables continuas. Por ejemplo, el **Porcentaje de Mujeres Embarazadas que Suben al Menos 1.0 Kg. de Peso por Mes en los Últimos Dos Trimestres del Embarazo** se deriva de la variable continua “aumento de peso entre las mujeres embarazadas durante los dos últimos trimestres del embarazo.” El uso de porcentajes simples (aunque la variable subyacente sea

fundamentalmente continua) es justificable para el propósito de rastrear cambios a lo largo del tiempo. Sin embargo, convertir en dicótomicas las variables continuas significa una pérdida de información; y los evaluadores deben utilizar las variables en su forma continua si desean determinar la causalidad o controlar por los factores de confusión.

## 7. Los Indicadores Cuantitativos frente a los Indicadores Cualitativos

La investigación cuantitativa ha dominado el campo de investigación de la condición de la salud y de las ciencias sociales; sin embargo, la investigación cualitativa ganó bastante aceptación durante la década de los años noventa, a tal punto que se ha vuelto parte integral de muchos proyectos de investigación de las ciencias sociales. La investigación cualitativa también se utiliza con bastante frecuencia en la evaluación de los programas, particularmente para evaluar los procesos (por ejemplo, para medir el nivel de satisfacción de los clientes o las reacciones al programa por parte de los participantes). Las metodologías que más se utilizan en este tipo de investigaciones son los grupos focales, las entrevistas en profundidad, la observación directa y las entrevistas con informantes clave. Algunos investigadores han tratado de cuantificar los resultados de las técnicas cualitativas, pero la tendencia general es tratar de captar las principales ideas de los entrevistados por medio de textos descriptivos en lugar de con porcentajes y otras estadísticas.

Los autores manifiestan un apoyo decidido al uso de los métodos cualitativos en la evaluación de los programas como complemento de las técnicas cuantitativas y de los indicadores cuantificables. La investigación cualitativa es particularmente útil para las siguientes cuatro áreas:

---

<sup>9</sup> Este ejemplo hipotético ilustra los elementos necesarios para demostrar la causalidad, pero no refleja con precisión las realidades prácticas y las dificultades de los experimentos de campo. Por ejemplo, los administradores de los niveles superiores suelen oponerse a dejar de ofrecer las ventajas potenciales a algún grupo de personas dentro de su área administrativa. Otro problema es que las áreas de comparación raras veces son “similares” en todas las variables sociodemográficas relevantes, así como en las conductas o prácticas en estudio.

- Para identificar las necesidades (con el fin de conocer más sobre la situación local antes de diseñar un programa);<sup>10</sup>
- Para comprender la terminología local sobre un tema específico (antes de terminar de preparar los instrumentos para la recopilación de los datos cuantitativos);
- Para evaluar los procesos (documentando la dinámica de cómo funciona un programa, así como sus fortalezas y sus debilidades); y
- Para desarrollar un mejor entendimiento de los resultados que se obtienen por medio de un instrumento cuantitativo (por ejemplo, para las actitudes, las creencias y los valores subyacentes de un hallazgo específico).

A diferencia de esto, la cuantificación es esencial para evaluar los resultados y el impacto.

Resumiendo, la investigación cualitativa juega un papel cada vez más importante en la evaluación de los programas y recomendamos a los evaluadores utilizar los datos, tanto cuantitativos como cualitativos, para comprender mejor la dinámica de un programa. Sin embargo, el uso de la investigación cualitativa está fuera del alcance de este *Compendio*, que centra su atención en los indicadores cuantificables que se utilizan para monitorear las actividades de un programa, para medir los resultados y para demostrar el impacto que este ha tenido.

## 8. El Esquema de Resultados

A mediados de los años noventa, la USAID incorporó un nuevo enfoque para monitorear sus programas a través de toda la agencia, el cual se conoce como el Plan para Monitorear el Desempeño (PMD). El eje central del PMD es el esquema de resultados (“*results framework*”), una herramienta para planificar, comunicar y administrar. El esquema de resultados incluye el objetivo estratégico y todos los resultados intermedios requeridos para lograr los objetivos del programa, ya sea que esté financiado por la USAID o por sus socios. El esquema también comunica las hipótesis de desarrollo implícitas de la estrategia y los vínculos causales entre los resultados intermedios y el objetivo. El esquema incluye las premisas subyacentes de la hipótesis de desarrollo que deben sustentarse para alcanzar el objetivo relevante. Como tradicionalmente suele hacerse, el esquema se presenta en forma de gráfica y se complementa con una parte narrativa

(PriceWaterhouseCoopers, 2001). Los evaluadores pueden aplicar este marco conceptual en los programas nacionales (todas las Misiones de la USAID a nivel mundial trabajan con este esquema de resultados), así como en los programas de proyectos específicos financiados por la USAID (algunos de los cuales se ilustran en la Parte II de este *Compendio*). Las otras agencias donantes tienen herramientas similares para monitorear los programas y, aunque aquí no se describen en detalle, estas herramientas también requieren hacer una selección de los indicadores más apropiados para rastrear el progreso de los programas.

El esquema de resultados tiene varias ventajas. Primero, ha desplazado el sentido del enfoque anterior de un “conteo” simplista de las actividades hacia el logro de los resultados. Ya no interesa el número de afiches producido o el número de enfermeras capacitadas. Lo que ahora se pregunta es: “¿Qué resultados logró el programa?” Una segunda ventaja del esquema de resultados es su amplia difusión a través de toda la USAID, lo que ha logrado incrementar en gran medida el nivel de entendimiento de lo que es el monitoreo y la evaluación entre el personal de los programas, en todos los niveles. Antiguamente, la evaluación solía dejarse a un solitario evaluador que trabajaba aislado en una apartada oficina, pero el esquema de resultados es hoy día parte integral del diseño de un programa. Los gerentes de los programas tienen hoy la responsabilidad de lograr los resultados que establecen en su esquema de resultados y el personal del programa comprende que estos indicadores representan los criterios sobre los cuales se evaluarán sus esfuerzos.

El esquema de resultados también tiene sus limitaciones. Primero, describe las rutas por las cuales las intervenciones programáticas contribuirán al logro de los resultados finales por medio de los vínculos causales (por ejemplo, los resultados en los niveles más bajos que contribuyen a lograr un objetivo estratégico). Sin embargo, el esquema pocas veces identifica, en la práctica, todas las posibles influencias,

<sup>10</sup> La identificación de las necesidades es un componente ampliamente reconocido para diseñar los programas, aunque algunas personas se preguntan si forma parte de la *evaluación* de un programa, ya que ocurre antes que este se establezca. La identificación de las necesidades forma parte de la investigación formativa.

por lo que otros factores ajenos al programa pueden influir en los resultados, en condiciones que no se perciben en el esquema de resultados (el problema de la atribución). Segundo, los resultados deseados no siempre son fáciles de medir en términos cuantificables. Por ejemplo, el Proyecto para la Prevención del VIH/SIDA en Centro América (Proyecto de Acción SIDA de Centro América, PASCA) se diseñó para cambiar el entorno político hacia los programas de VIH/SIDA y para fortalecer la capacidad institucional de las ONG para implementar programas de prevención en la región centroamericana. La necesidad de contar con indicadores para medir el entorno político dió origen al hoy ampliamente utilizado Índice para Medir el Ámbito Político Asociado al SIDA (API por sus siglas en inglés), pero esta medida no había sido puesta a prueba cuando se utilizó por primera vez. Del mismo modo, las medidas para evaluar el fortalecimiento de la capacidad institucional representan “las mejores suposiciones” del grupo de expertos reunido para desarrollarlas, pero no tienen el rigor metodológico de otros indicadores más comunes, como la prevalencia del uso de anticonceptivos o la lactancia materna.

Hay una tercera limitación que no es inherente al esquema mismo, pero sí a su aplicación. Por ejemplo, si los gerentes de los programas temen no lograr los objetivos de comportamiento del programa, en términos de producir un cambio de comportamiento a nivel de la población (por ejemplo, a nivel de resultados), pueden revertir la medida a nivel de los productos, según convenga más a sus propios intereses. En las áreas de la salud reproductiva en las que los productos y los resultados están estrechamente vinculados (por ejemplo, las inmunizaciones), esta falla es menos evidente. Pero, en las áreas en las que el vínculo entre los productos y los resultados es menos evidente (por ejemplo, la nutrición de la mujer), esta limitación adquiere más importancia.

## **9. Niveles de Presentación de Informes para las Agencias Cooperantes de la USAID**

Los indicadores de este *Compendio* se proponen para utilizarse a nivel nacional (por ejemplo, la prevalencia de la lactancia materna, el porcentaje de establecimientos que ofrecen atención postaborto, el porcentaje de mujeres embarazadas con anemia). Sin embargo, los evaluadores deben adaptarlos para que se ajusten a las necesidades específicas de la organización

que los utiliza. Por ejemplo, si un proyecto trabaja en un solo distrito regional o en un conjunto de distritos regionales, esta área geográfica se convierte en la población de interés para dicho proyecto. De manera similar, si un proyecto para adolescentes trabaja únicamente en la ciudad capital, los jóvenes de cierta edad que residen en la ciudad (lo que probablemente se defina por las variables socioeconómicas o geográficas), se convierten en la audiencia objetivo.

Las Agencias Cooperantes de la USAID (es decir, las organizaciones que reciben el apoyo financiero de la USAID y que brindan asistencia técnica a nivel mundial), tienen que satisfacer dos exigencias distintas para los indicadores. Primero, trabajan con las contrapartes del país anfitrión para establecer los mejores indicadores para evaluar el programa o el proyecto a nivel nacional, a nivel del distrito regional o a nivel de la ciudad capital, según se explicó antes. Segundo, probablemente tienen que reportar al donante (por ejemplo, a la USAID) los resultados agregados de su trabajo en los distintos países.

El siguiente ejemplo ilustra cómo pueden los evaluadores adaptar los “mismos” indicadores para evaluar un programa a nivel del país anfitrión y a nivel global. Una organización que ofrece capacitación y que trabaja en colaboración con contrapartes nacionales podría seleccionar los siguientes indicadores de la Parte II.D para medir la capacidad de capacitación que tiene el país:

- Capacidad para mantener un SIG funcional sobre el programa de capacitación;
- Existencia de una estrategia de capacitación para mejorar la calidad de los servicios, con base en las necesidades identificadas; y
- Evaluación sistemática de la organización a su programa de capacitación para mejorar la eficacia.

Esta misma organización puede utilizar los mismos indicadores para reportar a la USAID **el número de países** que lograron estos resultados (del total de países en los que trabajaron).

Este *Compendio* contiene más de 300 indicadores. Muchos de ellos están siendo ampliamente utilizados a nivel mundial para evaluar los programas de salud reproductiva. Otros se encuentran en la fase de experimentación, pues la salud reproductiva se ha ido expandiendo hacia nuevas áreas (por ejemplo, la participación del hombre, la mutilación genital femenina, la violencia contra la mujer). El resto del *Compendio* consiste en un “menú” de indicadores para evaluar la condición de múltiples factores del entorno de la oferta de los servicios (en la Parte II), así como los productos y los resultados de 11 áreas programáticas distintas de la salud reproductiva (en la Parte III).