

DISCUSIÓN SOBRE LA EVALUACIÓN DE IMPACTO DE PROGRAMAS Y PROYECTOS SOCIALES EN SALUD PÚBLICA

ORLANDO SCOPPETTA*
INSTITUTO COLOMBIANO DE BIENESTAR FAMILIAR

Recibido: Abril 5 de 2006

Revisado: Mayo 24 de 2006

Aceptado: Junio 12 de 2006

ABSTRACT

The objective of this article is to present a discussion on major problems related to impact evaluation of social and public health policies, plans and programs. Impact evaluation is a specific but interdisciplinary field, with an important amount of new knowledge in the last decades. Latin America's countries invest important funds in order to evaluate social interventions, following the multilateral agencies directions. Based on the newest experience in Colombia and other Latin American countries, this paper discusses conceptual and practical issues what evaluation teams have to deal.

Keywords: Evaluation, Impact evaluation, Public health.

RESUMEN

El objetivo de este artículo es presentar una discusión acerca de problemas prácticos y teóricos, relacionados con la evaluación de impacto de políticas, planes y programas sociales y de salud pública. La evaluación de impacto es un campo interdisciplinar de conocimiento, pero con características muy particulares y una importante cantidad de adelanto en el conocimiento en las últimas décadas. Los países latinoamericanos invierten grandes montos para evaluar sus intervenciones sociales, siguiendo las directrices de los organismos multilaterales de financiación. Este documento propone una discusión sobre asuntos conceptuales y prácticos que los equipos encargados de hacer evaluaciones de impacto deben abordar, con base en la experiencia más reciente en Colombia y América Latina.

Palabras clave: Evaluación, evaluación de impacto, salud pública.

* Correspondencia: Avenida carrera 68 # 64c-75 Piso 3 (Bogotá, Colombia). Correo electrónico: orlando.scoppetta@icbf.gov.co

Introducción

La evaluación del impacto es un campo interdisciplinario con un alto nivel de costo, especialización y complejidad. La experiencia reciente en América Latina demuestra que se requiere de profundización específica para apropiarse adecuadamente las metodologías desarrolladas hoy en el mundo.

A su vez, el campo de la salud pública abunda en programas y planes de carácter nacional, regional e incluso en pequeños grupos poblacionales. Pocos de estos esfuerzos cuentan con evaluaciones sistemáticas y menos de impacto. Adicionalmente, la misma política de salud pública nacional debería ser objeto de evaluación de su impacto. En este documento se presentará la discusión acerca de temas de gran relevancia para la evaluación de impacto, recogiendo las evaluaciones más importantes que se han realizado en el país recientemente.

Conceptos generales

La OMS define la evaluación de programas de salud como "...un medio sistemático de aprender empíricamente y de utilizar las lecciones aprendidas para el mejoramiento de las actividades en curso y para el fomento de una planificación más satisfactoria mediante una selección rigurosa entre distintas posibilidades de acción futura" (OMS, 1981, p.6). Según Saraceno y Levav (1992, p. 56), "evaluar implica hacer una comparación entre la realidad y lo que se tendría que hacer y lograr según las especificaciones de la misión y los programas desarrollados..."

Por impacto se entiende una dimensión de los resultados de la política, el plan o programa. Para Hernández (1996, p. 1099), esta es "la evaluación en sí de un programa". Por tanto "el juicio acerca de un programa descansa en la contrastación (sic) de que los resultados que se obtienen tras su implantación indican la mejoría en el problema que se pretendía abordar". Con la evaluación de impacto en particular se busca saber si se produjeron "los efectos deseados en las personas, hogares e instituciones y si esos efectos son atribuibles a la intervención..." (Baker, 2000). La noción más extendida de la evaluación de impacto hoy, precisa de la obtención de evidencia *contra fáctica* acerca de qué habría ocurrido si el programa no se hubiera implantado. A su vez, para que esta evidencia *contrafactual* sea válida, se requiere aislar los efectos del medio social para poder identificar con precisión los efectos de las intervenciones (Baker, 2000; Ravallion, 2004).

Esta concepción está bastante difundida y es impulsada por los organismos multilaterales de financiamiento; sin embargo, no está libre de discusiones y no debe entenderse como la única opción, especialmente cuando conduce a procesos muy costosos y de difícil repetición, tratándose de políticas, planes y programas de gran cobertura.

Opciones para la evaluación de impacto

La evaluación de impacto tiene diferentes modalidades y posibilidades no excluyentes unas de otras. Realmente al evaluar una política, plan o programa de salud pública, el escenario es de tal complejidad que requiere la adaptación de estrategias y procedimientos múltiples. La descripción detallada de las diferentes opciones va más allá del alcance de este artículo. Se mencionarán dos de ellas que representan realmente categorías amplias.

Evaluación basada en investigación

Se utiliza la investigación para hacer evaluación por diferentes motivos, uno de ellos es que la investigación le aporta rigor a la evaluación y que los métodos utilizados en la investigación pueden adaptarse para el logro de los fines evaluativos.

Aunque se cuente con un sistema de evaluación con sus diferentes componentes, la evaluación de impacto es un área específica que ofrece un conocimiento particular. Aun así, los tomadores de decisiones y los evaluadores deben estudiar con cuidado la necesidad y aplicabilidad de la evaluación de impacto mediante investigación u otros métodos (Habicht, Victora & Vaughan, 1999; Baker, 2004).

La evaluación basada en la investigación debe afrontar las dificultades teóricas y prácticas de utilizar modelos experimentales o cuasi experimentales, en situaciones que no son necesariamente las más propicias para estos métodos. De hecho, no debe aceptarse como dado el que sea plausible la separación de los efectos de un programa de los efectos de factores que inciden al tiempo en los escenarios sociales. Aunque las evaluaciones que utilizan estos métodos suelen estar rodeadas de toda la parafernalia científica, la supuesta separación de los efectos no pasa de ser una hipótesis de muy difícil contraste.

No es fácil utilizar métodos experimentales en la evaluación de impacto de programas sociales, especialmente por la imposibilidad de tener grupos aleatorios puros. Por otra parte, es frecuente que por problemas en el funcionamiento de los programas tampoco pueda garantizarse el control sobre la dosificación de la intervención, condición necesaria para la aplicación de métodos experimentales o cuasiexperimentales, como se observa en los programas colombianos de Empleo en Acción (DNP, 2004); y Desayunos Infantiles (Forero et al., 2005). Una alternativa a los estudios de este tipo, son aquellos que utilizan muestras relativamente pequeñas y que hacen seguimiento a los sujetos durante el programa y tiempo después de que los sujetos egresan, como en el caso del Programa de Prevención de Montreal, dirigido a disminuir la conducta agresiva y delinencial (Chaux, 2005).

Por lo regular no es posible la aplicación ortodoxa de los métodos experimentales o cuasi experimentales

en la evaluación de las intervenciones, principalmente porque: 1) muchos programas llevan bastante tiempo funcionando y no es posible construir grupos aleatorios; 2) hay muchas relaciones endógenas entre los factores que determinan la asistencia al programa y las variables dependientes; 3) no se previó el sistema de evaluación desde el inicio del programa o nunca se puso en práctica y 4) hay una distancia importante entre el funcionamiento ideal y el funcionamiento esperado.

Una alternativa a la realización de estudios específicos de impacto, es la utilización de datos de fuentes nacionales confiables. Por ejemplo, la Encuesta Nacional de Calidad de Vida (1997-2003), la Encuesta Continua de Hogares (ambas realizadas por el DANE) y la Encuesta Nacional de Demografía y Salud, especialmente la versión del año 2005 (Profamilia, 2005), constituyen fuentes de datos parcialmente explotadas sobre las cuales pueden realizarse estudios específicos, aprovechando que en las diferentes versiones de estas encuestas aparecen datos comparables.

Evaluación basada en indicadores

Una posibilidad, que no excluye la anterior, es la evaluación mediante indicadores. Dentro de las posibilidades de evaluación por indicadores, puede utilizarse la opción de indicadores trazadores. Este tipo de evaluación no se basa en investigaciones sino en sistemas de información que suministran datos.

Un indicador es una expresión matemática o estadística (total, promedio, proporción, tasa o cualquier otro) que sirve para dar cuenta del comportamiento de un evento. Un indicador trazador, es un indicador seleccionado por considerarse que es el que mejor expresa el comportamiento de un evento. Seleccionar indicadores trazadores es favorable en tanto hace más eficiente el proceso mismo de evaluación, puesto que ésta se concentra en menos variables.

Matemáticamente, la mejor manera de escoger indicadores trazadores sería el cálculo de factores, componentes principales o índices ya que estos sintetizarían el comportamiento de varios indicadores.

Una condición de la evaluación de impacto por indicadores es que el programa defina con claridad en cuáles indicadores se esperaría ver sus efectos, como por ejemplo, las disminuciones en los casos de malformación del tubo neural en Chile, mediante la fortificación de la harina con ácido fólico y otras medidas de salud pública (Nazer et al., 2001).

Los indicadores para evaluación de impacto en salud, suelen ser de tipo epidemiológico en la medida en que dan cuenta del seguimiento a enfermedades y factores protectores o de riesgo. Sin embargo, el campo de la promoción de la salud y del trabajo con la comunidad demanda la creación de indicadores diferentes, como los

«indicadores ambientales», los cuales cumplen dos funciones en la evaluación: sirven para medir las actitudes compartidas y los comportamientos colectivos; además, captan las características medioambientales que podrían integrarse a un modelo con los cambios en salud y las acciones de las políticas, planes o programas (Koepsell et al., 1996, p. 225).

Otro aspecto de la evaluación por indicadores es que se puede tener una mejor visión de los cambios sociales en su conjunto. La inversión social no debería estar dirigida a corregir tal o cual problema específico. Particularmente en los países en vías de desarrollo, la inversión social debería dirigirse a fomentar el desarrollo social integral, aumentando la equidad (Misión Social, DNP, PNUD, 2003). Por lo tanto, los indicadores sociales son un termómetro necesario para evaluar los resultados de las políticas.

Adicionalmente, conviene estudiar la utilización de indicadores de desarrollo que dan cuenta de los procesos sociales. La mejoría de las interacciones puede y debe ser uno de los focos de las políticas sociales. Una nueva ola de pensamiento propone incluir dentro del conjunto de indicadores tradicionales, indicadores de desarrollo que den cuenta de las interacciones (Dawes, Bray, Kvasvíg, Rama & Richter, 2004; SEP, SSA, DIF, UNAM, UNICEF & UNESCO, 2001).

También es frecuente que los datos requieran ajustes antes de presentarlos como representación de la realidad social. Por ejemplo, los datos de mortalidad requieren frecuentes ajustes por subregistro (Piñeros & Murillo, 2004), o en el caso de requerirse comparaciones entre áreas geográficas, es necesario aplicar técnicas para evitar sesgos por las diferencias en las estructuras de población (Rothman, 2002; Londoño, 2004). La comparación directa de tasas crudas puede ser un error frecuente entre psicólogos y otros analistas de lo social.

Podría discutirse que al evaluar por indicadores no se aíslan los efectos puros de la intervención de los demás factores del ambiente que afectan los resultados. Sin embargo, si hay suficiente evidencia acerca de los efectos de las intervenciones, si hay control acerca de la evolución de las intervenciones, habrá un sustento adecuado para atribuir los cambios a las intervenciones.

Principales problemas teóricos y prácticos de la evaluación de impacto

En cualquier evaluación de impacto se deben superar problemas fundamentales, que de no considerarse pueden dar lugar a esfuerzos baldíos en los cuales se inviertan una gran cantidad de recursos sin obtener información válida.

Uno de los aspectos más importantes es la definición del modelo teórico de la política, plan o programa. Si éste

parte del reconocimiento de un problema, debe existir un enlace teórico entre el problema y las acciones propuestas para su solución (Koepsell et al. 1996).

El modelo teórico de la intervención

Toda intervención supone o reconoce un problema y plantea un conjunto de intervenciones de las cuales se cree que tienen algún efecto sobre el problema. Debe existir algún sustento racional para apoyar el postulado de que las intervenciones propuestas afectan el problema en cuestión. Este sustento racional se encuentra en planteamientos teóricos o por lo menos hipotéticos (Becoña, 1999).

Es común que en las políticas, planes o programas no se especifique el tipo de relación entre el problema y las intervenciones. En este caso se acude al planteamiento simplista de que las intervenciones solucionan el problema de alguna manera.

Por ejemplo, es frecuente que en los programas de prevención del consumo de drogas, se incluya el refuerzo de la autoestima de los sujetos de la intervención y que no se demuestre en forma alguna, cómo la autoestima se relaciona con el consumo de drogas.

Es común que no se cuente con modelos explicativos ni investigaciones suficientes para adjudicar el cambio favorable o el desfavorable al programa, plan o política implantado. La importancia de los modelos radica en que al lograr la simplificación de los sistemas de la realidad, es posible aprehender sus características más importantes. Cualquier sistema de la naturaleza es de tal complejidad, que resulta imposible un acercamiento a éste *tal cual es*, por lo anterior tal reducción es necesaria.

Los rasgos del sistema se expresan en variables, lo que permite la formalización del comportamiento en términos lógicos o matemáticos, que es la finalidad última de la modelación. Cuando se logra esta formalización, es posible explicar el evento, predecir su comportamiento y explorar de manera sistemática diferentes posibilidades de control y prevención.

En el caso de las políticas, planes o programas en salud, los modelos deben establecer la relación entre el problema y la intervención en términos de variables que por lo regular son indicadores epidemiológicos o del conocimiento, actitudes y prácticas de los sujetos de la intervención. También deben incluir variables del ambiente, que incluyen cambios poblacionales, políticos y otros, siempre y cuando estos se consideren cruciales para el modelo.

Un modelo en salud requiere múltiples niveles de medición, correspondientes a múltiples niveles de conceptualización (Koepsell et al. 1996). Esto se debe a que entre un factor que supuestamente incide sobre el evento y el evento, por lo regular hay otros 'momentos'

y variables del sistema que determinan el efecto de la intervención. Por ejemplo, el modelo de la acción razonada de Fishbein y Ajzen (1980), propone que para lograr un cambio en la conducta es necesario un cambio en la actitud hacia la conducta que se quiere cambiar y un cambio en la norma subjetiva. Sin embargo, el efecto del cambio en las dos variables independientes, que debe observarse en la conducta, está mediado por el cambio en la intención del sujeto. En otras palabras, el modelo requiere varios niveles de medición.

No basta con proponer un modelo teórico y luego evaluar los cambios en las variables dependientes. Es necesario estudiar el comportamiento del modelo en sí mismo para establecer si se comporta como está predicho, en términos de los diferentes factores que lo componen. Esto implica tomar medidas de los cambios en las variables que explican un determinado comportamiento o resultado, según los términos del modelo, así como establecer si tales variables covarían según está dispuesto en el modelo. Por ejemplo, Muñoz, Gutiérrez y Guerrero (2004), informan que su modelo de prevención de la violencia se basa en el modelo de cambio de comportamiento propuesto por Bandura. Sin embargo, en la evaluación de impacto en tres ciudades de Colombia no se hacen las estimaciones correspondientes al funcionamiento en sí del modelo por lo que, aunque, según su propio reporte, los resultados de la intervención son positivos, no puede afirmarse que el modelo funcionó a partir de los datos publicados.

En otras circunstancias es posible establecer modelos donde la política, plan o programa figura como fuerza que explica el comportamiento del evento. El modelo, además de establecer la forma de las relaciones entre el problema y la intervención, permite calcular la cantidad de la variación del evento explicada por la política, plan o programa.

La evidencia acerca de la efectividad de la intervención

En el transcurso de este artículo se ha utilizado la palabra intervención para designar cualquier tipo de política, plan o programa en salud pública, o en general en el ámbito de lo social. Sin embargo, conviene introducir cierta diferencia importante entre éstos y las intervenciones puras. Por *intervención pura* se entiende el mecanismo concreto, del cual, al ser introducido, se espera que produzca un efecto.

Por ejemplo, si se suministran alimentos en la calidad y la cantidad suficiente, se esperará un mejoramiento de la condición nutricional de los beneficiarios. A su vez, si se hace llegar un subsidio en dinero a una familia, se esperará una mejora en el ingreso y en el gasto, con los consecuentes efectos colaterales positivos.

Un programa se basa típicamente en una intervención pura que se supone ha sido probada previamente. En términos de salud pública o de cualquier intervención social, debe haber evidencia previa acerca de la efectividad de las intervenciones antes de incurrir en los altos costos que implica su utilización masiva. Otra fuente de evidencia es la evaluación de la misma intervención en otros contextos. Por ejemplo, el programa colombiano Familias en Acción se parece mucho al programa Progresía de México, el cual cuenta con un importante acervo evaluativo.

Para realizar una evaluación de impacto, no basta asociar la implantación del programa con los cambios en el indicador. Se deben construir modelos que con el suficiente respaldo teórico, integren variables y factores, siendo el grado de implantación del programa uno de esos factores. Un ejemplo de lo anterior lo constituye la evaluación del programa de EDA-IRA. La incidencia de enfermedad diarreica y de infección respiratoria, dependen de la presencia de un conjunto de factores ambientales, de salud pública y de saneamiento básico de las comunidades. Las variaciones en EDA e IRA, pueden ser explicadas por tales factores y no por el programa (Rubio & Sánchez, 1999).

El tipo de relación entre el programa y el evento, puede anticiparse parcialmente cuando se ha demostrado previamente que las acciones que lo componen son efectivas. Esto desafortunadamente no siempre es así, puesto que es común que las acciones de un plan, política o programa se conformen sin que se haya demostrado su eficacia o efectividad.

Puede considerarse que la capacidad que tienen las acciones de la política, plan o programa de lograr resultados, corresponde a la eficacia de la política, plan o programa. En otras palabras, al iniciar, toda política, plan o programa, tiene una eficacia potencial que se hará efectiva en la medida en que las condiciones administrativas sean adecuadas; así, la eficacia se expresa en efectividad, que tiene que ver con el funcionamiento de las acciones eficaces en unas condiciones determinadas (eficacia real).

La evidencia previa acerca de la efectividad de la intervención no garantiza que su aplicación como política, plan o programa sea exitosa. Hay una distancia importante entre la efectividad de las intervenciones puras y la efectividad de las políticas o los programas. Al aplicar intervenciones eficaces a conjuntos amplios de población los efectos pueden estar mediados por problemas para garantizar la exposición adecuada.

En el caso del PAI (Programa Ampliado de Inmunizaciones), por ejemplo, las adecuadas coberturas de vacunación garantizan el cambio en los indicadores epidemiológicos que dan cuenta de los eventos inmunoprevenibles. Pero en Colombia, los problemas de la política de salud han generado reducciones en las

coberturas de vacunación, como lo corroboran diferentes fuentes (Profamilia, 2005; Carrasquilla, Porras e Ibáñez, 2005). Así, una eventual epidemia por enfermedades que deberían estar controladas, sería achacable a la política o a la aplicación del PAI en el país, no a una falla de la vacunación como mecanismo de salud pública.

El grado de estructuración y funcionamiento de los programas y su relación con la evaluación de impacto

Uno de los grandes problemas al desarrollar una evaluación de impacto, es saber si el programa a evaluar está implantado y cuál es su grado de desarrollo.

Evaluar un programa que no ha sido implantado, o por lo menos no adecuadamente, fue denominado *error del tipo III* en un artículo de Basch, Slipceovich, Gold, Duncan y Kolbe (1985). En el campo de la salud pública es altamente probable que los programas, planes o políticas existan sobre el papel y en los discursos, pero no estén suficientemente implantados en la práctica. También es posible que un conjunto de actividades con un mediano grado de integración sean denominadas programas o planes.

Se espera que una intervención implantada, cumpla por lo menos con las siguientes características:

Una descripción del problema que pretende abordar que incluya: 1) un recuento histórico sobre el comportamiento del evento; 2) una descripción de su naturaleza; 3) un recuento de las principales acciones desarrolladas en el país y en el mundo alrededor del mismo; 4) un planteamiento de objetivos del programa; 5) el planteamientos de las metas; 6) una descripción de las acciones y de los diferentes responsables por niveles; 7) la especificación de talento humano y recursos en los diferentes niveles del programa; 8) la especificación de los indicadores de evaluación del programa en los diferentes niveles; 9) la definición de las acciones, procesos y procedimientos de la política, plan o programa.

Adicionalmente, en cuanto a las condiciones para su funcionamiento, debe ser discernible una estructura que sustente las acciones del programa en cuanto a 1) talento humano en los diferentes niveles del sistema, con las condiciones de capacitación especificados; 2) la definición de las funciones del talento humano comprometido.

Por otra parte, no es aconsejable iniciar las evaluaciones de impacto cuando el programa está recientemente implantado, dado que es frecuente que el programa sufra adaptaciones antes de llegar a una especie de etapa de madurez.

Las evaluaciones de impacto realizadas recientemente en Colombia, muestran que el mal funcionamiento de los programas crea problemas

operativos y conceptuales para las evaluaciones de impacto, ya que es difícil asegurar la exposición mínima requerida. Por ejemplo, en las evaluaciones a los programas de subsidio como Familias en Acción y Empleo en Acción (DNP, 2004) y en otros en los cuales se dispensa algún bien social, como el programa de Desayunos Infantiles (Forero et al., 2005), se observa que los problemas en la frecuencia en los suministros interrumpen la dosificación necesaria para producir los efectos esperados.

El tiempo de la exposición real a una intervención es un aspecto crítico de las evaluaciones de impacto (Behrman, Cheng & Todd, 2004). De todos modos, considerar el tiempo de exposición no es suficiente puesto que la intensidad de la exposición (la verdadera exposición) es la que determina los efectos y medirla es muy complicado. Por ejemplo, en la evaluación de impacto de los Hogares Comunitarios de Bienestar que se realizó en el marco de la evaluación del Programa Familias en Acción, se tuvo en cuenta el tiempo en meses de asistencia al servicio (Attanasio & Vera-Hernandez, 2004). Esto puede constituir una buena aproximación pero no ser suficiente cuando se trata de servicios *continuos*, como aquellos que se basan en la permanencia en un determinado ambiente (a diferencia de los discretos, como la provisión de subsidios), puesto que la asistencia no garantiza que se hayan suplido los servicios en cantidad y calidad suficiente.

Por otra parte, En el caso del Programa Familias en Acción (Attanasio y Gómez, 2004), se incluyeron variables para controlar la discontinuidad en la entrega y recepción de los subsidios. No obstante, es aconsejable evaluar la eficiencia operacional del programa para introducir adecuaciones a la evaluación de impacto, como se hizo en la evaluación del programa *Progresá* en México (Skoufias & Behrman, 2000).

La relación temporal entre la implantación de la política, plan o programa y los efectos esperados

El tiempo es una variable que complica de manera extraordinaria la evaluación de impacto de las políticas, planes o programas, en diferentes sentidos. Por una parte, como ya se mencionó, es importante conocer los tiempos de exposición efectiva al programa, lo que requiere de un complejo sistema de registro de los individuos incluidos dentro de la evaluación.

Adicionalmente, en el campo de la salud pública es de esperarse que los indicadores epidemiológicos sean afectados por las políticas, planes o programas bastante tiempo después de la implantación de estos; por lo tanto, una evaluación cuando el programa está en funcionamiento o terminado recientemente, puede mostrar impactos que no se sostendrán; o no mostrar los resultados esperados. Grossman (1994) sugiere dejar

pasar entre 12 y 18 meses para poder evaluar adecuadamente los efectos de las intervenciones.

En la actualidad se hace mucho énfasis en la comparación entre un grupo intervenido y un grupo de control. La diferencia entre los dos grupos en las variables en las cuales se esperan los resultados de la intervención, se considera el impacto de la misma (Baker, 2000; Ravallion, 2004). Esta postura subestima la importancia del tiempo posterior a la intervención. En otras palabras, no responde a la pregunta fundamental sobre lo que sucede una vez los sujetos egresan de un programa, la cual es de gran importancia si lo que se espera es que las intervenciones sociales contribuyan al desarrollo, la equidad y a la independencia económica por parte de los beneficiarios (Misión Social, DNP & PNUD, 2003).

Surgen entonces otras complicaciones asociadas al tiempo. En la medida en que los cambios se dan más mediatizados, es necesario establecer diseños más complejos y es más difícil controlar la influencia de variables externas a la política, plan o programa. También se hace más difícil coleccionar la información necesaria para los análisis. Por lo tanto, la evaluación debe considerar los efectos del tiempo y en la medida de lo posible basarse en estudios longitudinales. Las buenas intervenciones sociales son aquellas que pueden demostrar efectos duraderos en el tiempo, no aquellas en las cuales los sujetos, una vez egresados, pierden lo que ganaron por su permanencia en el programa.

El aislamiento de los efectos de la intervención y la obtención de evidencia contrafactual

Pretender evaluar los efectos de las intervenciones es común en medios donde se privilegia la intervención social basada en programas y no en políticas que aborden el desarrollo integral. La separación de efectos es plausible en el contexto del pensamiento analítico; desde otra perspectiva podría entenderse que una vez que un factor es introducido en un ambiente social, produce cambios que no son del todo discernibles y actúa en conjunto con un todo. De la política social debería esperarse una acción benéfica en la calidad de vida de las personas y no una mejoría parcial en un único conjunto de factores. Adicionalmente, las variables del entorno socioeconómico pueden competir con la intervención contrarrestando sus efectos (Instituto Colombiano de Bienestar Familiar, 1996).

Para el propósito de reconocer los efectos de la intervención de entre el amasijo de factores y resultados que se presenta en la cotidianidad de los sujetos, se han desarrollado un conjunto de técnicas analíticas. Su aplicación, sin embargo, es compleja, frecuentemente requiere adaptaciones sobre la marcha y sus resultados no están libres de discusión.

Los diseños ideales para la evaluación de impacto, según las tendencias actuales, son los experimentales (obviando la discusión que se introdujo antes). Sin embargo, en la implantación de intervenciones sociales es muy difícil lograr que un grupo sea asignado de manera aleatoria a la intervención y otro no sea intervenido, salvo contadas excepciones como en el caso del programa Progresía en México (Skoufias & Behrman, 2000) y Familias en Acción en Colombia; (Attanasio & Gómez, 2004). Esto no implica la negación de servicios a quienes lo necesitan, sino una compaginación de la inclusión gradual en la intervención con la programación de la evaluación de impacto. Otra opción es la focalización precisa, dando prioridad de entre los elegibles para el beneficio a aquellos más necesitados (Baker, 2000).

No obstante, los diseños experimentales propiamente dichos, no garantizan por sí mismos que los errores y sesgos sean eliminados (Habicht, Victora & Vaughan, 1999). Se recomienda, en todo caso, la utilización de varias series de medidas (Baker, 2000). Así, los estudios cuasi experimentales se erigen como la mejor opción disponible en la mayoría de las oportunidades, aunque es necesario trabajar con modelos estadísticos y econométricos complejos para suplir diferentes sesgos. Por ejemplo, la participación de las personas en el programa puede estar determinada por una condición socioeconómica particularmente desfavorable, lo que haría parecer que el grupo intervenido nunca mejora en comparación con el grupo de comparación. Aun cuando se suponga que una determinada intervención se aplica a sujetos de condiciones económicas muy similares, hay diferencias observables y no observables que en su análisis revelan múltiples segmentos de población (ver por ejemplo, Scoppetta, Giraldo, Borda, Gutiérrez & Cepeda, 2003).

Para contrarrestar las deficiencias propias de los diseños de evaluación, se han desarrollado métodos como los de pareo (siendo muy recomendado en la actualidad el que utiliza la *correspondencia de puntuación de la propensión*); los métodos de diferencia en diferencia y los métodos de variables instrumentales (Baker, 2000; Behrman, Cheng & Todd, 2004).

Los métodos de variables instrumentales, por ejemplo, han sido utilizados con éxito en Colombia para demostrar efectos de los Hogares Comunitarios de Bienestar que con otros métodos estadísticos y econométricos convencionales, se subestiman. Variables que dan cuenta de la distancia entre el hogar del beneficiario y el punto donde se provee el servicio, sirvieron como instrumento en la medida en que intervienen en la decisión de participar mas no en los efectos de los programas (Attanasio & Gómez, 2004; Attanasio & Vera-Hernandez, 2004). A su vez, los métodos de pareo fueron los más ajustados según Behrman, Cheng y Todd (2004) para un programa similar en Bolivia (PIDI).

La descripción de cada uno de estos métodos supera las condiciones de espacio de este artículo. Una mejor aproximación se logra a partir de la revisión de las referencias bibliográficas aquí incluidas.

El nivel de cambio aceptable

Finalmente, uno de los aspectos a discutir es el nivel de cambio producido por una intervención que se debe considerar aceptable. Lamentablemente no hay ningún parámetro establecido para ello. El desarrollo en evaluación de impacto se ha centrado en las metodologías para observar sin sesgos los efectos, sin preguntarse hasta qué punto tales efectos son o no aceptables.

Una manera de resolver esta cuestión tiene que ver con la valoración de los efectos económicos de las intervenciones, en términos del ahorro que con ellas se permite. Por ejemplo, el Programa de Prevención de Montreal mostró ahorros de cinco millones de dólares por su implantación (Chaux, 2005). Sin embargo, la evaluación de este programa mostró que 21% de los participantes llegaron a tener antecedentes criminales (la cifra en el grupo de control fue del 30%) ¿Es este un resultado aceptable? A simple vista parece bastante alto el porcentaje de intervenidos con registros criminales (17% siguieron trayectorias de *agresión física alta o medianamente alta*), aunque haya evidencias de la efectividad del programa.

Todas las intervenciones en el campo social obtienen éxitos y fracasos relativos. La clave podría estar en un mejoramiento constante. No repetir los errores conocidos, avanzando sobre el conocimiento que se obtiene a partir de las evaluaciones. Las nuevas intervenciones deberían producir, gradualmente, mejores resultados.

Conclusiones

La discusión sobre la evaluación de impacto gira alrededor de asuntos conceptuales de fondo, de cuya resolución dependerá una respuesta metodológica. Sin embargo, la práctica ha demostrado que es en el escenario mismo de las evaluaciones donde se adaptan las metodologías y se toman las decisiones finales, en condiciones de alta complejidad social.

Hay una gama interesante de posibilidades para la valoración de evaluación de impacto de intervenciones en lo social que recogen avances muy particulares en el conocimiento metodológico, estadístico y econométrico. No obstante, cada evaluación de impacto es un reto particular.

Tal vez la pregunta más importante en cuanto a la evaluación de impacto, en cada caso, es si vale la pena hacerla. *Para qué* hacer evaluación de impacto y *por qué*, son preguntas que aunque implican respuestas que parecen obvias, deben ser abordadas con el máximo de

seriedad y compromiso en cada caso, en la medida en que tal vez no sea necesario incurrir en tan altos costos.

Lo que muestra la experiencia reciente en América Latina, salvo contadas excepciones, es que cada proceso de evaluación demanda probar diferentes opciones y que más que recetas, se requieren adaptaciones y posiciones flexibles. Los programas en América Latina suelen responder más a los intereses de los políticos (no de la política), lo que repercute en formulaciones pobres, flaco fundamento en la evidencia y evaluaciones que tienen que batirse en medio de múltiples dificultades.

Las evaluaciones de impacto basadas en investigaciones son muy costosas y frecuentemente las decisiones no se basan en ellas. Tal vez sería preferible contar con sistemas robustos de información, que permitan ajustar las intervenciones con el paso del tiempo.

La política social debería evaluarse en su conjunto, con base en los indicadores de pobreza y equidad, acompañados de los indicadores de salud, educación y participación. El desarrollo social no se logrará con programas que aumentan la dependencia de los hogares de los recursos del Estado. Esto constituye un retroceso a nuevas versiones de populismo que no contribuyen al desarrollo social.

El impacto de las intervenciones podría estimarse con metodologías de bajo costo y menos complejidad. En países como Colombia, la ausencia de una política social clara se oculta bajo la multiplicidad de programas que no resuelven los problemas de fondo. En este escenario, la evaluación de impacto de intervenciones puntuales poco aporta, cuando no se tiene información sobre qué esperar de tales acciones, y cuando estas serán modificadas por los grupos de turno al poder, sin considerar seriamente el resultado de las evaluaciones.

Referencias

- Ajzen I. & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, New Jersey: Prentice Hall.
- Attanasio, O. & Gómez, L. (2004). *Evaluación de impacto del programa Familias en Acción- Subsidios condicionados de la Red de Apoyo Social. Informe del primer seguimiento*. Bogotá DC: Departamento Nacional de Planeación.
- Attanasio O. & Vera-Hernandez, M., (2004). *Medium and Long Run Effects of Nutrition and Child Care. Evaluation of a Community Nursery Programme in Rural Colombia*. The Institute for Fiscal Studies. Mimeo.
- Baker, L. (2000). *Evaluación de impacto de los proyectos de desarrollo en la pobreza. Manual para profesionales*. Washington: Banco Mundial.
- Basch, C., Slipceovich, E., Gold, R., Duncan, D. & Kolbe, L., (1985). Avoiding Type III Errors in Health Education Program Evaluations: A Case Study. *Health Education Quarterly*, Vol 12(4): 315-331.
- Becoña, E. (1999). *Bases teóricas que sustentan los programas de prevención de drogas*. Madrid: Universidad de Santiago de Compostela, Plan Nacional sobre Drogas.
- Behrman, J. R., Cheng, Y. & Todd, P. E. (2004). Evaluating Preescolar Programs when Length of Exposure to the Programme Varies: A Nonparametric Approach. *The Review of Economics and Statistics*, 86. 108-132.
- Brown, F. (1980). *Principios de medición en psicología*. México. Manual Moderno.
- Carrasquilla, G., Porras, A. & Ibáñez, M. (2005). *Evaluación de los factores asociados al uso de servicios de salud en menores de edad en Colombia*. Bogotá: Instituto Colombiano de Bienestar Familiar.
- Chaux, E. (2005). El programa de prevención de Montreal: Lecciones para Colombia. *Revista de estudios sociales*, 21, 11-25.
- Chew, F., Torún B. & Viteri, F. (1998). Suplementación diaria y semanal para mejorar el estado de hierro y prevenir la anemia en niñas escolares. *IV Bienal de Pediatría, Caja Colombiana de Subsidio Familiar, COLSUBSIDIO*.
- Dawes, A., Bray, R., Kvalsvig, J., Rama, Sh. & Richter, L., (2004). *Chile & adolescent well-being indicators*. Cape Town: Human Science Research Council.
- DNP (2004). Programa empleo en acción. Condiciones iniciales de beneficiarios e impactos de corto plazo. *Evaluación de políticas públicas*, 2. Bogotá: Departamento Nacional de Planeación.
- Forero, A., Giraldo, M., Scopetta, O., Heredia, A., Borda, C., Fajardo, L. & Sánchez, H. (2005). *Evaluación de impacto del programa de desayunos infantiles del Instituto Colombiano de Bienestar Familiar. Primera medición*. Bogotá: Instituto Colombiano de Bienestar Familiar.
- Gofin, J., Levav, I. & Saraceno B. (1992) Vigilancia y evaluación en los programas de salud mental comunitario. En Levav, I. (1992). *Temas de salud mental en la comunidad*. Serie PALTEX para ejecutores de programas de salud, 19. Organización Panamericana de la Salud.
- Grossman, J. (1994). *Evaluating Social Policies: Principles and US Experience*. The World.
- Habicht, J. P., Victoria C. G. & Vaughan J. P. (1999). Evaluations designs for adequacy, plausibility and probability of public health programme

- performance and impact. *Internacional Journal of Epidemiology*, 28, 10-18.
- Hernández, J. (1996). Evaluación de programas de salud. En Buela-Casal, Caballo & Sierra (1996). *Manual de evaluación en psicología clínica y de la salud*. Madrid: Siglo XXI.
- Instituto Colombiano de Bienestar Familiar (2000). *Evaluación de impacto de hogares comunitarios de bienestar*. Bogotá DC: Autor.
- Koepsell, T. D., Wagner, A. C., Cheadle, D. L., Patrick, D. C., Martin, P. H., Diehr, E. B., Perrin, A. R., Kristal, C. H., Allan-Andrilla & Dey, L. J. (1996). Algunos aspectos metodológicos de la evaluación de los programas de promoción de la salud y prevención de las enfermedades basados en la comunidad. En *Promoción de la salud: Una antología*, 557. Washington DC: Washington. Organización Panamericana de la Salud – Organización Mundial de la Salud.
- Labonte, R. (1996). Estrategias para la promoción de la salud en la comunidad. En *Promoción de la salud: Una antología*, 557. Washington DC: Washington. Organización Panamericana de la Salud - Organización Mundial de la Salud.
- Luengo, M., Romero, E., Gómez, J., Guerra, A. & Lence, M. (1999). *La prevención del consumo de drogas y la conducta antisocial en la escuela: análisis y evaluación de un programa*. Madrid: Ministerio de Educación y Cultura, Ministerio de Sanidad y Consumo, Ministerio del Interior, Universidad Santiago de Compostela.
- Londoño J. L. (2004). *Metodología de la investigación epidemiológica*. (3 ed.) Bogotá DC: Manual Moderno.
- Misión Social, DNP & PNUD (2003). *¿Quién se beneficia del Sisbén?* Evaluación integral. Bogotá: Misión Social, Departamento Nacional de Planeación, Programa de Naciones Unidas para el Desarrollo.
- Muñoz, E., Gutiérrez, M. & Guerrero R., (2004) *Evaluación del impacto de una estrategia de información, educación y comunicación para la prevención de la violencia en Bogotá, Cali y Medellín, 2003-2004*. Santiago de Cali: Ministerio de la Protección Social, Universidad del Valle, Cisalva.
- Nazer, H., Julio, Cifuentes, O., Lucía, Rodríguez C., Mariela et al. (2001). Malformaciones del sistema nervioso central en el Hospital Clínico de la Universidad de Chile y maternidades chilenas participantes en el Estudio Colaborativo Latinoamericano de Malformaciones Congénitas (ECLAMC). *Revista médica de Chile*, vol. 129, (10), 1163-1170.
- Nutbeam, D., Smith, C. & Catford, J. (1996). La evaluación en la educación para la salud: una revisión de sus progresos, posibilidades y problemas. En *Promoción de la salud: Una antología*, 557. Washington DC: Organización Panamericana de la Salud – Organización Mundial de la Salud.
- OPS-OMS (1991). Estratificación epidemiológica de la malaria en la Región de las Américas. *Boletín epidemiológico, Organización Panamericana de la Salud, Vol. 12 No. 4*, 1-4.
- Organización Mundial de la Salud (1981). *Evaluación de programas y servicios sociales: normas fundamentales para su aplicación en el proceso de gestión para el desarrollo nacional de la salud*. Ginebra: Autor.
- Piñeros M. & Murillo R., (2004) Incidencia de cáncer en Colombia: importancia de las fuentes de información en la obtención de cifras estimativas. *Revista Colombiana de Cancerología, Vol 8. No 1*, 5-14.
- Profamilia (2005) *Salud sexual y reproductiva. Encuesta Nacional de Demografía y Salud 2005*. Bogotá: Autor.
- Ravallion M. (2004). *Principles and methods of ex-post impact evaluation WB.I Workshop on Impact Evaluation*, Mexico City, April 2004.
- Rothman, K. (2002). *Epidemiology. An introduction*. New York: Oxford University Press.
- Rubio, M. & Sánchez A. (1999). *Factores de riesgo asociados a la mortalidad por neumonía en menores de un año en Santa Fe de Bogotá*. Bogotá: Colección de Investigaciones. Secretaría Distrital de Salud.
- Saraceno, B. & Levav, I. (1992) La evaluación de servicios de salud mental en la comunidad. En I. Levav (1992). *Temas de salud mental en la comunidad*. Serie PALTEX para ejecutores de programas de salud, número 19. Organización Panamericana de la Salud.
- Scoppetta, O., Giraldo, M., Borda, C., Gutiérrez, A. & Cepeda, A. (2003). *Relaciones entre el refuerzo alimentario escolar y la alimentación en el hogar*. Bogotá: Instituto Colombiano de Bienestar Familiar.
- Skoufias, E. & Behrman, J. (2000). *¿Está dando resultados PROGRESA? Informe de los resultados obtenidos de una evaluación realizada por el IFPRI*. México: Instituto Internacional de Investigación en Políticas Alimentarias.
- SEP, SSA, DIF, UNAM, UNICEF & UNESCO (2001). *Memoria: Foro sobre indicadores de bienestar en la primera infancia en México*. Noviembre 22-23 de 2001. México: UNICEF.

