



GOBIERNO DE CHILE
MINISTERIO DE HACIENDA
Dirección de Presupuestos

ANEXO

METODOLOGÍA EVALUACIÓN DE IMPACTO

División de Control de Gestión
Santiago, Mayo 2009
CHILE

PRESENTACIÓN*

El anexo que a continuación se presenta tiene por objeto describir aspectos metodológicos necesarios de considerar en el desarrollo de “Evaluaciones de Impacto” de programas públicos. El documento se centra en los ámbitos de resultados de eficacia, en sus diferentes niveles, y en los aspectos de eficiencia y economía.

Se trata de un documento de carácter referencial. Su elaboración ha tenido como base diferentes documentos preparados con anterioridad en la División de Control de Gestión, los que se han sido complementados con elementos conceptuales y metodológicos extraídos de la literatura especializada.

Heidi Berner H.

Jefa de la División de Control de Gestión
Dirección de Presupuestos

* Este trabajo ha sido elaborado y actualizado por la División de Control de Gestión de la Dirección de Presupuestos del Ministerio de Hacienda. Participaron en su elaboración Heidi Berner, Paula Darville, Nelson Guzmán, Ana María Montoya y Sebastián Izquierdo.

METODOLOGÍA EVALUACIÓN DE IMPACTO ASPECTOS RELEVANTES

1. ¿Qué es la Evaluación de Impacto?

Dependiendo de los ámbitos a evaluar la línea de Evaluación de Impacto (EI) en aplicación en la Dirección de Presupuesto comprende dos modalidades; las Evaluaciones Módulo de Impacto y las Evaluaciones en Profundidad. Las Evaluaciones Módulo de Impacto integran la evaluación de los resultados de corto, mediano y largo plazo de los programas (eficacia), con los principales aspectos de la eficiencia y economía en el uso de los recursos. Las Evaluaciones en Profundidad, además de las dimensiones revisadas por la modalidad de Módulo de Impacto, realizan un análisis de los aspectos relativos a la gestión de los procesos internos de los programas.

La evaluación de los resultados de un programa público requiere desarrollar un proceso de trabajo sistemático e independiente orientado a responder preguntas tales como: ¿Se ha contribuido significativamente a resolver el problema que dio origen al programa? ¿Los receptores de los bienes y servicios que genera el programa han resultado realmente beneficiados? ¿Los beneficios recibidos por los participantes son los que se había propuesto lograr el programa o son otros no previstos? ¿Los beneficios están llegando a la población que se pretende atender? ¿Los usuarios del programa están satisfechos con los bienes y servicios recibidos? ¿Los beneficiarios han experimentado un mejoramiento significativo en su condición-problema inicial, como consecuencia de su participación en el programa?

De las preguntas anteriores se desprende que la evaluación de impacto comprende la “efectividad” de las medidas tomadas o los servicios entregados por el programa. Por otra parte, sí se considera que los recursos necesarios para la ejecución de un programa son escasos y de uso alternativo, los mismos beneficios no pueden ser logrados a cualquier costo, entonces la “eficiencia” también es una dimensión a abordar en una evaluación de impacto.

Hacer una evaluación de eficiencia implica responder entre otras las siguientes preguntas: ¿Los recursos que se han invertido han sido socialmente rentables?, es decir, ¿el conjunto de beneficios generados por el programa, valorados socialmente, es mayor que los gastos financieros y no financieros de éste? ¿La unidad ejecutora del programa conoce y está considerando todos los gastos relevantes? ¿Los componentes del programa están siendo producidos al mínimo costo posible? ¿Cuál es el costo real de producción de dichos componentes?

Por su parte, la evaluación de la economía de un programa público debe responder interrogantes como las siguientes: ¿Está el programa administrando adecuadamente los recursos financieros que recibe del Estado? ¿La gerencia del programa ha mostrado capacidad para atraer recursos de otras fuentes? ¿El programa está utilizando algún mecanismo que le permita recuperar parte

de los costos de producción de sus componentes?, de no ser así, ¿es viable plantear este tipo de mecanismos a futuro?

Finalmente, para el caso de la Evaluación en Profundidad, evaluar la gestión interna del programa consiste en analizar y emitir juicios evaluativos acerca de aspectos tales como: proceso de toma de decisiones del programa, su estructura organizacional, asignación de responsabilidades, uso de mecanismos de evaluación y monitoreo, mecanismos utilizados para generar los componentes, criterios de focalización y selección de los beneficiarios, etc. Esta etapa de la evaluación en profundidad debe expresarse en una revisión y análisis detallado del proceso de producción de cada uno de los componentes del programa, sus etapas o sub-componentes, sus productos intermedios, los plazos y responsabilidades involucrados, así como la detección de actividades posibles de ser eliminadas. Dicha información servirá de insumo para analizar los resultados que posteriormente se obtengan al evaluar el impacto de los programas.

2. Los Resultados.

2.1. Aspectos cuantitativos y cualitativos versus mediciones cuantitativas.

Una evaluación de resultados debe abarcar tanto los aspectos cuantitativos como cualitativos de un programa. Tomando como ejemplo un programa de asistencia técnica para familias campesinas, los aspectos cuantitativos se relacionarán con variables tales como productividad de los cultivos, precio de venta de los productos, inversión predial, ingreso promedio de las familias, etc; por su parte los aspectos cualitativos se relacionarán con temas tales como la actitud de los campesinos frente a la innovación, grado de aprobación o satisfacción con la asistencia recibida, percepción optimista (o pesimista) respecto a su situación futura, etc.

Las dimensiones cuantitativas de la evaluación se expresan a través de variables cuantitativas (quintales de trigo por hectárea, miles de pesos reales mensuales, etc.). No obstante, los aspectos cualitativos también son susceptibles de mediciones cuantitativas. Por ejemplo sí el grado de aceptación de la asistencia técnica por parte de los campesinos es sistematizado por medio de una encuesta, el aspecto cualitativo “aceptación” tendrá una medición y expresión cuantitativa, por ejemplo: “porcentaje de campesinos que califican la asistencia técnica como MUY BUENA”.

2.2. Hipótesis causales.

La evaluación de impacto debe identificar las relaciones de causa-efecto entre los componentes que produce el programa y los resultados esperados. ¿Los bienes y servicios que entrega el programa contribuyen a lograr los beneficios esperados en la población usuaria?

Con relación a este punto se debe considerar que al evaluar un programa que

ya está en marcha, el equipo evaluador se encontrará con que los bienes y servicios que se producen son justificados por la gerencia del programa sobre la base de que contribuyen a lograr el resultado deseado, es decir, la gerencia del programa maneja explícita o implícitamente ciertas hipótesis de causa-efecto¹. El rol del equipo evaluador es identificar con precisión estas hipótesis e indagar sobre la real existencia de las relaciones causa-efecto en la aplicación del programa, validarlas si corresponde o rechazarlas. En este último caso el resultado de la evaluación será negativo puesto que el equipo evaluador habrá demostrado empíricamente que las relaciones causa-efecto supuestas por la gerencia no se han cumplido y, por lo tanto, no se han alcanzado los beneficios esperados.

De lo anterior se desprende que, en cualquiera de los dos casos, tanto si el resultado de la evaluación es positivo como negativo, el juicio evaluativo debe fundamentarse en la medición objetiva de los beneficios esperados.

Asimismo, el equipo consultor deberá corroborar la causalidad sobre posibles externalidades que generen los distintos componentes del programa a evaluar. Se debe definir la hipótesis de causalidad en base a información otorgada por el programa en cuestión o por información levantada mediante encuestas a los beneficiarios.

2.3 La dimensión temporal de los resultados (horizonte de evaluación).

Muchas veces la relación causa-efecto entre los componentes del programa y los beneficios esperados puede ser distinta dependiendo del período que se estudie. Es por esto que se reconoce la existencia de resultados de corto, mediano y largo plazo. El proceso de evaluación debe incluir estos distintos tipos de resultados e identificar y justificar el momento en que corresponde que éstos sean medidos y analizados.

Resultados a nivel de producto

Los productos se refieren a resultados del programa a nivel de componentes y responden a preguntas como: ¿Tuvieron las actividades realizadas los productos esperados, beneficiando al grupo objetivo predefinido?, ¿en qué magnitud?. Es decir, en una línea de tiempo corresponden al primer tipo de resultados que debe generar un programa y al efecto más inmediato que tiene la producción de los componentes sobre la población beneficiaria: familias subsidiadas, urgencias atendidas, textos escolares entregados, trabajadores capacitados, etc.

La calidad de los bienes y servicios entregados por el programa también puede considerarse como un resultado a nivel de producto. En este punto conviene distinguir, para fines evaluativos, entre la calidad objetiva y la calidad subjetiva. La primera se refiere a la calidad técnica de los bienes y servicios entregados; por ejemplo: calidad pedagógica de los textos escolares, aporte calórico de las raciones, tiempo promedio y máximo para la entrega de un servicio, etc.; en

¹ Por lo general las hipótesis causa-efecto que maneja la gerencia son declaradas o se puede deducir de los objetivos del programa.

tanto que, la segunda dice relación con la percepción que tienen los usuarios respecto al bien o servicio recibido; por ejemplo: porcentaje de aceptación de las raciones, grado de satisfacción de los usuarios, etc.

Resultados Intermedios

Los resultados intermedios corresponden al segundo tipo de resultados que se produce en una secuencia de tiempo. Se refieren a cambios en el comportamiento o actitud de los beneficiarios o certificación o cambio en el estado de éstos, una vez que han recibido los bienes y servicios que entrega el programa. Ejemplos de este tipo de resultados son: porcentaje de jóvenes con actitud crítica frente a las drogas después de participar en talleres de sensibilización, innovaciones organizacionales después de un programa de capacitación, agricultores que están desarrollando nuevos cultivos después de acceder a transferencia tecnológica, alumnos egresados, mediaciones laborales exitosas, etc.

Resultados Finales

Los resultados finales (impacto) son resultados a nivel de propósito o fin del programa. Implican un mejoramiento significativo y, en algunos casos, perdurable o sustentable en el tiempo, en alguna de las condiciones o características de la población objetivo que se plantearon como esenciales en la definición del problema que dio origen al programa. Un resultado final suele expresarse como un beneficio de mediano y largo plazo obtenido por la población atendida. En el caso en que la intervención del programa es una categoría de inversión en capital (físico, humano o social²), su beneficio no debería desaparecer, al menos en el corto o mediano plazo, si se deja de ejecutar o participar en el programa; por ejemplo: aumento en el nivel de remuneraciones de los jóvenes capacitados, aumento en la productividad promedio de las tierras manejadas y/o aumento en el nivel socio-económico de las familias en un programa de asistencia técnica, reducción de la tasa de reincidencia de jóvenes con problemas delictuales, etc³.

2.4 La influencia de factores externos.

Los resultados intermedios son eventos, comportamientos y cambios en las condiciones que ocurren como consecuencia directa de la acción del programa y, por esa razón, su logro depende principalmente de factores internos al programa, debido a que las variables involucradas están por lo general bajo el control de la gerencia. Lo mismo puede afirmarse respecto de los resultados a nivel de producto.

² En capital social lo constituyen las instituciones, set de relaciones, actitudes y valores que determinan las interrelaciones entre las personas, que contribuyen al desarrollo económico y social. El capital social emerge cuando hay valores sociales compartidos que sustentan relaciones cotidianas, generando sentimientos de identificación y pertenencia.

³ Para mayor información respecto de la sustentabilidad de los resultados del un programa ver documento "Notas Técnicas. Evaluación de Programas. XI. Resultados Finales o Impacto" en sitio web de la Dirección de Presupuestos, www.dipres.cl, Control de Gestión, publicaciones.

Los resultados a nivel de productos e intermedios tienen dos ventajas desde el punto de vista de la gestión del programa: generalmente ocurren antes que los resultados finales y además son más visibles y, por consiguiente más fácil su medición y evaluación. Por lo anterior, los resultados a nivel de productos e intermedios constituyen una importante fuente de retroalimentación sobre el desempeño del programa.

Los resultados finales en cambio, están más sujetos a la influencia de variables externas, es decir, es más probable encontrar que este tipo de resultados además responden a variables que escapan al control de la gerencia del programa, tales como: comportamiento general de la economía, cambios culturales, situaciones climáticas o medioambientales, situación de los mercados internos o externos, etc.

No obstante, lo anterior no significa que la influencia de los factores externos no pueda ser considerada en una evaluación, puesto que existen varias técnicas que permiten aislar y medir este tipo de efectos (ver punto 3.3. de este documento).

Es importante distinguir sin embargo, entre aquellos factores externos al programa que realmente están fuera del control de la gerencia y aquellos que mediante la acción del programa pueden hacerse más controlables. Por ejemplo, el efecto sobre los resultados finales de un programa que pueda tener la labor de otra agencia o programa público, no podría considerarse un factor externo incontrolable, puesto que bastaría generar los mecanismos de coordinación necesarios con esas instituciones para compatibilizar o complementar el efecto sobre los resultados finales.

3. Medición y evaluación de Resultados (Eficacia).

3.1 Medición de los resultados a nivel de producto.

La medición y evaluación de los resultados a nivel de producto se lleva a cabo principalmente mediante técnicas cuantitativas⁴. Lo más común es el uso de indicadores de gestión o indicadores de desempeño⁵, que miden la evolución de los niveles de producción y/o de los atributos o características más relevantes de los bienes o servicios entregados por el programa.

La evaluación de este tipo de resultados debe realizarse a partir de una definición amplia de producto. Lo anterior implica no limitarse a lo netamente físico o productivo, sino que incluir dentro del concepto de bien o servicio sus características o atributos más relevantes, por ejemplo: grado de aceptación por parte de los beneficiarios, satisfacción de los usuarios, oportunidad en la entrega, cobertura de la demanda potencial, grado de focalización efectiva, etc.

Para cada uno de los atributos o características del producto se debe generar

⁴ Aunque las variables involucradas sean de tipo cualitativo. Ver punto 2.1.

⁵ Para mayor detalle ver Notas Técnicas Evaluación de Programas, en sitio web de la Dirección de Presupuestos, www.dipres.cl, Sistema de Evaluación y Control de Gestión, publicaciones.

información que permita construir los indicadores y evaluar los resultados a este nivel. Una vez que se cuenta con la base de indicadores existen varias posibilidades de análisis.

La primera alternativa es comparar el desempeño de la institución respecto a sí misma.

Una segunda posibilidad es evaluar los resultados en función de las metas que se ha planteado el programa. Este tipo de análisis, debe tener en cuenta que la calidad de las conclusiones y juicios evaluativos que se puedan realizar, dependerá directamente de la existencia y pertinencia de dichas metas⁶.

Finalmente, se pueden comparar los resultados del programa con estándares técnicos o de calidad preestablecidos. Por ejemplo, en varias áreas del quehacer de los programas públicos, existen estándares internacionales para ciertos tipos de servicios (estándares de la OPS, normas ISO 9000, etc.). Por otra parte, un programa público puede auto-imponerse estándares de calidad en función de las expectativas de sus usuarios o beneficiarios; para este fin, existen mecanismos como la *carta ciudadana* y otros similares, mediante los cuales el prestador garantiza a sus usuarios ciertos niveles mínimos de calidad en los servicios entregados.

3.2. Medición de los resultados intermedios.

La medición de resultados intermedios apunta a evaluar los cambios en el comportamiento de los beneficiarios que se producen una vez que estos han utilizado y asimilado los componentes del programa. No mide los beneficios finales alcanzados, puesto que éstos requieren un período de tiempo mayor para ser logrados (resultados finales), sino que más bien analiza la adquisición de ciertas conductas, habilidades o capacidades que presumiblemente facilitarán la generación de los beneficios de largo plazo.

No todos los programas públicos deben presentar resultados intermedios. Es probable que existan iniciativas que, por su naturaleza, solamente exhiban resultados finales y a nivel de productos, por ejemplo los proyectos de infraestructura vial, las campañas de vacunación, etc.

En términos generales, se esperan resultados intermedios en los programas que implican la formación de algún tipo de capital humano o social, por ejemplo: ¿los beneficiarios de una obra de regadío están administrando adecuadamente las nuevas instalaciones?, ¿los profesores están utilizando adecuadamente en sus clases los textos de estudio entregados a los estudiantes?, ¿los beneficiarios de un programa de transferencia tecnológica están aplicando los conocimientos y las técnicas adquiridas?, etc.

⁶ La pertinencia de las metas se refiere a que éstas, para que sean buenos referentes en el proceso de evaluación, deben ser un fiel reflejo de los objetivos que persigue el programa. Además, deben ser exigentes en el sentido de representar mejoramientos significativos en los factores evaluados y, a la vez, ser lo suficientemente realistas como para que puedan ser logradas con los recursos con que cuenta el programa.

La medición de los resultados intermedios implica necesariamente indagar, a nivel de los beneficiarios efectivos, si se han producido los cambios de comportamiento deseados, por esta razón generalmente se requiere utilizar técnicas de investigación de campo, tales como encuestas o inspecciones en terreno.

Lo anterior no significa que para la medición de los resultados intermedios no sea necesaria la construcción de indicadores de desempeño. La diferencia radica en que en el caso de los resultados a nivel de producto los indicadores serán cuantificados principalmente en base a información existente en el programa, mientras que en caso de los resultados intermedios, en algunos casos se requerirá del uso de instrumentos de recolección de información que no forman parte del proceso regular de recolección de información que maneja el programa⁷.

3.3 Medición de resultados finales⁸.

La medición de resultados finales o impacto debe establecer en primer lugar cuales son los efectos de mediano y largo plazo sobre la población beneficiaria que espera obtener el programa. ¿Cuáles son los beneficios con que quedará la población beneficiaria una vez que concluya su participación en el programa?, por ejemplo: incremento en el nivel de remuneraciones de trabajadores capacitados, reducción de la tasa de morbilidad de menores beneficiados con un programa de vacunación, aumento del rendimiento escolar de los alumnos de escuelas participantes en un programa de reforzamiento educativo, etc. Lo anterior implica identificar y separar los beneficios logrados como consecuencia de la intervención del programa, de aquellos que de todas maneras hubieran ocurrido sin la ejecución de éste, por efecto de la evolución normal de las condiciones del entorno o la acción de otros programas públicos o privados.

Para ello, en primer lugar, corresponderá identificar la(s) variable(s) que deben ser medidas para determinar si el programa ha generado realmente los beneficios o resultados finales o impacto esperados. Cabe destacar que esta misma metodología debe ser aplicada para evaluar los resultados intermedios y externalidades del programa en cuestión.

De este modo, el impacto del programa corresponde a la diferencia en la variable resultado (Y) que registra el individuo i con y sin la intervención del programa.

⁷ Para mayor información respecto de la construcción y uso de indicadores de desempeño ver documento "Notas Técnicas. Evaluación de Programas. XII. Indicadores de Desempeño" en sitio web de la Dirección de Presupuestos, www.dipres.cl, Sistema de Evaluación y Control de Gestión, publicaciones.

⁸ Las metodologías de estimación de resultados finales aquí descritas pueden ser también válidas para la estimación de los resultados intermedios o externalidades de un programa, si corresponde al tipo de resultados intermedios esperados que permite aplicar estas herramientas.

Y_{1i} = resultado potencial del individuo i con tratamiento (resultado que el individuo i obtendría si se le expusiera al tratamiento).

Y_{0i} = resultado potencial del individuo i sin tratamiento (resultado que el individuo i obtendría si no se le expusiera al tratamiento).

$$\alpha_i = Y_{1i} - Y_{0i} \quad (1)$$

Donde α_i es el impacto del programa (tratamiento) sobre la persona i . Sin embargo, no es posible observar Y_{0i} e Y_{1i} para un mismo individuo en un momento del tiempo, es decir, un individuo i puede presentar sólo uno de los dos estados posibles, intervención en el programa (tratamiento) o no intervención en el programa. Esto es conocido como el “problema fundamental de identificación”.

Una solución a este problema sería medir el impacto esperado o promedio del programa sobre el conjunto de los individuos:

$$\alpha = E[Y_1 - Y_0] \quad (2)$$

Donde α es el impacto promedio del programa (Average Treatment Effect (ATE)), Y_1 es el resultado potencial, con posterioridad a la intervención del programa, de los individuos que han participado en dicho programa; e Y_0 es la variable resultado, con posterioridad a la intervención del programa, de los individuos que no han participado en el programa, es decir

$$E[Y | D=1] - E[Y | D=0] = E[Y_1 | D=1] - E[Y_0 | D=0] = E[Y_1 - Y_0] + \{E[Y_0 | D=1] - E[Y_0 | D=0]\} \quad (3)$$

Donde, $D = 1$ si el individuo ha participado en el programa (es decir, recibe tratamiento), y $D = 0$ de lo contrario.

El último término de la ecuación anterior probablemente no será igual a cero debido a la existencia de sesgo de selección, es decir, la selección de individuos que participan en el programa responde a ciertos criterios de focalización u otros. Por ejemplo, en general los beneficiarios de un programa de capacitación poseen ingresos menores antes de la capacitación que aquellos individuos que no participan del programa. Lo anterior posiblemente implica que, en ausencia del programa, los ingresos potenciales de los beneficiarios serían menores a los de los no beneficiarios en el mismo período, por lo tanto, en este ejemplo, si se utiliza la diferencia de medias como el impacto del programa este resultado estaría sub estimado.

$$(E[Y_0 | D=1] - E[Y_1 | D=0] < 0).$$

Para resolver el problema ocasionado por la existencia de sesgo de selección, y por ende, aislar los efectos que sobre la variable resultado (Y) tienen factores

externos al programa, se requiere que la selección de individuos beneficiarios del programa sea independiente de sus resultados potenciales. El cumplimiento de esta condición implica que $E[Y_0 | D = 1] = E[Y_0 | D = 0]$, y por lo tanto, el resultado promedio o impacto esperado del programa (tratamiento) será:

$$E[Y_1 - Y_0] = E[Y_1 - Y_0 | D = 1] = E[Y | D = 1] - E[Y | D = 0] \quad (4)$$

Para que se cumpla la condición de independencia, se requiere que el tratamiento sea asignado aleatoriamente, es decir, que los individuos que se beneficien del programa sean seleccionados de forma aleatoria, lo cual se conoce como diseño experimental o aleatorio.

3.3.1 Diseño experimental o aleatorio

Es considerada la metodología de evaluación técnicamente más robusta y consiste en la selección aleatoria de los beneficiarios dentro de un grupo de individuos elegibles, el proceso de asignación aleatoria de las intervenciones o servicios del programa crea dos grupos estadísticamente idénticos entre sí, uno que participa en el programa (grupo de tratamiento, $D_i = 1$) y otro que, cumpliendo todas las condiciones para participar, está fuera de él (grupo de control⁹, $D_i = 0$).

La medición de impacto consiste entonces en cuantificar, una vez transcurrido el tiempo pertinente de intervención del programa, la(s) variable(s) de impacto, tanto para el grupo de tratamiento como para el grupo de control – en ambos simultáneamente y para el mismo período de tiempo – para luego analizar las diferencias que existen entre ambos.

En términos operativos, esta medición se debe hacer sobre muestras estadísticamente representativas¹⁰ de ambos grupos, y calcular el impacto promedio del programa sobre una variable resultado (Y).

La asignación aleatoria de los beneficios es la principal ventaja de este método, debido a su fortaleza estadística. En dicho caso, un **mecanismo aleatorio** divide a los individuos en dos grupos: un grupo de TRATAMIENTO y un grupo de CONTROL. Es este mecanismo al azar es el que permite estimar el efecto promedio.

Lo anterior implica una gran facilidad para interpretar los resultados de las evaluaciones, puesto que se puede estimar el impacto de un programa mediante una simple diferencia de medias (para la o las variables de resultado relevantes) entre el grupo de control y el grupo de tratamiento.

⁹ Un grupo de control es básicamente un segmento de no beneficiarios que son parte de la población objetivo, es decir, un grupo que comparte las características sociales, económicas, educacionales, etc. del grupo beneficiario. La única diferencia entre los individuos del grupo de control y los beneficiarios efectivos es su participación en el programa.

¹⁰ Para mayor detalle ver “Protocolo de Levantamiento de Información en Terreno” en www.dipres.cl Sistema de Evaluación y Control de Gestión, publicaciones. (revisar si corresponde al número del capítulo de la última versión de Notas Técnicas)

$$\hat{\alpha} = \frac{1}{n_1} \sum_{D_i=1} Y_i - \frac{1}{n_0} \sum_{D_i=0} Y_i \quad (5)$$

Donde, n_1 es el número de individuos en el grupo tratamiento y n_0 el número de individuos en el grupo control.

Basados en el teorema del límite central, se puede mostrar que el anterior es un estimador insesgado y consistente, es decir, asintóticamente¹¹:

$$\hat{\alpha} \Rightarrow \alpha \quad (6)$$

El estimador del efecto promedio del tratamiento también puede ser obtenido a través de la estimación vía Mínimos Cuadrados Ordinarios de la siguiente expresión:

$$Y_i = \beta_0 + \beta_1 * Z_i + v_i \quad (7)$$

En efecto:

$$\hat{\beta}_1 = \hat{\alpha} \quad (8)$$

La principal desventaja del diseño experimental es de orden práctico y no metodológico, y se refiere a la viabilidad política de excluir de la cobertura de un programa, aleatoriamente, a un grupo de beneficiarios elegibles y, que como tales, necesitan los servicios del programa. Por lo tanto muchas veces se cuestiona la asignación aleatoria desde el punto de vista público en base a razones éticas¹².

En la práctica, muchos programas entregan los servicios del programa diferidos en el tiempo a distintos grupos de beneficiarios seleccionados aleatoriamente. En este caso el grupo de control es el grupo de beneficiarios que aún no ha participado en el programa y el grupo de tratamiento es el grupo que ha recibido los beneficios, siempre y cuando, la diferencia de tiempo en la entrega de estos servicios entre ambos grupos sea suficiente para poder medir los resultados finales o de impacto en el momento correspondiente. Este caso se puede dar cuando hay lista de espera en la postulación a un beneficio, cuando hay restricciones presupuestarias que establecen la entrada en vigencia del tratamiento de manera gradual en el tiempo y en el contexto de un programa piloto.

Por último, se debe tener presente que un diseño experimental es una técnica muy intensiva en tiempo. La postergación de los resultados de la evaluación dependerá del tiempo que se estime pertinente para que un programa produzca impactos en la población beneficiaria, lo que en general significa

¹¹ Para conocer la significancia estadística se utiliza un simple test t, comúnmente usado para evaluar las diferencias entre dos muestras.

¹² Ver Heckman y Smith (1995) y Burtless (1995) para mayor discusión al respecto.

planificar una evaluación con varios años de anticipación.

Desde el punto de vista metodológico, los principales problemas asociados a un diseño no experimental provienen de **los posibles cambios que puede sufrir el grupo de control o tratamiento durante el período de evaluación**. Por ejemplo: i) algunos de los miembros del grupo de control pueden abandonar el área geográfica durante la evaluación, lo que impediría contar con información de éstos sesgando, de esta forma, los resultados, ya que el grupo de control no será comparable con el grupo de tratados; ii) los individuos del grupo de tratamiento no continúan con la intervención, en este caso tan sólo se podría evaluar el impacto de una oferta de determinado beneficio, más que de recibir el tratamiento, en caso de hacer esta inferencia, los resultados estarán sesgados. Ambos problemas se denominan desgaste o atrición de la muestra y condicionan **la validez interna** de los resultados, es decir, ponen en duda su valor. Lo mismo puede ocurrir si los individuos del grupo de control o tratamiento han recibido otro tipo de intervenciones con similares objetivos al programa a evaluar, durante el período de evaluación que no están siendo controladas, lo cual sesgaría los resultados y condicionaría la validez interna de éstos.

Otro problema que se puede presentar con esta metodología es el de **validez externa**, que se refiere a la imposibilidad de generalizar los resultados de la evaluación para la población objetivo en su conjunto. Por ejemplo, cuando las muestras no son representativas¹³, o los programas no son representativos, ya sea por un efecto de escala o si el tratamiento difiere de la implementación planificada.

3.3.2 Métodos Cuasi-Experimental

Por lo general la selección de los beneficiarios de un programa no es aleatoria, es decir, no es posible tener una aleatorización de la asignación del tratamiento. Por el contrario, su selección se basa en la aplicación de criterios de elegibilidad y focalización que establecen diferencias, tanto observables como no observables¹⁴, entre éstos (grupo de tratamiento) y los no-beneficiarios (grupo control).

Esto significa que el impacto del programa no podrá ser estimado a través de la

¹³ Al plantear un estudio experimental, lo primero que se debe evaluar es si el tamaño muestral a considerar permitirá detectar el efecto esperado del tratamiento al nivel que se desea obtener inferencia causal, por ejemplo: a nivel de beneficiarios a nivel nacional, regional por género, entre otros.

¹⁴ Las diferencias observables proviene de la aplicación de los propios criterios de selección y focalización aplicados por el programa, puesto que al concentrar las intervenciones sobre la población que presenta en mayor magnitud el problema o es de más alto riesgo, se distorsiona cualquier comparación posterior entre los grupos de beneficiarios y no beneficiarios. Las diferencias no observables son todas aquellas características de la población que afectan su participación en el programa, pero para las cuales no se cuenta con información (habilidad individual, voluntad de participar y en general procesos subjetivos y no formales que determinan que una persona acceda a los beneficios del programa).

simple diferencia de medias entre la variable resultado del grupo de tratamiento y el grupo control (ecuación 5), ya que las diferentes características observables y no observables de los beneficiarios y no beneficiarios implicará la existencia de sesgo de selección y por ende, $E[Y_0 | D = 1] \neq E[Y_0 | D = 0]$, la medición del impacto del programa resultará sesgada. Dependiendo de la magnitud y el signo del sesgo, se puede llegar a subestimar o sobrestimar el impacto de un programa. En el extremo, se pueden evaluar positivamente los resultados finales de una intervención cuando éstos son negativos o viceversa.

Inspirados en el modelo experimental, se han desarrollado diversos métodos, los cuales se denominan métodos cuasi-experimentales, estos son: i) método de experimentos naturales, ii) construcción de grupos artificiales de control vía procedimientos de “matching”, iii) identificación de efectos causales por **medio de variables instrumentales** y iv) regression discontinuity design.

Estos permiten controlar por las distintas características que posee el grupo de beneficiarios del programa (tratamiento) respecto de los no beneficiarios (grupo control) que pudiesen afectar el resultado o impacto del programa.

Las diversas alternativas metodológicas, coinciden en intentar simular las condiciones de un diseño experimental a partir de bases de datos de beneficiarios y no-beneficiarios y usando avanzadas técnicas estadísticas. Se aplican una vez que la intervención del programa ya tuvo lugar, es decir, una vez que ya se seleccionaron los beneficiarios del programa a través de sus propios criterios sin considerar métodos aleatorios.

Por esta razón, en un diseño cuasi-experimental los grupos de “control” y “tratamiento” no son iguales entre sí y, por lo tanto, el corazón de esta metodología de evaluación consiste en aplicar sofisticados controles estadísticos para eliminar o minimizar estas diferencias.

Lo anterior implica la construcción de dos escenarios. Un escenario sin programa o contrafactual y un escenario con programa. La construcción del contrafactual se logra a través de un grupo de control formado por individuos, iguales o muy parecidos a los beneficiarios, y cuya única diferencia con éstos sea “no haber participado en el programa”.

De este modo se busca responder lo más certeramente posible a la pregunta: ¿Cuál sería la situación actual de estos beneficiarios si no hubieran participado en el programa?, comparando los resultados finales del programa en los beneficiarios con respecto al grupo de control, de modo de calcular el efecto neto o impacto atribuible al programa, “limpio” del efectos de otros factores externos al programa.

Una condición necesaria para el desarrollo de un diseño cuasi-experimental, es la existencia de una base de datos (o varias bases que permitan formar una común) de beneficiarios y no-beneficiarios, que contenga para ambos grupos información sobre los criterios de elegibilidad del programa (por ejemplo: condición socioeconómica, tamaño de la familia, escolaridad, ubicación geográfica o cualquier otra característica que se haya utilizado para seleccionar

a los beneficiarios) e información sobre las variables que permitirán medir los resultados finales o impactos esperados de la intervención (por ejemplo: nivel de ingresos, condición laboral, nivel de salud, etc.).

Una vez contruidos los grupos control y de tratamiento, y teniendo observaciones de las variables de resultados y las variables que caracterizan al individuo y su entorno en dos momentos del tiempo (antes y después de la intervención del programa), por lo general se cuantificará el impacto del programa a través de la estimación econométrica del siguiente modelo. El estimador de Diferencias en Diferencias se puede obtener estimando la siguiente ecuación testeando **las propiedades de Mínimos Cuadrados Ordinarios**, donde el estimador del parámetro α ($\hat{\alpha}$) corresponde al impacto del programa:

$$Y_{it} = \beta_1 + \beta_2 DG_i + \beta_3 DT_i + \alpha(DG_{it} * DT_{it}) + X_{it} * \gamma + \varepsilon_{it} \quad (9)$$

Donde,

Y_{it} = variable resultado para el individuo i en el período t.

DG_i = variable igual a 1 si individuo i pertenece al grupo de tratamiento y 0 si pertenece al grupo control.

DT_i = variable igual a 1 si el dato de la variable resultado para el individuo i pertenece al período post-programa y 0 de lo contrario.

X_{it} = conjunto de variables que caracterizan la individuo y su entorno.

ε_{it} = término de error aleatorio

La estimación de modelo de la ecuación (6) es lo que se conoce como método de doble diferencia, donde el impacto promedio del programa (ATE) se estima primero midiendo el cambio en las variables de resultado experimentado por cada grupo, o más precisamente por sus muestras, entre la línea base y la encuesta de seguimiento (primera diferencia), y luego comparando el cambio mostrado por el grupo de beneficiarios con el registrado en el grupo de no-beneficiarios (segunda diferencia).

Este es el método técnicamente más robusto puesto que elimina el efecto de características no observables que afectan la selección de los beneficiarios (por ejemplo habilidad, motivación, etc) y por lo tanto, es la mejor forma de tratar el sesgo de selección, dado que se puede comparar el cambio registrado en el grupo de tratamiento con el cambio exhibido por el grupo de control. Esto permite eliminar los efectos generados por factores exógenos (efectos de incremento en desempleo, crisis económica, factores de la naturaleza) en la variable de resultado, que impactan en igual magnitud al grupo de control y tratamiento en el período evaluado¹⁵. No obstante, es el más exigente en términos de información puesto que los estudios de línea de base no son muy comunes y los estudios de seguimiento, por lo general, enfrentan el problema

¹⁵ En un diseño experimental el estimador de diferencias en diferencias, debiese ser similar al de una diferencia, en el caso de no presentarse ningún problema de validez interna.

de que es difícil localizar a los individuos de la muestra original, especialmente en el caso de los no beneficiarios.

La ventaja de la formulación anterior es que se puede agregar un vector de variables explicativas que permite controlar por diferencias observables existente entre el grupo de tratamiento y el de control antes del programa, que aún persistan (particularmente relevante cuando el grupo de control no viene de una asignación aleatoria). Sin embargo, si el coeficiente de la variable interactiva varía enormemente una vez que se controla por nuevas variables puede ser indicativo de que se tiene un deficiente grupo de control.

En algunos estudios es imposible poder construir grupos de control de calidad, lo cual ha llevado a los investigadores a proponer un grupo de control adicional para verificar los resultados obtenidos, es decir, contar con dos grupos de control en la misma estimación. Esto da lugar a lo que se ha denominado **método de diferencias en diferencias en diferencias**¹⁶ (**D-D-D**). Por ejemplo, cuando un programa se aplica en una determinada área (área 1) y no en otra área (área 2), este método (D-D-D) consistiría en calcular un estimador de diferencias en diferencias (DD) entre el grupo que obtuvo tratamiento del área 1 y el grupo de control del área 1 y adicionalmente en utilizar el estimador de DD para la segunda área entre el grupo tratamiento, que en realidad no ha recibido tratamiento pero posee las mismas características observables que el grupo de tratamiento del área 1, y el grupo de control del área 2 como un segundo “contrafactual”, es decir, con las mismas características observables que el grupo de control del área 1.

Esta estimación se obtiene de la siguiente forma:

$$\hat{\beta}_8 = \frac{[\bar{Y}_t^T - \bar{Y}_{t-1}^T] - [\bar{Y}_t^C - \bar{Y}_{t-1}^C]}{[\bar{Y}_t^T - \bar{Y}_{t-1}^T] - [\bar{Y}_t^C - \bar{Y}_{t-1}^C]} E1 - \frac{[\bar{Y}_t^T - \bar{Y}_{t-1}^T] - [\bar{Y}_t^C - \bar{Y}_{t-1}^C]}{[\bar{Y}_t^T - \bar{Y}_{t-1}^T] - [\bar{Y}_t^C - \bar{Y}_{t-1}^C]} E2$$

$$Y_{it} = \beta_1 + \beta_2 DG_{iet} + \beta_3 DT_{iet} + \beta_4 DE_{iet} + \beta_5 DG_{iet} DT_{iet} + \beta_6 DG_{iet} DE_{iet} + \beta_7 DT_{iet} DE_{iet} + \beta_8 (DG_{iet} * DT_{iet} * DE_{iet}) + \varepsilon_{it} \quad (10)$$

$$\alpha^{d-d-d} = \hat{\beta}_8$$

$DE_{iet} = 1$ si la observación corresponde al área E1

0 si la observación corresponde al área E2

$DG_{iet} = 1$ si el individuo pertenece al grupo que recibe tratamiento

0 si el individuo no pertenece al grupo que recibe tratamiento

$DT_{iet} = 1$ si la observación corresponde al periodo T después del tratamiento

0 si la observación corresponde al periodo T-1 antes del tratamiento

¹⁶ Para mayor detalle sobre esta metodología ver, Gruber, J (1994) The Incidence of Mandated Maternity Benefits, The American Economic Review.

3.3.2.1 Experimentos Naturales.

La metodología de estimación de experimentos naturales, es decir, cuando la asignación del tratamiento está ***dada por variaciones exógenas producidas por cambios legales, fenómenos naturales u otras, puede ser obtenida a partir de una estimación de diferencias en diferencias*** con controles tal como está explicitado en los modelos anteriores.

3.3.2.2 Método de Matching.

Consiste en una construcción de grupos de control mediante procedimientos estadísticos de “matching”. Esta metodología corrige las diferencias observables entre el grupo de tratamiento (beneficiarios del programa) y el grupo de control (no beneficiarios), buscando para cada individuo de la muestra del grupo de tratamiento a la unidad muestral más parecida de la muestra de no beneficiarios, los cuales finalmente conformarán el grupo de control.

El principal supuesto de esta metodología es que la selección se basa en elementos observables. Si este no es el caso los resultados obtenidos a través de esta metodología estarán sesgados. El sesgo se origina de la correlación entre las variables no-observables que afectan la selección para el programa y que afectan la variable de interés de la evaluación. En el caso en que la heterogeneidad no varíe en el tiempo, la estimación en doble diferencia puede eliminar esta limitación.

La aplicación de este método requiere información de corte transversal para ambos grupos, que incluya **las variables que se utilizaron para seleccionar a los participantes del programa**, las que deben ser variables de pre-tratamiento, con el fin de garantizar que no han sido afectadas por este, y las variables de resultado relevantes. Se necesita por lo general una base de datos amplia como la obtenida de una encuesta de carácter nacional (CASEN, Censo Agropecuario, Encuesta de Ingresos Familiares, etc.).

El más conocido de estos métodos es el Propensity Score Matching, que consiste en modelar estadísticamente la participación en el programa (variable P en la ecuación 11) y luego calcular para los individuos de ambas muestras, tanto la de del grupo de tratamiento como la del grupo de comparación, la probabilidad de participar en el programa dadas sus características de elegibilidad (pobreza, educación, tamaño de la familia y cualquier otro criterio de selección del programa para el cual se cuente con información).

$$P_i = \phi + \delta X_i + v_i \quad (11)$$

Donde:

P_i : es igual a 1 si el individuo i participa en el programa y 0 de lo contrario.

X_i : son las variables pre-tratamiento afectan la participación en el programa.

v_i : es un término de error aleatorio

Al estimar el modelo de la ecuación 11 mediante un modelo probit o logit, se obtiene para cada individuo de la muestra un score que refleja su probabilidad condicional de participar en el programa. Luego, se debe verificar que se cumpla la condición de "soporte común", vale decir, que exista un rango común para los p-score de ambos grupos que permita la comparación. Posteriormente, para cada uno de los beneficiarios se busca al o los no-beneficiarios con un score más cercano, se calculan para ambos el valor esperado de las variables de resultado (valor esperado de Y) esto se hace mediante las técnicas de matching existentes: vecino más cercano, kernel, y metodología de radios. Una vez obtenido un grupo de comparación se puede estimar el impacto del tratamiento vía Diferencias-en-Diferencias, lo que requeriría contar con datos de panel, y si esto no es posible la estimación se deberá hacer sólo con una diferencia.

La principal ventaja de este método es que permite hacer evaluaciones de impacto una vez que el programa ya está en marcha y usando información de corte transversal (single cross section data), por lo cual, si se cuenta con la información apropiada, se pueden obtener mediciones de impacto en tiempos relativamente cortos. Metodológicamente es robusta puesto que permite eliminar el sesgo *causado por las diferencias observables*, aunque no soluciona el sesgo de selección causado por las diferencias no observables, las cuales debiesen ser controladas para obtener resultados insesgados y consistentes.

3.3.2.3 Método de Variables Instrumentales.

Para abordar el sesgo derivado de las características no observables, aun habiendo controlado por todas las características pre-existentes observables entre participantes y no participantes consideradas en el proceso de selección, es probable que atributos no observados lleven a que el grupo de participantes y no participantes en el programa difieran antes de que se aplique el tratamiento, lo cual llevaría a obtener una estimación del impacto sesgada. Para solucionar este problema generalmente se utilizan variables instrumentales (VI).

Este sesgo se deriva de la existencia de factores que afectan la participación en el programa (P_i) y que el investigador desconoce, por ejemplo disposición a participar en el programa de capacitación por parte de individuos de mayor habilidad. En términos estadísticos lo anterior implica que en la ecuación:

$$Y_i = \lambda_1 + \lambda_2 DG_i + X_i * \varphi + \mu_i \quad (12)$$

El término de error μ_i contiene variables no especificadas en la matriz X_i que afectan la participación en el programa (P_i).

Para corregir este sesgo se necesita utilizar el valor estimado de $P = \hat{P}_i$, de acuerdo a la expresión 11, pero incluyendo una variable Z que determine la participación en el programa (condición de relevancia), y que no tenga incidencia en los resultados del mismo. Es decir, una variable que no esté en X y que no esté correlacionada con el término de error μ_i (condición de

exogeneidad), esta variable es la denominada variable instrumental. Luego el estimador de este instrumento se reemplaza por DG_i en la ecuación 12 y su efecto es el que permite obtener un estimador insesgado $\hat{\lambda}_2$ del impacto del programa.

$$Y_i = \lambda_1 + \lambda_2 \hat{P}_i + X_i * \varphi + \mu_i \quad (13)$$

Las ventajas de esta opción son evidentes, si se cuenta con una buena variable instrumental - **y se aplican paralelamente técnicas que permitan eliminar el sesgo observable** - se puede, a partir de información de corte transversal, obtener una estimación insesgada del impacto de un programa. Lamentablemente, en la práctica, los estudios de impacto por lo general enfrentan restricciones de información para obtener buenos instrumentos.

3.3.2.4 Regression-discontinuity¹⁷.

La principal característica que debe presentar un programa para hacer uso de esta metodología es que se utilice una discontinuidad en la regla de asignación del tratamiento al grupo de beneficiarios, ya que ésta será utilizada como una asignación aleatoria en un vecindario para estimar el impacto del programa.

La intuición que está detrás de esto es que los individuos que están justo arriba del corte utilizado para asignar el tratamiento, son muy parecidos a los individuos que han sido clasificados justo por debajo del corte. Por lo tanto, para la evaluación de impacto se compararán los resultados obtenidos entre ambos grupos de individuos tratados y no tratados.

Cabe destacar que para poder utilizar esta metodología de evaluación debe ser posible ordenar a los beneficiarios y no-beneficiarios en base a un índice construido a partir de variables objetivas y cuantificables.

La principal limitación de esta metodología es que entrega un estimador del impacto del tratamiento que es de carácter local, específicamente para los individuos que están en torno a los puntos de corte.

La ecuación a estimar debe ser:

$$Y_i = \lambda_1 + \lambda_2 D_i + \varphi \text{ Puntaje} + \mu_i \quad (14)$$

$D_i = 1$ si el postulante recibe tratamiento.

$D_i = 0$ si el postulante no recibe tratamiento

¹⁷ Este diseño metodológico fue introducido en la literatura por Thistlewaite y Campbell (1960).

Si el parámetro λ_2 es significativo y positivo, personas por debajo del puntaje de corte presentan un resultado menor que las personas que se ubican por encima (y cerca) de éste, es decir, el programa tuvo un impacto positivo.

En el caso en que el tratamiento dependa también de otras variables endógenas y la discontinuidad esté altamente correlacionada con el tratamiento, se debe utilizar el índice como una Variable Instrumental para la asignación al tratamiento, ya que al ser una variable endógena los parámetros estimados serán sesgados e inconsistentes¹⁸.

Estimación por variable instrumental¹⁹:

Primera etapa:

$$D_i = \lambda_1 + \lambda_2 DP_i + \varepsilon_i \quad (15)$$

$DP = 1$ si el individuo posee un puntaje mayor que el corte
 0 si el individuo posee un puntaje menor que el corte

Segunda etapa:

$$Y_i = \lambda_1 + \lambda_2 \hat{D}_i + \varphi \text{ Puntaje} + \mu_i \quad (16)$$

Finalmente, cabe destacar que independiente del método que se utilice, la validez de los resultados obtenidos por medio de herramientas cuantitativas dependerá de la calidad de la información con que se cuente.

Heterogeneidad en el Impacto de un programa

La generación de un grupo de control (escenario contrafactual) y de un grupo de tratamiento generalmente se utiliza para estimar el impacto promedio sobre la población de beneficiarios, sin embargo también pueden ser utilizados para indagar respecto de la heterogeneidad de efectos causales del programa para distintos sub-grupos de la muestra.

¹⁸Un buen instrumento debe estar correlacionado con la variable endógena, en este caso, la variable endógena es si el postulante participa o no del programa que será evaluado (condición de relevancia, se debe testear mediante el test F (significancia global) del modelo de la primera etapa). Asimismo el instrumento que se seleccione debe ser exógeno, es decir, no debe estar correlacionado con el término de error (condición de exogeneidad, se debe testear con Test de sobreidentificación). Mediante esta forma de estimación se elimina la correlación que existe entre la participación en programa y el término de error.

¹⁹ Un ejemplo de esto se puede encontrar en "Efecto del tamaño de clases sobre el rendimiento escolar" (Angrist y Lavy, 1999).

La estimación del impacto promedio de un programa pudiese significar promediar impactos positivos y negativos de sub-grupos de la población beneficiaria, no permitiendo, en ocasiones, observar impactos de distinto signo para subgrupos de la muestra. Obviar la posible heterogeneidad en resultados impide conocer si existen sub conjuntos de hogares o personas que pueden estar siendo perjudicados por el programa, cuando el resto de los beneficiarios goza de los resultados positivos de éste.

Las dos metodologías más utilizadas para evaluar la posible heterogeneidad de resultados de un programa sobre una población de beneficiarios son el Quantil Treatment Effect (QTE) y el Marginal Treatment Effect (MTE).

Quantil Treatment Effect (QTE):

El QTE para un percentil p se estima como la diferencia sobre la condición de tratamiento entre el percentil p de la distribución de tratados con el mismo percentil de aquellos no tratados. Por ejemplo, si se toma la mediana de la distribución del grupo tratado y se le resta la mediana de la distribución del grupo de control, se tiene el QTE en el percentil 0.5. Los demás QTE para otros percentiles se estiman análogamente.

Los QTE son características de las distribuciones de probabilidades marginales $F_0(y)$ y $F_1(y)$. Para la asignación de tratamiento t , el percentil p de la distribución F_t está definido como $y_p(t) \equiv \inf\{y : F_t(y) \geq p\}$. El efecto causal para el percentil p es de $\Delta_p = y_p(1) - y_p(0)$. En términos intuitivos, la técnica de QTE es equivalente a una estimación ATE para cada uno de los percentiles de la distribución, pero estimada a través de una función no paramétrica.

Si bien la estimación del efecto causal QTE es bastante simple, no es posible decir lo mismo con respecto a la estimación de la matriz de varianza-covarianza de los parámetros estimados.

Por el hecho de que se estiman efectos causales para pequeños grupos de la muestra total, el número de datos de la estimación se reduce especialmente en las colas de la distribución. Además, producto de que es posible que hayan efectos con signos diferentes en distintos tramos de la distribución de la variable evaluada, también es posible que existan cambios en el ranking de individuos tratados. Esto se traduce en que la estimación puede estar computando resultados de percentiles con características diferentes entre grupo de control y de tratamiento, independientemente de que las muestras totales de grupo de control y tratamiento sea estadísticamente comparables²⁰. Estos dos fenómenos se traducen en la necesidad de estimar errores estándar mediante métodos de simulación numérica, tales como bootstrap.

El método de bootstrap asume que la muestra observada es la poblacional y mediante una iteración de muestreo aproxima las propiedades de muestras

²⁰ La asignación del tratamiento haya sido aleatoria.

pequeñas de cada percentil, permitiendo realizar inferencia estadística confiable a través del uso de intervalos de confianza.

Es importante considerar que el nivel de desagregación (subgrupos) de los beneficiarios para la estimación del efecto causal dependerá de la cantidad de datos de la muestra original. Mientras menos datos existan, menor es el grado de heterogeneidad posible de analizar, dado los requisitos mínimos de grados de libertad necesarios para la obtención de estimaciones mínimamente precisas. En el caso de muestras pequeñas, puede que solamente sea posible estimar por deciles o quintiles.

Marginal Treatment Effect (MTE):

Otra forma de medir la heterogeneidad del efecto causal de un programa es mediante la estimación Marginal Treatment Effect.

A diferencia del QTE, esta metodología no estima una función para toda la distribución sino que estima parámetros para percentiles.

Considerando el modelo de potenciales resultados:

$$Y^1 = X\beta^1 + U^1, \quad Y^0 = X\beta^0 + U^0, \quad D^* = Z\theta + U^D,$$

siendo las dos primeras especificaciones ecuaciones de resultado en dos posibles estados²¹, y donde la tercera especificación es la regla de decisión²², el parámetro MTE mide el efecto causal para individuos con un valor dado de U^D ,

$$\begin{aligned} MTE(x, u^D) &= E(\Delta / X = x, U^D = u^D) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 / U^D = u^D, X = x) \\ &= x(\beta^1 - \beta^0) + E(U^1 - U^0 / U^D = u^D) \end{aligned}$$

La evaluación del parámetro MTE a valores bajos de u^D ²³ entrega el impacto promedio para aquellos con valores de no observables que los hacen poco

²¹ Superíndice 1 representa el estado tratamiento y 0 representa el estado sin tratamiento.

²² $D(Z)$ denota la decisión observable de tratamiento, donde $D(Z) = 1$ denota el estado de tratamiento y $D(Z) = 0$ denota la situación sin tratamiento. La variable D^* es una variable latente que genera $D(Z)$ de acuerdo a la siguiente regla:

$D(Z) = 1[D^*(Z) \geq 0] = 1[Z\theta + U^D \geq 0]$ donde U^D es el determinante no observable de la decisión.

²³ Los valores de u^D son teóricos dado que por definición el término de error es no observado por el investigador. En la práctica la estimación marginal se realiza para sub grupos de la distribución ordenados según su probabilidad de participar (equivalente al propensity score). Se asume que el grupo marginal es aquel que está indiferente entre participar y no participar condicional a su u^D , es decir, en el margen se cumple que $u^D = Z\theta$, donde $Z\theta$ es el

probables de recibir el tratamiento, mientras que la evaluación del parámetro MTE a valores altos de u^D es la ganancia para aquellos individuos con no observables que les dan la mayor probabilidad de participar.

3.4. Uso de Instrumentos Cualitativos.

Las herramientas cualitativas en una evaluación de impacto se utilizan principalmente para entender e incorporar la visión de los beneficiarios en la evaluación, partiendo de la base de que un mismo servicio puede ser valorado de manera distinta dependiendo de la percepción que de este tengan los beneficiarios.

Las herramientas cualitativas son más apropiadas para entender y evaluar los procesos sociales que rodean la implementación de un programa (conflictos suscitados en torno al programa, razones por las cuales los beneficiarios no usan los servicios ofrecidos) o comportamientos organizacionales (cultura, clima organizacional).

Ejemplos de herramientas cualitativas son métodos tales como la evaluación participativa – técnica en la que se incorpora a los beneficiarios en el diseño, la ejecución y la evaluación de un programa -, los focus group y las entrevistas a informantes clave.

Las principales ventajas de las técnicas cualitativas son su flexibilidad, su rapidez en la recolección y procesamiento de la información y su capacidad para mejorar las conclusiones de una evaluación mediante el aporte de un conocimiento más preciso de las prioridades y las percepciones del usuario.

Sus principales desventajas son la subjetividad envuelta en la recolección de información y su carencia de representatividad estadística, la cual se traduce en una limitación para generalizar las conclusiones de este tipo de análisis a toda la población beneficiaria.

3.5. Integrando herramientas cuantitativas y cualitativas en una evaluación de impacto.

Del análisis de las características de las herramientas cualitativas se deduce que estas, por si solas, no permiten obtener conclusiones sobre los impactos del programa. No obstante, su integración con herramientas cuantitativas puede enriquecer de manera importante los resultados de una evaluación de impacto.

Una primera manera de integrar estas herramientas en una evaluación es en la etapa de diseño de los cuestionarios, puesto que un estudio cualitativo, por ejemplo un focus group, puede ayudar a precisar las preguntas relevantes.

propensity score del grupo marginal. Para una revisión más detallada de la metodología MTE ver Heckman, Tobías y Vytlačil 2001.

Una segunda forma de integración es en la etapa final del estudio, cuando hay resultados del análisis estadístico que necesitan ser interpretados o chequeados por el investigador. En este caso volver a terreno y llevar a cabo entrevistas más focalizadas a los beneficiarios puede aportar importante información interpretativa.

4. Medición de la eficiencia y economía de un programa.

Evaluar la eficiencia de un programa consiste en realizar un juicio técnicamente válido acerca de si los recursos que dicho programa tiene a su disposición están siendo maximizados: ¿Se está alcanzando el nivel actual de producción de los componentes al mínimo gasto posible o utilizando la menor cantidad de insumos posible (físicos o humanos)?, o alternatively, ¿Con el presupuesto vigente se está logrando el máximo nivel de producción de los componentes?

Por otra parte, se requiere analizar la economía en el manejo de los recursos del programa, es decir, la capacidad de la gerencia para administrar adecuadamente los recursos presupuestarios, para atraer recursos de otras fuentes y para abordar una eventual recuperación de costos del programa, si corresponde.

4.1 Cuantificación de los gastos y costos²⁴ relevantes²⁵.

El primer paso en esta dirección consiste en la identificación y cuantificación de todos los gastos y costos relevantes del programa.

a) Costos de Producción de Componentes

Corresponden a los gastos directamente asociados a la producción de los componentes (bienes y/o servicios) que entrega el programa a sus usuarios o beneficiarios (gasto becas, monto subsidios, precio de raciones alimenticias, remuneraciones del personal que labora directamente en la entrega de los servicios, etc.).

Costos Financieros

Corresponden a los gastos del programa que tienen una expresión monetaria. En este sentido, se solicita calcular el gasto devengado²⁶ más los desembolsos no contemplados en la Ley de Presupuestos devengados por otras instituciones públicas, privadas, o los mismos beneficiarios.

²⁴ Los términos “costo” y “gasto” se usan indistintamente.

²⁵ Para mayor información respecto de la cuantificación de costos del programa ver documento “Notas Técnicas. Evaluación de Programas. Capítulo VII. Conceptualización de Antecedentes Presupuestarios y de Gastos”, en sitio web de la Dirección de Presupuestos, www.dipres.cl, Sistema de Evaluación y Control de Gestión, publicaciones; y Anexo Información Financiera.

²⁶ Gasto devengado corresponde a todas las obligaciones en el momento que se generen, independientemente de que éstas hayan sido o no pagadas (Fuente: Normativa del Sistema de Contabilidad General de la Nación - Oficio C.G.R. N° 60.820, de 2005).

Se debe tener cuidado en computar todos los costos financieros que afectan el presupuesto público, puesto que en muchos casos parte de los gastos de un programa no aparecen en su asignación presupuestaria sino en otros concepto de gasto de la institución responsable de su administración²⁷. Ejemplo de esto son programas que son co-ejecutados por más de una institución pública, o en programas que son administrados por el nivel central pero devengados por instancias regionales o locales, las que por lo general absorben parte de los gastos de funcionamiento.

Lo mismo puede aplicarse respecto de aquellos gastos en que deben incurrir los beneficiarios para acceder a los bienes y servicios que genera el programa. Por ejemplo, un programa de capacitación de trabajadores, debiera considerar entre sus costos financieros, el gasto en movilización que realizan los beneficiarios para llegar al lugar donde se imparte la capacitación²⁸.

Costos no financieros

Los costos no financieros son gastos que no tienen una expresión monetaria, debido a que no representan desembolsos y por ende no afectan el flujo de caja del programa, tal es el caso de la depreciación de los vehículos, de las oficinas y del resto de las instalaciones, y el valor del tiempo que los trabajadores ocupan en trasladarse al lugar donde se imparte la capacitación.

Otro ejemplo en este tipo de costos, dice relación con el costo de oportunidad del patrimonio fiscal. Por lo general los análisis de eficiencia que se realizan en el sector público no toman en cuenta que los fondos fiscales tienen un costo de oportunidad financiero (la tasa de interés que podrían estar ganando si se mantuvieran en algún instrumento del mercado de capitales) o que los bienes inmuebles fiscales que se traspasan a particulares tienen un costo alternativo, puesto que podrían ser tasados y vendidos como cualquier otro inmueble.

b) Gastos de Administración del Programa²⁹

Los gastos de administración de un programa pueden definirse como todos aquellos desembolsos financieros que están relacionados con la generación de los servicios de apoyo a la producción de los componentes del programa.

Los gastos de administración de un programa pueden tener un componente fijo y otro variable. Ejemplo de los primeros son las inversiones necesarias para el funcionamiento del programa: vehículos, oficinas, computadores, mobiliario y, en general, todos aquellos gastos que, en el corto plazo, ocurrirán independientemente del nivel de producción de los componentes. El componente variable de los gastos de administración puede estar asociado a la ejecución de actividades de supervisión (viáticos, pasajes), difusión y

²⁷ Otro concepto de gasto de acuerdo con la clasificación presupuestaria.

²⁹ Para mayor información respecto de los costos administrativos del programa ver documento "Notas Técnicas. Evaluación de Programas. Capítulo VIII. Conceptualización de los Gastos Administración del Programa", en sitio web de la Dirección de Presupuestos, www.dipres.cl, Sistema de Evaluación y Control de Gestión, publicaciones; y Anexo Información Financiera.

promoción, selección de beneficiarios y, en general, todos aquellos gastos que aumentan cuando se incrementa el nivel de producción de los componentes o viceversa.

En general dentro de un programa es más fácil identificar y calcular los gastos directamente asociados a la generación de los componentes. Una manera práctica de aproximarse a estimar los gastos administrativos consiste en hacer una buena estimación del costo de producción de los componentes y considerar el gasto administrativo como el complemento del primero (el monto que falta para llegar a los costos totales).

Los gastos administrativos por lo general tienen una expresión financiera, es decir, se materializan en un desembolso de carácter monetario. Lo importante es tener en cuenta que no necesariamente todos los gastos administrativos de un programa están incluidos en su presupuesto. Es común, por ejemplo, que parte de los gastos de administración se realicen con cargo a clasificaciones presupuestarias más amplias y distintas a la del programa. También es frecuente que se traspase una parte importante de los gastos de administración a las instancias locales y/o regionales, según la modalidad de ejecución.

Ante situaciones como las descritas anteriormente, se vuelve necesario realizar estimaciones de los gastos administrativos en base a información entregada por la gerencia del programa. Lo anterior no siempre es fácil de realizar, en ocasiones, por ejemplo, el personal administrativo presta apoyo a más de un programa de la institución responsable. En tal caso, una estimación adecuada debe apuntar a determinar el porcentaje de la jornada que este personal destina al programa, y ponderar las remuneraciones respectivas para obtener una buena aproximación a los gastos en personal.

c) Externalidades.

La ejecución de algunos programas genera lo que en términos económicos se conoce como externalidades. Estas externalidades pueden ser positivas (beneficios) o negativas (costos), y afectan a personas o grupos que no están involucrados en el programa, es decir, al resto de la sociedad. Por ejemplo: un programa destinado a reducir la prevalencia de una enfermedad contagiosa, representa una externalidad positiva para toda la población, puesto que reduce el riesgo de contraer la enfermedad; mientras que un programa que fomenta el desarrollo forestal sobre la base de la plantación de las especies más rentables, puede implicar una externalidad negativa si es que se traduce en una reducción de la superficie de bosque nativo.

Una evaluación en profundidad debe identificar, y sí es del caso incorporar, las externalidades negativas más relevantes en el análisis de los costos del programa y las externalidades positivas por el lado de los beneficios.

4.2 Análisis de eficiencia³⁰.

³⁰ Para mayor información respecto de los costos administrativos del programa ver documento "Notas Técnicas. Evaluación de Programas. Capítulo IX. Profundizando en el Concepto de

Una vez computados todos los gastos relevantes que la información disponible permita, lo que sigue es relacionarlos con los niveles de producción de los componentes para obtener las medidas de eficiencia.

La eficiencia describe la relación entre dos magnitudes: la producción física de un producto y los insumos o recursos (humanos, físicos y monetarios) que se utilizaron para alcanzar ese nivel de producto. En otros términos, se refiere a si las actividades de un programa se ejecutaron, administraron y organizaron de tal manera que se haya incurrido en el menor gasto posible para generar los productos y/o componentes esperados (asignación y uso óptimo de recursos de manera de maximizar la entrega de productos y/o componentes).

Para medir eficiencia comúnmente se utilizan indicadores que relacionan insumos con productos/resultados. Las medidas de eficiencia se pueden agrupar en tres categorías:

- a) Las que dicen relación con la productividad física y la productividad media de los distintos tipos de factores (especialmente recursos humanos), y que relacionan el nivel de actividad, expresado como número de acciones, beneficios o prestaciones otorgadas, con el nivel de recursos utilizados, por ejemplo: consultas diarias por médico, índice de ocupación de camas hospitalarias, relación horas de supervisión por escuela, número de docentes por alumno, promedio de fiscalizaciones por fiscalizador, etc.
- b) Las que dicen relación con los gastos del programa, tales como: gastos medios por unidad de producción del bien o servicio que entrega el programa, gastos medios por beneficiario, gastos medios por componente del programa, etc.
- c) Las que dicen relación con la estructura de gastos del programa, ejemplo: el porcentaje de gastos administrativos en relación al gasto total del programa.

Idealmente, estas medidas deben ser comparadas con algún referente técnicamente válido para poder evaluar la eficiencia del programa. Una primera posibilidad es compararlas con los gastos de bienes, servicios, actividades o procesos similares a los estudiados, que sean desarrollados por otras organizaciones públicas o privadas. Esta técnica conocida como benchmarking implica conocer muy bien el proceso productivo del programa evaluado o de aquella parte que se quiere comparar, y contar con la información de gastos detallada, así como también contar con la información de gastos de aquellos identificados como líderes o con un desempeño aceptable desde el punto de vista de la eficiencia de sus procesos³¹.

Eficiencia del Programa”, en sitio web de la Dirección de Presupuestos, www.dipres.cl, Sistema de Evaluación y Control de Gestión, publicaciones.

³¹ Lo anterior debido a que no tiene sentido tomar como punto de referencia a una organización que sea ineficiente o muestre mal desempeño.

Otra alternativa es la de generar estándares de eficiencia internos para los procesos más importantes del programa. En este enfoque, se parte de la premisa de que existe una función de producción para el conjunto o para cada uno de los componentes y, por lo tanto, existe un set de alternativas técnicamente eficientes para generarlos. La idea es identificar las etapas que sigue la producción de cada componente, los insumos que se utilizan y cualquier otro aspecto que se considere relevante (identificar la función de producción), para luego medir y/o estimar el costo que tendrían éstos procesos si se ejecutaran de manera óptima. El resultado de este trabajo es costo estándar unitario para cada producto (bien o servicio), que puede ser comparado con los gastos devengados que surjan del análisis.

Considerando las restricciones en la disponibilidad de información, también es posible evaluar el programa respecto a sí mismo analizando el comportamiento histórico de las distintas mediciones de eficiencia desarrolladas e incorporando, además otros antecedentes que permitan realizar un juicio evaluativo acerca de la tendencia observada.

4.3 Análisis de economía.

Evaluar la economía de un programa implica, a partir de una recopilación y análisis de información financiera, estudiar de manera más precisa los aspectos relacionados con la formulación del presupuesto, la programación y ejecución del gasto y, en general, todo lo relativo al manejo adecuado de los recursos financieros del programa. Además, en este punto se deben analizar temas como la capacidad mostrada por la gerencia del programa para obtener recursos de otras fuentes de financiamiento – especialmente no gubernamentales – y la manera en que se ha abordado el tema de la recuperación de costos del programa, si corresponde a la naturaleza del programa.

5. Mediciones globales de desempeño³².

Aunque se reconoce que son difíciles de construir debido a la gran cantidad de información y análisis que necesitan, existen algunos instrumentos que relacionan los resultados finales o beneficios de un programa con sus gastos y costos relevantes y, por lo tanto, constituyen una medida de desempeño global de un programa.

La más conocida de estas mediciones de costo-beneficio es el VAN social, que representa el valor actual del flujo de beneficios netos de un programa (beneficios valorados menos costos valorados), la principal característica del VAN es que considera en su construcción un concepto amplio de costos (y también de beneficios), que es el concepto de costo económico o social, el cual envuelve la totalidad de los costos asociados a la ejecución de un proyecto:

³² Las mediciones globales de desempeño señaladas en este punto, se utilizan por lo general en la evaluación ex ante de programas y/o alternativas de intervención. No obstante, también pueden cumplir un rol muy importante en una evaluación ex post, como es el caso de la evaluación en profundidad, puesto que permiten verificar empíricamente si se dieron los flujos de beneficios y costos supuestos en el análisis inicial.

internos y externos, monetarios y no monetarios, valoración económica de las externalidades, etc. Existen varias herramientas similares o derivadas del VAN, tales como la TIR social o Tasa Interna de Retorno Social y la TIR marginal.

Debido a la dificultad práctica que representan las mediciones de costo beneficio, se pueden construir otros indicadores que relacionen los beneficios expresados en sus propias unidades de medida (jóvenes rehabilitados, puntaje promedio del SIMCE, familias que superaron la línea de pobreza, etc.) con los gastos asociados a generar esos resultados, e indicadores de costo-utilidad que relacionen los gastos con la valoración que los individuos afectados por la intervención tienen de los beneficios recibidos³³.

³³ En este caso, por lo general, la valoración que tienen los usuarios respecto de los beneficios recibidos se mide a través de su disposición a pagar. ¿Cuánto estaría dispuesto a pagar Ud. por el beneficio X?

BIBLIOGRAFÍA.

Angrist, J. D. y Krueger, A. B. (1999). "Empirical Strategies in Labor economics". *Handbook of Labor Economics*. Vol. 3. Elsevier Science.

Angrist, J. D., Imbens, G. W., y Rubin, D. B. (1996). "Identification of Causal Effects using Instrumental variables". *Journal of the American Statistical Association*. Vol. 91. pág. 444-472.

Angrist, J. D. y Pischke, J. K. (1999). "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement". *The Quarterly Journal of Economics*, Vol. 114, No. 2. (May, 1999), pp. 533-575.

Baker Judy L. (2000). *Evaluating the Impact of Development Projects on Poverty. A Handbook for Practitioners*. The World Bank. Washington D.C. 2000

Banco Interamericano de Desarrollo. (1997). "Evaluación un herramienta de gestión para mejorar el desempeño de los proyectos". Oficina de Evaluación (EVO), BID, Washington D.C. 1997.

Banco Mundial (1998). "Manual para la gestión del gasto público". The World Bank. Washington D.C. 1998.

Bitler, M., Gelbach, J. y Hoynes, H. (2003). "What Mean Impacts Miss: Distributional Reform Experiments". NBER Working Paper Series. Working paper 10121.

Bravo, D., Contreras, D. y Crespi, G. (2000). "Evaluación de Impacto en Formación Empresarial. El Caso de FUNDES". FUNDES - Mc Graw Hill Interamericana.

Card, D. y Krueger, A. B. (1994). "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania". *American Economic Review*. Vol. 84. pág. 772-793.

CEPAL (1995). "Focalización y Pobreza". *Cuadernos de la CEPAL N° 71*. Naciones Unidas.

Gruber, J (1994) "The Incidence of Mandated Maternity Benefits", *The American Economic Review*. 1994.

Guzmán M. (2001). *Evaluación de Programas e Indicadores de Desempeño. Transparencia y Mejoramiento de los Procedimientos para la Elaboración y Discusión Presupuestaria*. Dirección de Presupuesto. Noviembre de 2001. www.dipres.cl.

Heckman, J., Ichimura, H. y Todd, P. (1997). "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme". *Review of Economic Studies*. Vol. 64, pág. 605-654.

Heckman, J., Tobias, J., Vytlačil, E. (2001). "Four Parameters of Interest in the Evaluation of Social Programs". *Southern Economic Journal*, Vol 68, No. 2 (Oct., 2001), 210-223.

Instituto Internacional de Gobernabilidad (2001) *Una conceptualización del capital social desde la economía compromisos y efectos colaterales*. Colección de Reseñas. Biblioteca de Ideas.

www.gov.igov/pnud/bibliote/resenas/tema5/resena0121.htm.

LaLonde, R. (1986). "Evaluating the Econometric Evaluation of training Programs with Experimental Data". *American Economic Review*. Vol. 66, pág. 604-629.

Mackay, K (ed) (1998). "Public Sector Performance – The Critical Role of Evaluation". Operations Evaluation Department. The World Bank. Washington D.C. 1998.

Marcel, M. (1999). "Indicadores de Desempeño como Instrumento de Modernización en el Estado de Chile". *Perspectivas 1999*.

Mayne, J. y Zapico-Goñi, E. (2000). "Seguimiento de los Resultados de la Gestión en el Sector Público. Perspectivas desde la Experiencia Internacional". Instituto de Estudios Fiscales, Ministerio de Hacienda. Madrid, España.

Mokate, K.M. (2000). Convirtiendo el "monstruo" en aliado: la evaluación como herramienta de la gerencia social. Instituto Interamericano para el Desarrollo SOCAL – NDES. Abril del 2000.

Mokate, K. M. (1999). "Eficacia, eficiencia, equidad y sostenibilidad: ¿qué queremos decir?. Instituto Interamericano para el Desarrollo SOCAL – NDES. 1999.

Van der Klaauw, W. (2002) "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression Discontinuity Approach. *International Economic Review*, Vol. 43, No.4. November 2002