

Sinopsis: Estrategias de Atribución Causal

Patricia Rogers

CENTRO DE INVESTIGACIONES INNOCENTI DE UNICEF

El Centro de Investigaciones Innocenti es la oficina de UNICEF especializada en investigación científica. El objetivo principal del Centro de Investigaciones es mejorar la comprensión internacional de una serie de cuestiones relacionadas con los derechos de la infancia, a fin de facilitar la plena aplicación de la Convención sobre los Derechos del Niño mediante su promoción en todo el mundo. El centro tiene el cometido de establecer un marco integral de investigación y conocimiento dentro de la organización para brindar apoyo a sus programas y políticas mundiales. Las publicaciones producidas por el centro contribuyen al debate global sobre la infancia y los derechos del niño e incluyen una amplia gama de opiniones.

Las opiniones expresadas corresponden a los autores o editores y se publican para estimular un mayor diálogo sobre métodos de análisis de impacto. Esta publicación no refleja necesariamente las políticas o perspectivas de UNICEF sobre ciertos temas.

SINTESIS METODOLOGICAS DEL CENTRO DE INVESTIGACIONES

Las síntesis metodológicas del Centro de Investigaciones de UNICEF pretenden compartir prácticas de investigación, métodos, diseños y recomendaciones de reconocidos investigadores y analistas. La audiencia a la que van dirigidas es principalmente personal de UNICEF que lleve a cabo, encargue o interprete los resultados de investigación y análisis para la toma de decisiones sobre programas y políticas de sensibilización.

Esta síntesis metodológica ha seguido un proceso de revisión por pares interna.

El texto no ha sido editado de acuerdo con los estándares de publicación oficiales y UNICEF declina toda responsabilidad por posibles errores

Se permite la reproducción de cualquier parte de la presente publicación siempre que se incluya referencia a la presente. Si se desea utilizar una parte sustancial de la publicación dirijan su solicitud al Departamento de Comunicación en la dirección de correo electrónico: florence@unicef.org

Para consultas o descargas, pueden encontrar estas síntesis metodológicas en <http://www.unicef-irc.org/KM/IE/>

Recomendamos la siguiente cita para cualquier referencia al presente documento:

Rogers, P. (2014). Sinopsis: Estrategias de atribución causal, *Síntesis metodológica n.º6*, Centro de Investigaciones de UNICEF, Florencia.

Agradecimientos: Varios autores han proporcionado orientación en la preparación de esta síntesis. El autor y el Centro de Investigaciones de UNICEF desean agradecer a todos aquellos que han participado en la preparación de la presente publicación, especialmente a:

Por su contribución: Simon Hearn, Greet Peersman, Jessica Sinclair Taylor, Howard White

Por su revisión: Nikola Balvin, Sudhanshu Handa, Debra Jackson, David Parker

© Fondo de las Naciones Unidas para la Infancia (UNICEF), septiembre de 2014
Centro de Investigaciones Innocenti de UNICEF

Piazza SS. Annunziata, 12
50122 Florencia (Italia)
Tel: (+39) 055 20 330
Fax: (+39) 055 2033 220
florence@unicef.org
www.unicef-irc.org

1. ATRIBUCIÓN CAUSAL: BREVE DESCRIPCIÓN

Uno de los aspectos esenciales de una [evaluación de impacto](#) es que no solo mide o describe cambios que han ocurrido, sino que también procura entender la función de determinadas intervenciones (es decir, programas o políticas) en la generación de estos cambios. Este proceso se suele conocer como atribución causal, contribución causal o inferencia causal. Esta síntesis expone una visión general de las distintas formas de examinar la atribución causal, utilizando una combinación de diseño de investigación y determinadas estrategias de recolección y análisis de datos.

La definición de «impacto» del Comité de Asistencia para el Desarrollo (CAD) de la Organización de Cooperación y Desarrollo Económicos (OCDE) deja claro que la evaluación de impacto debe establecer la causa de los cambios observados: «Efectos de largo plazo positivos y negativos, primarios y secundarios, producidos directa o indirectamente por una intervención para el desarrollo, intencionalmente o no»¹.

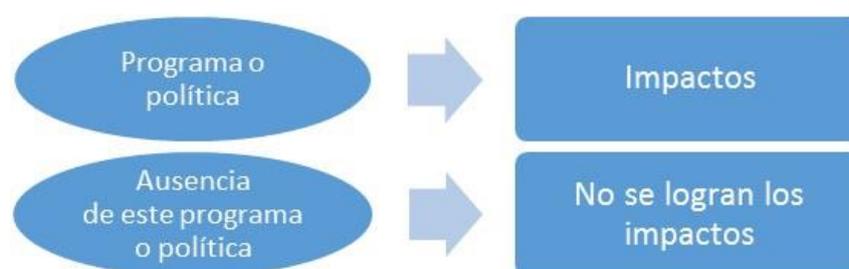
El CAD define la atribución causal como: «Imputación de un vínculo causal entre cambios observados (o que se espera observar) y una intervención específica»².

Esta definición no exige que los cambios sean producidos exclusiva o totalmente por el programa o la política que se investiga³. En otras palabras, tiene en cuenta que pueden haber intervenido otras causas, por ejemplo otros programas/políticas en la esfera de interés o determinados factores contextuales (a menudo denominados «factores externos»).

Las evaluaciones generan constataciones más sólidas y útiles si no investigan únicamente los vínculos entre las actividades y los impactos, sino también los vínculos a lo largo de toda la cadena causal entre [actividades](#), [productos](#), [resultados](#) intermedios e [impactos](#). Por lo tanto, es conveniente definir una «[teoría del cambio](#)»⁴ (véase la Síntesis n.º 2 (Teoría del cambio)) para orientar la atribución causal en una evaluación de impacto. La evaluación puede confirmar la teoría del cambio o sugerir mejoras basadas en el análisis de las pruebas empíricas.

Puede resultar útil tener presentes las siguientes tres conceptualizaciones de causa-efecto a la hora de planificar una evaluación de impacto.

1. **Atribución causal exclusiva:** Cuando un programa o política es necesario y suficiente para producir los impactos, con independencia (o relativa independencia) de los factores contextuales u otras intervenciones.



¹ Organización de Cooperación y Desarrollo Económicos – Comité de Asistencia para el Desarrollo, *Glosario de los principales términos sobre evaluación y gestión basada en resultados*, OCDE, París, 2010. Véase <http://www.oecd.org/development/peer-reviews/2754804.pdf>.

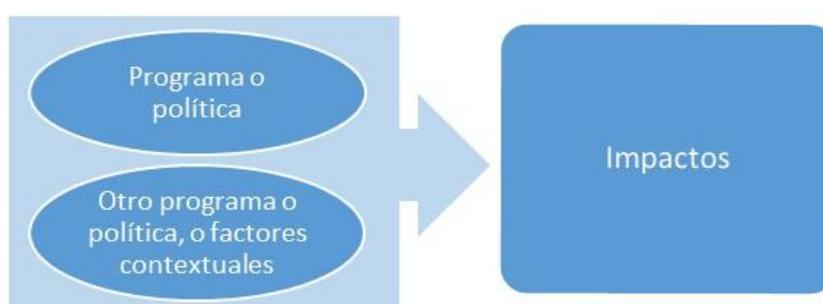
² *Ibíd.*

³ Grupo de Evaluación de las Naciones Unidas, *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, documento orientativo, UNEG, Nueva York, 2013. Véase http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=1434.

⁴ La teoría del cambio explica cómo se entiende que las actividades produzcan una serie de resultados que contribuyen a lograr los impactos finales previstos.

En la práctica, los programas o las políticas rara vez se bastan por sí solos para producir los impactos previstos; a menudo existen formas alternativas de lograrlos. Como consecuencia, este no suele ser un modelo útil de causa-efecto para la evaluación de impacto.

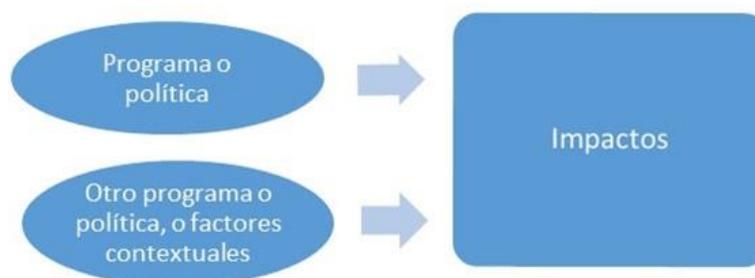
2. **Atribución causal conjunta:** Cuando el programa o la política produce los impactos junto con otros programas o políticas o determinados factores contextuales. Pueden ser programas complementarios que sientan las bases para el programa que se está evaluando o lo refuerzan. También pueden ser factores relacionados con el entorno de ejecución, como las cualificaciones e infraestructuras de los asociados en la ejecución, o con los participantes, por ejemplo en cuanto a niveles de motivación y conocimientos previos. Cuando estos factores coadyuvantes están ausentes o son negativos, los impactos no se lograrán o se lograrán en menor medida.



Se trata de una situación muy común, en la que el programa solo funcionará eficazmente si se dan condiciones favorables o si se eliminan las condiciones desfavorables. Tiene importantes implicaciones para la forma de realizar la evaluación de impacto y de utilizar las constataciones. Por ejemplo, si se constata que un programa solo funciona en un lugar donde hay transparencia respecto a las asignaciones de fondos públicos, solo debe ejecutarse en situaciones donde ya exista esta transparencia o en combinación con actividades para lograrla. Para identificar estos factores coadyuvantes, es importante desarrollar una buena teoría del cambio que los incluya —sobre la base de investigaciones anteriores y conocimientos existentes— y recopilar datos sobre ellos.

3. **Vías causales alternativas (o múltiples):** Cuando el programa o la política es solo una de las diversas formas posibles de lograr los impactos. Un determinado programa o política puede generar los impactos, pero estos también pueden producirse a través de otras intervenciones o factores externos.

Puede ocurrir cuando los participantes tienen la posibilidad de acceder a los servicios a través de un proveedor alternativo o cuando distintos programas de distintos proveedores tienen por objeto producir el mismo impacto. Por ejemplo, varios programas podrían tener como objetivo reducir la mortalidad infantil, algunos mejorando la nutrición, otros reduciendo las enfermedades transmitidas por el agua y otros mediante la inmunización.



Este supuesto tiene importantes implicaciones para la evaluación de impacto. Si se utiliza un [diseño contrafáctico](#), en el que se compara a los participantes con no participantes, es importante investigar a qué servicios acceden los no participantes. Si el objetivo de un programa es producir impactos que también son el objetivo de otros programas, es especialmente importante definir resultados intermedios en la teoría del cambio y recopilar datos sobre ellos. De esta forma se podrá determinar en qué medida los impactos se han debido al programa que se está evaluando.

Existen tres estrategias generales de atribución causal en las evaluaciones de impacto:

- estimar lo contrafáctico (es decir, qué habría pasado en ausencia de la intervención en comparación con la situación observada);
- comprobar la coherencia de las pruebas empíricas respecto a las [relaciones causales](#) explicitadas en la teoría del cambio;
- descartar explicaciones alternativas mediante [procesos con base empírica](#) lógicos.

El uso de una combinación de estas estrategias normalmente puede ayudar a reforzar las conclusiones obtenidas (para más información, véase la sección 3).

Es importante que todas las evaluaciones de impacto lleven a cabo sistemáticamente una atribución causal. Sin embargo, también debe reconocerse lo que se puede lograr de manera realista con una única evaluación, especialmente cuando hay un tiempo limitado para la recolección y el análisis de datos iterativos. Una evaluación única de impacto puede añadirse a la base empírica, pero, aun cuando exista un grado razonable de seguridad acerca de que el programa o la política causaron los impactos observados, esto no significa que la evaluación haya examinado todos los aspectos de la intervención o cómo funcionará en otros entornos y en otros momentos. Así, la elaboración de una síntesis de las constataciones de todas las evaluaciones de impacto proporciona pruebas más sólidas que una única evaluación sobre la [extrapolabilidad](#) de los efectos observados.

Puntos principales

1. La atribución causal investiga los vínculos causales entre un programa u otra intervención y los cambios observados.
2. La atribución causal es un elemento esencial de la evaluación de impacto.
3. Existen varias estrategias para examinar la atribución causal, todas las cuales mejoran si se basan en una teoría del cambio sólida.
4. La estrategia más adecuada para la atribución causal depende del contexto de evaluación, así como de lo que se está evaluando.

2. ¿CUÁNDO SE APLICA LA ATRIBUCIÓN CAUSAL EN LA EVALUACIÓN DE IMPACTO?

La atribución causal es un elemento esencial de la evaluación de impacto. Permite que una evaluación informe no solo de que se ha producido un cambio, sino también que se debió, al menos parcialmente, al programa o la política objeto de evaluación.

No puede afirmarse que un cambio sea un impacto salvo que exista un vínculo demostrable entre él y la intervención, lo cual se apoya en conexiones demostrables entre los insumos y productos de la intervención, los resultados e impactos que se observan (véase la Síntesis n.º 1 (Sinopsis de la evaluación de impacto)).

Si una evaluación de impacto no lleva a cabo sistemáticamente una atribución causal, se corre un mayor riesgo de que la evaluación produzca constataciones incorrectas y lleve a decisiones incorrectas, por ejemplo, decidir ampliar un programa cuando en realidad es ineficaz o únicamente eficaz en determinadas situaciones, o decidir interrumpirlo cuando el programa podría funcionar si se corrigiesen los factores restrictivos.

3. CÓMO LLEVAR A CABO UNA ATRIBUCIÓN CAUSAL

En todos los casos es conveniente empezar por definir o revisar la vía de cambio mediante la que se entiende que la *intervención* (programa o política) contribuye a los impactos previstos u observados.

La vía de cambio puede describirse en una cadena causal —o modelo lógico de programa— y puede resultar útil formularla utilizando las categorías estándar de [insumos](#), productos, resultados e impacto de la gestión basada en resultados. La evaluación puede dilucidar la vía de cambio cuando no se ha formulado explícitamente una, o puede validar (o refutar) una vía hipotética, lo que suele denominarse una teoría del cambio (véase la Síntesis n.º 2 (Teoría del cambio)).

Pueden utilizarse distintas estrategias para llevar a cabo la atribución causal, todas las cuales presentan sus propios puntos fuertes, limitaciones y grados de idoneidad en función del programa y contexto de evaluación específico. Aunque existen diversas formas de clasificar los diseños y métodos de atribución causal en la evaluación de impacto⁵, con el fin de dar una visión general en esta síntesis, solo se analizarán tres enfoques generales:

- **Enfoques contrafácticos:** se elabora una estimación de lo que habría ocurrido en ausencia del programa o la política y se compara con lo que se ha observado en presencia de la intervención. Este enfoque requiere el uso de un [grupo de control](#) o un [grupo de comparación](#).
- **Coherencia de las pruebas empíricas con la relación causal:** se definen patrones que serían coherentes con una relación causal, que normalmente se basa en una teoría del cambio bien desarrollada, y a continuación se intenta confirmar o desmentir las pruebas.
- **Descarte de explicaciones alternativas:** se definen posibles explicaciones causales alternativas y se busca información para ver si se pueden descartar.

Estos enfoques se examinan con más detalle en esta síntesis y otras síntesis de la serie (véanse la Síntesis n.º 7 (Ensayos controlados aleatorios); Síntesis n.º 8 (Diseño y métodos cuasi experimentales) y Síntesis n.º 9 (Estudios de caso comparativos)).

⁵ Véase, por ejemplo, el marco definido por Elliott Stern *et al.*, «Broadening the range of designs and methods for impact evaluations», *DFID Working Paper* No. 38, Departamento de Desarrollo Internacional, Londres, 2012. Este documento distingue entre estudios experimentales, estadísticos, basados en teorías, basados en casos, participativos y de síntesis.

Enfoques contrafácticos

Un [enfoque contrafáctico](#) consiste en elaborar una estimación de lo que habría ocurrido en ausencia de un programa o una política y compararlo con lo que se ha observado en presencia de la intervención. Hay cuatro tipos de diseño de evaluación que aplican un enfoque contrafáctico, que varían en función de cómo se calcula el efecto «contrafáctico» (no intervención).

Diseños experimentales

Consisten en la asignación aleatoria de participantes para poner a prueba los efectos de una intervención. Los participantes se distribuyen de forma aleatoria entre uno o más grupos que participan en la intervención y un grupo de control que no recibe la intervención. Estas evaluaciones se llaman [ensayos controlados aleatorios](#) (véase también la Síntesis n.º 7 (Ensayos controlados aleatorios)) por la rigurosa forma de aplicar la intervención, similar a la forma de someter a prueba los productos farmacéuticos.

Los ensayos controlados aleatorios reducen el riesgo de [sesgo de selección](#), cuando los participantes y los no participantes presentan diferencias sistemáticas además de que reciban o no la intervención, por lo que el aparente impacto del programa o la política puede deberse a estas diferencias en vez (o además) de a su efecto real.

Diseños cuasi experimentales⁶

Consisten en crear grupos de comparación de diversas maneras sin incluir la [asignación aleatoria](#) (véase también la Síntesis n.º 8 (Diseño y métodos cuasi experimentales)). Estos diseños suelen ser más viables en una evaluación de impacto y puede considerarse que proporcionan una comparación suficientemente válida entre quienes reciben y no reciben la intervención. Entre las opciones para crear grupos de comparación está el emparejamiento crítico⁷ y las comparaciones emparejadas⁸, así como otras opciones más comúnmente utilizadas:

- **Emparejamiento crítico:** crear un grupo de comparación encontrando una pareja para cada persona o lugar del [grupo de tratamiento](#) según el criterio del investigador respecto a las [variables](#) que son importantes.
- **Comparaciones emparejadas:** emparejar a cada participante (persona, organización o comunidad) con un no participante en función de variables que se consideran relevantes.
- **Emparejamiento por puntuación de la propensión:** crear estadísticamente grupos comparables sobre la base de un análisis de los factores que influyeron en la propensión de las personas a participar en la intervención.
- **Diseño de regresión discontinua:** comparar los resultados de los individuos que están por debajo y por encima de un valor de corte.
- **Asignación secuencial:** se crean un grupo de tratamiento y un grupo de comparación mediante asignación secuencial (por ejemplo, una de cada tres personas en la lista).

⁶ Algunos materiales publicados se refieren a estos diseños como «no experimentales» en lugar de «cuasi experimentales». Los autores de la síntesis prefieren utilizar el término «cuasi experimental», ya que los diseños se asemejan de muchas formas a experimentos y los enfoques no experimentales de la atribución causal son bastante diferentes y, por lo tanto, normalmente se clasifican por separado.

⁷ Crear un grupo de comparación encontrando una pareja para cada persona o lugar del grupo de tratamiento sobre la base del juicio del investigador respecto a las variables que son importantes.

⁸ Emparejar a cada participante (persona, organización o comunidad) con un no participante en función de variables que se consideran relevantes.

Algunas opciones de análisis en los diseños cuasi experimentales son:

- **Diferencia única:** comparar los resultados en el grupo de tratamiento con los resultados en el grupo de comparación una única vez tras la intervención.
- **Diferencia doble (también conocida como «diferencia en diferencias» o «DID»):** comparar la diferencia entre «el antes y el después» del grupo que recibe la intervención (no asignados aleatoriamente) con la diferencia del antes y el después del grupo que no la recibe.

Contrafáctico hipotético

A veces es posible construir un «contrafáctico hipotético» de lo que habría ocurrido en ausencia del programa o la política demostrando que las condiciones habrían seguido siendo las mismas.

Por ejemplo, en una evaluación de impacto de un proyecto hídrico que consistía en estudiar su impacto en el tiempo empleado en transportar agua, fue suficiente demostrar que el tiempo empleado se había reducido después de que se instalase una bomba central, puesto que era razonable suponer que en ausencia de la bomba este tiempo no habría cambiado. En términos prácticos, esto representa una comparación del antes y el después (que puede ser entre la misma [población](#)), en vez de una comparación de distintos grupos en un momento dado.

Elaboración de modelos

Un contrafáctico creado estadísticamente supone desarrollar un modelo estadístico como un análisis de [regresión](#) para calcular qué habría pasado en ausencia de una intervención (véase la Síntesis n.º 13 (Elaboración de modelos)).

Coherencia de las pruebas empíricas con la relación causal

Este enfoque de atribución causal consiste en identificar qué pruebas empíricas serían coherentes con una relación causal y posteriormente recabar y analizar datos de distintos grupos para determinar si las pruebas coinciden con ello. Este enfoque se guía normalmente por una teoría del cambio, que o bien se haya explicado de forma pormenorizada o bien se encuentre implícita en el modelo lógico del programa o la política.

Pueden utilizarse distintos métodos de recolección y análisis de datos para reunir estas pruebas. Es recomendable combinar varios métodos en una única evaluación de impacto en función del nivel de certeza requerida y de las posibles contraexplicaciones definidas. Además, pueden utilizarse las pruebas empíricas de anteriores investigaciones y evaluaciones; así por ejemplo, las evaluaciones de impacto de los programas de vacunación no necesitan comprobar cada vínculo en la cadena causal habida cuenta del conjunto de conocimientos disponibles de anteriores investigaciones.

Algunas opciones son:

- **Logro de resultados intermedios:** comprobar si en todos los casos que se lograron los impactos finales se lograron también los resultados intermedios definidos en la teoría del cambio.
- **Contrastar los resultados con predicciones de expertos:** formular predicciones basadas en la teoría del cambio o una nueva teoría de factores más amplios que contribuyen a los resultados y realizar un seguimiento posterior para ver si estas predicciones se materializan verdaderamente o no a lo largo del tiempo.
- **Comprobar el momento en el que se producen los impactos:** determinar si el momento en el que se producen los impactos es coherente con una relación causal, una vez más con

referencia a la teoría del cambio. Por ejemplo, el impacto ocurre un tiempo razonable después de que haya surtido efecto el programa o política.

- **Estudios de caso comparativos:** comparar sistemáticamente [estudios de caso](#) para entender la serie de factores que pueden ser responsables de los impactos.
- **Patrones dosis-respuesta:** examinar el vínculo entre la «dosis» (la intensidad o el nivel al que se ha aplicado la intervención) y la «respuesta» (el efecto observado) en el marco de las actividades para determinar si la intervención ha producido o no el resultado. Este es otro método extraído del mundo de los ensayos biológicos que se aplica cada vez más a la investigación y evaluación en ciencias sociales.
- **Comprobar la coherencia con la literatura existente:** contrastar los resultados con lo que se conoce mediante la revisión de la literatura en el área específica de interés con el fin de detectar congruencias e incongruencias. Debe hacerse con cautela y mencionando explícitamente las limitaciones de la literatura existente (o subconjunto de la misma).
- **Entrevistas a informantes clave:** no se trata de preguntar a los informantes clave si creen que la intervención ha producido los impactos (lo que puede verse afectado por su nivel de conocimiento de los procesos causales y su intención respecto a la continuación de la intervención). Se trata de pedirles que expliquen los procesos causales tras su participación. Por ejemplo, en el caso de un programa cuyo objetivo sea prestar apoyo al gobierno nacional para la elaboración de nuevas políticas con base empírica, en la entrevista a informantes clave se podría preguntar por el proceso de elaboración de las políticas y después retrotraerse al programa, evitando preguntas dirigidas que presupongan que el programa ha sido la principal causa. Estas entrevistas pueden proporcionar pruebas que expliquen, por ejemplo, cómo un programa de capacitación ha marcado la diferencia en términos de la posterior capacidad del gobierno para elaborar y negociar determinadas políticas.
- **Modus operandi:** tomar como base la experiencia previa de los participantes y las partes interesadas para determinar qué serie o patrón de efectos es típico de una intervención. Por ejemplo, un programa de capacitación podría tener una terminología o práctica «distintiva» que es evidente entre los participantes.
- **Seguimiento de procesos:** desarrollar hipótesis alternativas y posteriormente recabar pruebas (pistas) para determinar si son compatibles o no con las hipótesis. (El enfoque detectivesco de Sherlock Holmes ejemplifica esta lógica.)
- **Análisis comparativo cualitativo:** relacionado con los estudios de caso comparativos, consiste en comparar las configuraciones de los distintos estudios de caso para identificar los componentes que parecen ser los principales responsables de producir resultados específicos.
- **Análisis realista de hipótesis comprobables:** utilizar una teoría del cambio realista (lo que funciona, con quién, en qué circunstancias y a través de qué mecanismos causales) para definir contextos específicos en los que cabría esperar y no cabría esperar resultados positivos, y contrastarlos con las situaciones observadas.

Para más información, véanse la Síntesis n.º 9 (Estudios de caso comparativos) y la Síntesis n.º 13 (Elaboración de modelos). Para más información sobre los demás métodos descritos anteriormente, véase «[Check the results support causal attribution](#)» en el sitio web de BetterEvaluation.

Descarte de explicaciones alternativas

La tercera estrategia de atribución causal consiste en definir posibles explicaciones alternativas del logro de los impactos y a continuación recopilar datos para ver si pueden descartarse. Esta estrategia

resulta especialmente útil cuando las pruebas disponibles pueden bastar solo para indicar «[correlación](#)», pero no «[causalidad](#)».

Algunas opciones son:

- **Entrevistas a informantes clave:** pedir a expertos en el tipo específico de programa, miembros de la comunidad u otras partes interesadas que definan otras posibles explicaciones y, si resulta viable, evaluar si estas pueden descartarse.
- **Seguimiento de procesos:** utilizar las pruebas empíricas para descartar variables explicativas alternativas en cada paso de la teoría del cambio.
- **Descarte de explicaciones técnicas:** definir e investigar posibles formas de que los resultados observados reflejen limitaciones técnicas de los datos (por ejemplo, regresión al [promedio](#) o medición no fiable) en vez de relaciones causales.
- **Elaboración de modelos:** investigar explicaciones alternativas por medio de análisis estadísticos como la regresión o la regresión logística para controlar factores de desviación.
- **Metodología de eliminación general:** se lleva a cabo en dos fases: 1) definir posibles explicaciones (incluso que los cambios observados se deben efectivamente a la intervención, más todas las explicaciones alternativas posibles) utilizando una *combinación* de opciones como las enumeradas anteriormente (por ejemplo, entrevistas a informantes clave e [intercambio de ideas](#), y revisión de evaluaciones/investigaciones anteriores); y 2) recopilar y analizar datos para ver si pueden descartarse las explicaciones alternativas. Esta opción de atribución causal es más eficaz cuando se combina con otras estrategias.

Puesto que el uso de la atribución causal sistemática no experimental en la elaboración de evaluaciones de impacto es poco común, este enfoque se ilustra mejor con un ejemplo de otro ámbito de evaluación. La evaluación de impacto de la nueva ley sobre el uso obligatorio del casco entre los ciclistas⁹ determinó que el número de lesiones craneoencefálicas de ciclistas había disminuido tras la implantación de la ley. Sin embargo, esto no constituía por sí mismo una prueba de relación causal. La evaluación no pudo basarse en un enfoque contrafáctico, puesto que no había un grupo de comparación adecuado. Solo pudo basarse parcialmente en la coherencia de las pruebas con la relación causal, es decir, confirmar que se habían comprado más cascos desde que entró en vigor la ley y que las observaciones demostraban un mayor nivel de cumplimiento de la nueva legislación.

La principal estrategia de atribución causal en este caso consistió en definir e investigar posibles explicaciones alternativas, con el objetivo de descartarlas. Por ejemplo, la disminución de las lesiones craneoencefálicas entre los ciclistas podría haberse logrado mediante la reducción de la práctica de ciclismo, al decidir la población que no utilizaría la bici si tenía que llevar casco. No fue factible realizar la gran encuesta necesaria para recopilar nuevos datos con el fin de investigar esta explicación alternativa. Sin embargo, existían datos que eran pertinentes y suficientemente precisos. Si la disminución de lesiones craneoencefálicas se hubiese debido a la reducción de la práctica de ciclismo, también cabría esperar que hubiese disminuido el número total de lesiones entre los ciclistas. En cambio, los datos mostraron que solo había disminuido el número de lesiones craneoencefálicas, reforzando el argumento de que la reducción no se había debido a la reducción de la práctica del ciclismo.

⁹ Walter, Scott R., *et al.*, «The impact of compulsory cycle helmet legislation on cyclist helmet head injuries in New South Wales, Australia», *Accident Analysis and Prevention*, 43, 2011, págs. 2064–2071.

4. CÓMO ELEGIR LA MEJOR ESTRATEGIA DE ATRIBUCIÓN CAUSAL

Existen opiniones divergentes sobre los méritos relativos de las distintas estrategias y opciones de atribución causal, que repercuten en su elección cuando se llevan a cabo evaluaciones conjuntas con otros organismos. Algunas organizaciones clasifican estos enfoques por orden de mérito, exigiendo un enfoque contrafáctico y etiquetando los diseños experimentales con un grupo de control como la «regla de oro».

Por ejemplo, la política de evaluación de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID) exige claramente la creación de un grupo contrafáctico, y se prefiere el uso de un grupo de control para que una evaluación se considere evaluación de impacto:

«Las evaluaciones de impacto miden el cambio en un resultado de desarrollo que es atribuible a una intervención definida; las evaluaciones de impacto se basan en modelos de causa-efecto y requieren un contrafáctico fiable y rigurosamente definido para controlar los factores»¹⁰.

Otras organizaciones pueden estar de acuerdo en los méritos técnicos de distintas estrategias, pero dejan que la mejor opción responda, según la situación, a la naturaleza de lo que se está evaluando y las circunstancias de la evaluación.

Por ejemplo, el documento de debate sobre evaluación de impacto de AusAID (ahora Departamento de Asuntos Exteriores y Comercio de Australia) propone utilizar una combinación de métodos y diseños que se adecúen a la situación:

«AusAID reconoce que **para la evaluación de impacto puede utilizarse una serie de enfoques metodológicos**. La necesidad de evaluar el impacto empieza por la necesidad de pruebas empíricas (como se ha detallado anteriormente). La finalidad y las preguntas de la evaluación, la complejidad de la intervención y el contexto de la intervención determinarán los tipos de métodos a utilizar. Las evaluaciones de impacto en AusAID no deben estar dirigidas por métodos»¹¹.

El reciente documento orientativo para la evaluación de impacto del Grupo de Evaluación de las Naciones Unidas (UNEG) aboga por escoger una mezcla de métodos y diseños adecuados a las circunstancias de la evaluación de impacto y los tipos de preguntas que busca responder:

«Al realizar una evaluación de impacto, más que en otros tipos de evaluación, es importante no limitarse solo a enumerar un conjunto de métodos y asumir que esto equivale a una metodología. Más bien, dentro del diseño de evaluación escogido debe existir un marco metodológico general explícito, que permita unir los métodos individuales para producir un análisis general significativo que pueda evaluar si una intervención ha generado impactos o ha contribuido a ellos. Es esencial adaptar el diseño de evaluación y la combinación de métodos a la situación, el contexto, las preguntas y cuestiones que suscitan preocupación»¹².

¹⁰ Agencia de los Estados Unidos para el Desarrollo Internacional, Evaluación, Aprendizaje de la experiencia, Política de evaluación, USAID, Washington, D.C., 2011. Véase <http://www.usaid.gov/sites/default/files/documents/1868/USAIDEvaluationPolicy.pdf>.

¹¹ Oficina de Eficacia del Desarrollo de AusAID, «Impact Evaluation: A Discussion Paper for AusAID Practitioners», Documento de debate, AusAID ODE, 2012. Véase <http://www.ode.dfat.gov.au/publications/impact-evaluation.html>.

¹² Grupo de Evaluación de las Naciones Unidas, *Impact Evaluation in UN Agency Evaluation Systems: Guidance on Selection, Planning and Management*, documento orientativo, UNEG, Nueva York, 2013. Véase http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=1434.

UNICEF sigue las orientaciones del UNEG, por lo que la elección de la estrategia o estrategias de atribución causal debe basarse en una valoración de lo que es más conveniente para la situación. Esta valoración debe tener en cuenta la naturaleza de:

- el programa o la política que se está evaluando: ¿con qué poblaciones o en qué zonas se está aplicando el programa o la política?, ¿cuáles son las unidades de análisis?, ¿cómo de variable es la ejecución del programa?
- la evaluación: ¿en qué fase se encuentra el programa o la política?, ¿han participado ya los participantes en el programa o la política?, ¿qué datos están ya disponibles?, ¿de qué recursos se dispone en cuanto a tiempo, fondos para evaluadores externos y capacidades?, ¿tienen las otras organizaciones que participan en la evaluación requisitos particulares para el diseño de la evaluación de impacto?

También es posible combinar las tres estrategias básicas en una única evaluación. El análisis de contribución¹³ es un enfoque general que puede englobar distintas estrategias. Funciona mediante un proceso estructurado de documentación y compilación de lo que ya se conoce en términos de la teoría del cambio actual, pruebas empíricas de anteriores investigaciones y evaluaciones, y datos de un programa o una política ya existentes; valora la calidad del relato de contribución y, a continuación, recaba pruebas sistemáticamente para corregir las carencias y puntos débiles identificados.

5. EJEMPLO DE BUENAS PRÁCTICAS

La evaluación de la Declaración de París sobre la Eficacia de la Ayuda al Desarrollo¹⁴ proporciona un ejemplo de atribución causal sistemática en una situación en la que no era posible definir o construir un contrafáctico fiable. El objetivo era evaluar si la Declaración de París se había aplicado satisfactoriamente o no, y en qué medida había contribuido a los resultados de desarrollo, cómo y por qué.

La teoría del cambio describía cómo las acciones programáticas llevarían a los resultados intermedios (en eficacia de la ayuda mejorada) y a los impactos a largo plazo (en contribuciones a resultados de desarrollo mejorados), e incluía varias vías complejas diseñadas para buscar posibles factores contextuales importantes que pudieran impulsar o inhibir el cambio.

Se hizo hincapié en la manera estructurada en que los equipos de evaluación debían utilizar un enfoque de métodos mixtos para evaluar «contribuciones plausibles» de la Declaración de París a los resultados de desarrollo en cada contexto, y también se hizo hincapié en proporcionar «pruebas empíricas claras de los cambios y conexiones observados y otras explicaciones plausibles»¹⁵.

Un marco de evaluación global determinaba los tipos de pruebas empíricas que debían buscar los evaluadores y los métodos o formas de análisis que podían aplicarse. También incluía un sistema de calificación para indicar, por un lado, la importancia de las pruebas halladas para las preguntas clave de evaluación, el grado en que podrían triangularse, y por tanto considerarse fiables, y el grado en que los datos procedían de fuentes recientes y fiables; y, por otro, la medida en que los métodos de recolección de datos y los análisis proporcionaban una base razonable para las constataciones y conclusiones extraídas.

¹³ Mayne, John, «Contribution analysis: An approach to exploring cause and effect», ILAC Brief 16, Institutional Learning and Change, 2008. Véase http://www.cgjar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf.

¹⁴ Wood, Bernard, *et al.*, *The Evaluation of the Paris Declaration*, Informe final, Instituto Danés de Estudios Internacionales, Copenhague, mayo de 2011. Véase <http://www.oecd.org/derec/dacnetwork/48152078.pdf>.

¹⁵ White, Howard y Daniel Phillips, «Addressing attribution of cause and effect in small impact evaluations: towards an integrated framework», *International Initiative for Impact Evaluation Working Paper* No. 15, 3ie, Nueva Delhi, 2012, pág. 12. Véase http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf.

Entre los métodos empleados estaba la revisión de literatura y documentos, el análisis cuantitativo/estadístico de los datos disponibles más relevantes, instrumentos de encuesta, entrevistas, grupos de discusión y análisis de partes interesadas.

6. EJEMPLOS DE POSIBLES PROBLEMAS

Las evaluaciones de impacto de UNICEF a menudo presentan deficiencias en la atribución causal. Todos los siguientes ejemplos provienen de evaluaciones de impacto de UNICEF finalizadas y anónimas.

No se lleva a cabo sistemáticamente la atribución causal

Algunas evaluaciones de impacto no intentan examinar la atribución causal, pero aun así insinúan o afirman una atribución. Otras evaluaciones de impacto no llevan a cabo la atribución causal si no es posible crear un grupo de comparación o un grupo de control. Esto no solo representa una oportunidad perdida —ya que hay varias formas alternativas de abordar la atribución causal, como se describió anteriormente—, sino que contradice la finalidad de la evaluación de impacto. Si no se intenta realizar la atribución causal o se hace de forma inadecuada, pueden extraerse conclusiones falsas sobre la eficacia del programa o política y enfocar equivocadamente la toma de decisiones.

Por ejemplo, una evaluación reconoció que no había investigado las relaciones causales, sin embargo afirmaba que el programa había sido eficaz puesto que se había logrado el cambio previsto (mayor uso de los servicios).

No se describe con precisión el tipo de contrafáctico utilizado

Algunas evaluaciones de impacto utilizan el término «grupo de control» para referirse a cualquier «grupo de comparación» —incluidos aquellos que son malos equivalentes del [grupo de intervención o de tratamiento](#)—, y proporcionan información insuficiente para poder juzgar la calidad del emparejamiento. Por ejemplo, una evaluación se refería a «zonas de control»¹⁶ y reconocía que estas zonas se habían construido utilizando técnicas cuasi experimentales en vez de por asignación aleatoria, pero no proporcionaba información acerca de cómo se habían seleccionado o construido o si se había comprobado su compatibilidad con las «zonas experimentales»¹⁷.

Esto resulta problemático porque los grupos de control y de comparación se construyen de formas muy diversas (como se explicó anteriormente). Repercute en lo bien que puede emparejarse cada grupo construido con el grupo de intervención de referencia (véase la Síntesis n.º 7 (Ensayos controlados aleatorios) y la Síntesis n.º 8 (Diseño y métodos cuasi experimentales)) y, por tanto, en la representación precisa del impacto real de un programa o política.

No se busca o se intenta explicar las pruebas empíricas que no encajan en la teoría del cambio

Algunas evaluaciones de impacto no buscan o intentan explicar las pruebas empíricas que son incoherentes con una relación causal. Por ejemplo, una evaluación afirmaba que un programa de creación de capacidades había producido determinados impactos en una organización, aunque informaba de que el encargado del programa había pasado poco tiempo con la organización y no

¹⁶ No se proporcionan fuentes porque las evaluaciones de impacto se han anonimizado.

¹⁷ *Ibíd.*

había prestado asistencia al personal o la dirección, y que otro programa había impartido capacitación a la organización que se consideraba que había desarrollado las capacidades.

7. LECTURAS Y ENLACES CLAVE

Otras síntesis especialmente pertinentes en esta serie:

- Ensayos controlados aleatorios (Síntesis n.º 7);
- Diseño y métodos cuasi experimentales (Síntesis n.º 8);
- Estudios de caso comparativos (Síntesis n.º 9);
- Elaboración de modelos (Síntesis n.º 13).

BetterEvaluation, «Understand Causes», página web, BetterEvaluation, <http://betterevaluation.org/plan/understandcauses>.

Comisión Europea, *Evalsed Sourcebook: Methods and Techniques*, Comisión Europea, 2013. Véase http://ec.europa.eu/regional_policy/sources/docgener/evaluation/guide/evaluation_sourcebook.pdf.

Gaarder, Marie y Jeannie Annan, «Impact Evaluation of Conflict Prevention and Peacebuilding Interventions», *Policy Research Working Paper* No. 6496, Grupo de Evaluación del Sector Público, Grupo de Evaluación Independiente, Grupo del Banco Mundial, 2013. Véase <http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-6496>.

Hughes, Karl, y Claire Hutchings, «Can we obtain the required rigour without randomisation? Oxfam GB's non-experimental Global Performance Framework», *International Initiative for Impact Evaluation Working Paper* No. 13, 3ie, Nueva Delhi, 2011. Véase http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_13.pdf.

Mayne, John, «Contribution analysis: An approach to exploring cause and effect», ILAC Brief 16, Institutional Learning and Change, 2008. Véase http://www.cgiar-ilac.org/files/ILAC_Brief16_Contribution_Analysis_0.pdf.

Rogers, Patricia J., RMIT University y BetterEvaluation, «Introducción a la evaluación de impacto», *Notas sobre la Evaluación de Impacto* N.º 1, InterAction, 2012. Véase http://www.interaction.org/sites/default/files/3%20-%20Introduction%20to%20Impact%20Evaluation%20-%20Spanish_0_0.pdf.

Rogers, Patricia J., «Matching Impact Evaluation Design to the Nature of the Intervention and the Purpose of the Evaluation», en Chambers, Robert *et al.*, «Designing impact evaluations: different perspectives», *International Initiative for Impact Evaluation Working Paper* No. 4, 3ie, Nueva Delhi, 2009. Véase http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_4.pdf.

Stern, Elliott, *et al.*, «Broadening the range of designs and methods for impact evaluations», *DFID Working Paper* No. 38, Departamento de Desarrollo Internacional, Londres, 2012.

White, Howard y Daniel Phillips, «Addressing attribution of cause and effect in small impact evaluations: towards an integrated framework», *International Initiative for Impact Evaluation Working Paper* No. 15, 3ie, Nueva Delhi, 2012, pág. 43. Véase http://www.3ieimpact.org/media/filer/2012/06/29/working_paper_15.pdf.

White, Howard, «Some Reflections on Current Debates in Impact Evaluation», *International Initiative for Impact Evaluation Working Paper* No. 1, 3ie, Nueva Delhi, 2009. Véase http://www.3ieimpact.org/media/filer/2012/05/07/Working_Paper_1.pdf.

GLOSARIO

<u>Actividad</u>	<i>Acciones emprendidas o procesos mediante los que se movilizan insumos para generar productos determinados. Por ejemplo, sesiones de asesoramiento que se adhieren a unas normas de calidad.</i>
<u>Asignación aleatoria</u>	<i>Proceso de incluir a los participantes o sectores objeto de investigación en un grupo de intervención o un grupo de control de forma que cada individuo o sector de cada grupo sea asignado enteramente al azar. Es decir, cada individuo o sector tiene las mismas probabilidades de ser incluido en cada uno de los grupos.</i>
<u>Causalidad</u>	<i>Principio según el cual una variable (X) produce un cambio en otra variable (Y). Se basa en la suposición de que los hechos ocurren de una forma predecible, no aleatoria, y de que un hecho lleva a otro o lo provoca. Para establecer la causalidad, las dos variables deben asociarse o correlacionarse entre sí; la primera variable (X) debe preceder a la segunda (Y) en tiempo y espacio; y las explicaciones alternativas no causales de la relación (como las explicaciones falsas) deben eliminarse. Los acontecimientos en los mundos físico y social son en general demasiado complejos para ser explicados por un único factor. Por este motivo, los científicos se guían por el principio de causalidad múltiple, que afirma que un hecho se produce como consecuencia de varios factores que funcionan u ocurren en combinación.</i>
<u>Correlación</u>	<i>Relación mutua o asociación de dos o más conceptos o variables, de manera que cuando cambia el valor de una también cambia el de la otra. Las variables pueden correlacionarse de forma positiva (cambian en la misma dirección) o negativa (cambian en direcciones opuestas). La correlación es necesaria pero no suficiente para demostrar la causalidad.</i>
<u>Diseño contrafáctico</u>	<i>Diseño en el que se busca entender las causas comparando los resultados observados con los que cabría esperar si no se hubiese ejecutado la intervención; implica el uso de un grupo de control o un grupo de comparación.</i>
<u>Enfoque contrafáctico</u>	<i>Elaborar una estimación de lo que habría pasado en ausencia del programa o política, lo cual implica utilizar un grupo de control o un grupo de comparación.</i>
<u>Ensayos controlados aleatorios</u>	<i>Diseño de investigación o evaluación con dos o más grupos seleccionados de forma aleatoria (un grupo experimental y un grupo de control) en los que el investigador controla o introduce una intervención (por ejemplo un nuevo programa o política) y mide su impacto en la variable dependiente al menos dos veces (medición anterior y posterior al ensayo). En concreto, los ensayos controlados aleatorios —que tiene su origen en el contexto clínico y se conocen como la «regla de oro» de la investigación médica y sanitaria— se suelen utilizar para responder a las preguntas de investigación de la evaluación, que buscan evaluar la eficacia de las intervenciones de un programa o política en entornos de desarrollo.</i>

<u>Estudio de caso</u>	<i>Examen exhaustivo de uno o más casos (por ejemplo, individuos, grupos, instituciones, países, procesos) diseñado y llevado a cabo para entender a fondo y de manera bien organizada a los sujetos que se están examinando. Los estudios de caso pueden abordar la microsituación de una única persona en su día a día o la macrosituación de un Estado, o incluso procesos mundiales. Los resultados pueden utilizarse como constataciones autónomas o pueden integrarse como contribución en síntesis o análisis comparativos más amplios.</i>
<u>Evaluación de impacto</u>	<i>Evaluación que proporciona información sobre los impactos que produce una intervención. Puede realizarse una evaluación de impacto de un programa o una política o del trabajo preliminar, como la creación de capacidad, la promoción de políticas y el apoyo a la creación de un entorno propicio. Esto supone examinar no solo las metas y objetivos, sino también los impactos no previstos.</i>
<u>Extrapolabilidad</u>	<i>Capacidad de aplicar los resultados de un determinado estudio a grupos o situaciones externos a los que se han estudiado.</i>
<u>Grupo de comparación</u>	<i>En un diseño de investigación cuasi experimental, es el grupo de participantes en la investigación que, a efectos de comparación, no recibe el tratamiento o la intervención dados al grupo de intervención. Los sujetos del grupo de comparación normalmente no se distribuyen de forma aleatoria por su condición, como ocurriría con los sujetos del grupo de control en un estudio de diseño experimental. Véase: grupo de control, grupo de tratamiento.</i>
<u>Grupo de intervención</u>	<i>También llamado «grupo experimental» o «grupo de tratamiento». Es un grupo de participantes en la investigación que recibe alguna forma de tratamiento o intervención, es decir, está expuesto a la variable independiente. Véase: grupo de comparación, grupo de control.</i>
<u>Grupo de tratamiento</u>	<i>Participantes expuestos a la variable independiente; también llamado «grupo experimental».</i>
<u>Grupo de control</u>	<i>Participantes en un estudio de investigación/evaluación que no reciben el tratamiento/intervención experimental.</i>
<u>Impacto</u>	<i>Efectos de largo plazo positivos y negativos, primarios y secundarios, producidos directa o indirectamente por una intervención para el desarrollo, intencionalmente o no¹⁸.</i>
<u>Insumos</u>	<i>Recursos financieros, humanos y materiales empleados en un programa o una política. Por ejemplo, los materiales de capacitación elaborados.</i>
<u>Intercambio de ideas</u>	<i>Estrategia mediante la cual se generan ideas a través de una charla intensiva y desenfadada. El objetivo es generar tantas ideas como sea posible —incluso las que inicialmente puedan parecer absurdas— y analizarlas y evaluarlas en una fase posterior.</i>
<u>Población (objeto de estudio)</u>	<i>Grupo de individuos (o instituciones, programas u otros sujetos de estudio) sobre los que un investigador intenta generalizar. Para generalizar sobre una población, los investigadores estudian una muestra que se pretende que sea representativa de la población. Véase: muestra.</i>
<u>Proceso con base empírica</u>	<i>Identificación y valoración sistemáticas de las «mejores» pruebas, sobre la base de su nivel, calidad, puntos fuertes y pertinencia para un determinado tema de interés.</i>

¹⁸ CAD-OCDE, *Glosario de los principales términos sobre evaluación y gestión basada en resultados*, OCDE, París, 2010. Véase <http://www.oecd.org/development/peer-reviews/2754804.pdf>.

<u>Producto</u>	<i>Efectos inmediatos de las actividades de un programa o política, o los productos o entregables directos de estas actividades. Por ejemplo, el número de vacunas administradas.</i>
<u>Promedio</u>	<i>Medida de la tendencia central, llamada comúnmente «media», que se calcula dividiendo la suma de todos los valores entre el número de valores.</i>
<u>Regresión</u>	<i>Procedimiento estadístico para predecir valores de una variable dependiente sobre la base de los valores de una o más variables independientes.</i>
<u>Relación causal</u>	<i>Relación establecida que muestra que una variable independiente provoca un cambio en una variable dependiente. Establece también cuánto cambio se muestra en la variable dependiente.</i>
<u>Resultado</u>	<i>Efectos intermedios de los productos de un programa o política, como por ejemplo un cambio en los niveles de vacunación o los comportamientos clave.</i>
<u>Sesgo de selección</u>	<i>Sesgo en la forma de seleccionar el grupo experimental y el grupo de control o de comparación, que da lugar a diferencias preexistentes entre los grupos que pueden actuar como factores de desviación.</i>
<u>Teoría del cambio</u>	<i>Explica cómo se entiende que las actividades produzcan una serie de resultados que contribuyen a lograr los impactos finales previstos. Puede elaborarse para cualquier nivel de intervención: un acontecimiento, un proyecto, un programa, una política, una estrategia o una organización.</i>
<u>Variable</u>	<i>Medida claramente definida y cuantificable de una cantidad que puede variar en el tiempo para una unidad de observación específica o en el espacio, es decir, entre distintas unidades de observación.</i>