

Curso en Metodologías Rigurosas para la Síntesis de Evidencia

Guía de estudio y
sistematización
del curso¹

Lima, Perú
Marzo del 2018

¹ Este curso se llevó a cabo en Lima, Perú del 17 al 21 de Abril de 2017 gracias al apoyo del TTI -Think Tank Initiative- y la red ILAIPP -Iniciativa Latinoamericana de Investigación para las Políticas Públicas-, en el marco de su programa de fortalecimiento de capacidades para el año 2017.

Índice

Objetivo del curso	6
Introducción al curso	6
Experiencia de aprendizaje del trabajo práctico	8
Sesión 1. Introducción a las metodologías rigurosas para la síntesis de evidencia	9
<i>Objetivos de la sesión</i>	9
<i>Síntesis rigurosas de evidencia</i>	9
<i>Uso e importancia de las síntesis rigurosas de evidencia</i>	9
<i>Tipos de síntesis rigurosa de evidencia que utilizan métodos de revisión sistemática</i>	10
<i>Pasos para el desarrollo de una síntesis rigurosa de evidencia</i>	11
<i>Diferencias entre las síntesis rigurosas y las revisiones de literatura convencionales</i>	12
<i>Importancia de definir adecuadamente el problema a investigar</i>	13
<i>Uso de síntesis rigurosas de evidencia para la práctica, la toma de decisiones de política y el debate público</i>	13
<i>Consideraciones finales</i>	14
<i>Bibliografía de la sesión</i>	15
Sesión 2. Otras metodologías de síntesis rigurosa de evidencia	16
<i>Objetivos de la sesión</i>	16
<i>Metodología de síntesis rigurosa de evidencia</i>	16
<i>Métodos alternativos de síntesis rigurosa de evidencia</i>	17
<i>Bibliografía de la sesión</i>	24
Sesión 3. Teoría de cambio	25
<i>Objetivos de la sesión</i>	25
<i>Preguntas de investigación</i>	25
<i>Teoría de cambio</i>	25
<i>Elementos de una teoría de cambio</i>	26
<i>Utilidad de una teoría de cambio</i>	28
<i>Etapas para el desarrollo de una teoría de cambio</i>	28
<i>La teoría de cambio y las síntesis rigurosas de evidencia</i>	28
<i>El marco lógico y su relación con la teoría de cambio</i>	31
<i>Bibliografía de la sesión</i>	33
Sesión 4. Búsqueda sistemática de literatura, identificación de fuentes y sistemas de generación de mecanismos de búsqueda	34
<i>Objetivos de la sesión</i>	34
<i>Consideraciones generales</i>	34
<i>Cómo realizar búsquedas rigurosas</i>	35

Uso de bases de datos Vs. Google Scholar	39
Bibliografía de la sesión	40
Sesión 5. Criterios de Inclusión y Exclusión de Pautas Para la Evaluación	
Critica de Estudios	41
<i>Objetivos de la sesión</i>	41
<i>La evaluación crítica de estudios</i>	41
<i>Criterios “PICO”</i>	41
<i>Jerarquía de métodos cuantitativos para la estimación de efectos causales</i>	44
<i>Validez de los estudios cuantitativos</i>	45
<i>Resumen del proceso multietápico o de embudo</i>	45
<i>Consideraciones finales</i>	47
<i>Bibliografía de la sesión</i>	48
Sesión 6. Extracción de datos: meta-análisis (cuantitativo)	49
<i>Objetivos de la sesión</i>	49
<i>Objetivos del meta-análisis</i>	49
<i>Cálculo del tamaño de efecto</i>	49
<i>Combinando los Tamaño de Efecto (Pooled Effect Sizes)</i>	54
<i>Análisis de sensibilidad, heterogeneidad y “forest plot”</i>	57
<i>Sesgo de Publicación</i>	59
<i>Bibliografía de la sesión</i>	60
Sesión 7. Extracción de datos: meta-síntesis (cualitativo)	61
<i>Objetivos de la sesión</i>	61
<i>Uso e importancia de las síntesis rigurosas de estudios cualitativos</i>	61
<i>Criterios y estrategias para evaluar la calidad de la investigación cualitativa</i>	62
<i>Bibliografía de la sesión</i>	64
Sesión 8. Síntesis rigurosas y su incidencia en las políticas públicas	65
<i>Objetivos de la sesión</i>	65
<i>Ventajas y limitaciones de las síntesis rigurosas de evidencia en términos de incidencia</i>	65
<i>Ejemplos internacionales del uso de evidencia rigurosa</i>	67
<i>Requisitos conducentes a lograr altos niveles de incidencia</i>	67
<i>Plan de comunicaciones, incidencia e involucramiento de actores</i>	68
<i>La importancia y el rol del comité consultivo</i>	69
<i>Bibliografía de la sesión</i>	70

Índice de tablas

Tabla nº 1: Diferencias entre síntesis rigurosas y revisiones de literatura convencionales	12
Tabla nº 2: Ejemplo de revisión rápida de evidencia sobre los determinantes de la violencia en Sudáfrica. Davies, et al (2011)	19
Tabla nº 3: Orden ascendente de las metodologías rigurosas de síntesis	23
Tabla nº 4: Resumen metodologías de síntesis rigurosas	23
Tabla nº 5: Características del marco lógico y la teoría de cambio	31
Tabla nº 6: Diferencias entre el marco lógico y la teoría de cambio	32
Tabla nº 7: Clasificación de fuentes de información para la búsqueda sistemática de estudios sobre entrenamiento laboral femenino	37
Tabla nº 8: Diferencias entre el uso de base de datos especializadas y Google Scholar	40
Tabla nº 9: Hoja resumen “PICOs”	48
Tabla nº 10: Tamaños de efecto más usados en estudios de investigación	50
Tabla nº 11: Tamaños de efecto más usados en evaluaciones de impacto	51
Tabla nº 12: Fórmulas para calcular información para estimar d-Cohen o g-Hedges	52
Tabla nº 13: Fórmulas para el cálculo del tamaño de efecto promedio	55
Tabla nº 14: Ejemplo de Comité Consultivo. Chinen, et al (2017)	70

Índice de figuras

Figura nº 1: Ejemplo de “evidence gap map” para intervenciones orientadas a niños	18
Figura nº 2: Ejemplo de revisión rápida de evidencia: Why is crime in South Africa so violent? Explicaciones teóricas	20
Figura nº 3: Ejemplo de revisión de alcance de evidencia: análisis de intervenciones para mitigar el riesgo agrícola	22
Figura nº 4: Distintos niveles de cambio: de actividades a resultados	26
Figura nº 5: Elementos de una teoría de cambio	27
Figura nº 6: Ejemplo de teoría de cambio. “Breakfast and school performance”. Cueto (2001)	29
Figura nº 7: Ejemplo de teoría de cambio: Programa “Young Lives”	30
Figura nº 8: Flujograma del proceso de revisiones sistemáticas	35
Figura nº 9: Jerarquía de métodos para la estimación de efectos causales	43
Figura nº 10: Ejemplo de proceso multietapa	46
Figura nº 11: Ejemplo de tamaño de efecto. Asistencia docente. Guerrero, et al (2013)	56
Figura nº 12: Ejemplo de tamaño de efecto. Logro escolar del estudiante. Guerrero, et al (2013)	56
Figura nº 13: Ejemplo de “Forest Plot”. Estudios experimentales - Efecto de programas de entrenamiento vocacional sobre el empleo	57
Figura nº 14: Ejemplo de “Forest Plot”. Estudios cuasi experimentales - Efecto de programas de entrenamiento vocacional sobre el empleo	58
Figura nº 15: Ejemplo de análisis de heterogeneidad. Estudios experimentales - Programas vocacionales en África/Asia Vs. otros continentes	58
Figura nº 16: Ejemplo de análisis de heterogeneidad. Estudios experimentales - Programas vocacionales con enfoque de género Vs. sin enfoque de género	59
Figura nº 17: Funnel Plot para evaluar el sesgo de publicación	60
Figura nº 18: Contexto en el cual busca incidir las revisiones sistemáticas	66

Objetivo del curso

El curso consistió en plantear una mirada integral a la utilidad y relevancia de las metodologías de síntesis rigurosa de evidencia, su uso para la investigación y su función en el contexto de la formulación de políticas públicas. En este marco, se discutieron diferentes metodologías de síntesis rigurosa como lo son las revisiones sistemáticas, los mapeos de vacíos en la evidencia (evidence gap-maps), las revisiones rápidas de evidencia (rapid evidence reviews) y las revisiones para determinar el alcance de la evidencia existente sobre algún tema (scoping reviews).

Se proporcionaron herramientas conceptuales y técnicas para la construcción de modelos de cambio y preguntas de investigación, así como para el manejo de fuentes y búsqueda de literatura en sistemas de información. Se cubrieron temas cuantitativos y cualitativos relevantes a lo largo de todo el proceso de síntesis y, finalmente, se relacionó el mismo con las labores de incidencia que deben producirse antes, durante y después de los ejercicios de síntesis rigurosa.

El curso fue dictado a través de un módulo presencial con una duración de cinco días en los que se desarrollaron sesiones teóricas y prácticas para los que se armaron equipos y se puso en práctica los contenidos desarrollados por el equipo docente en cada componente del curso, mediante el desarrollo de un esquema de protocolo de revisión sistemática. Luego del curso, algunos participantes continuaron el proceso de aprendizaje con mentorías sobre temas específicos basados en la elaboración de protocolo de síntesis rigurosas de evidencia sobre temas relevantes para la región. Estas mentorías se realizaron a distancia pero incluyendo además una visita presencial y tuvieron una duración de hasta seis meses.

La presente guía de estudio surge como el resultado de la sistematización de las experiencias y los aprendizajes obtenidos a través de la participación en este curso. Esta guía recoge todos los contenidos cubiertos a lo largo del curso y provee ejemplos prácticos para profundizar los conocimientos y potenciar los aprendizajes significativos. Además, pretende servir como un manual de consulta práctica para profesionales desempeñando labores en centros de investigación social y Organizaciones no Gubernamentales (ONG) abocadas al estudio del desarrollo.

Introducción al curso

A principios de la década de los 90, las ciencias médicas y de la salud experimentaron un cambio en la forma de abordar el conocimiento científico, algunas veces contradictorio, derivado de múltiples estudios de interés en torno a un mismo tema. Durante aquellos años se observa una notoria expansión en la formulación de lineamientos, prácticas y metodologías basadas en evidencia, en la medida en que se hizo cada vez más clara la necesidad de recopilar, resumir y generalizar tal conocimiento. En los años posteriores, las ciencias sociales fueron reconociendo el atractivo y la utilidad práctica de aplicar este abordaje a las problemáticas sociales, de manera de contribuir a un diseño acertado de políticas sociales y evaluar la eficacia de intervenciones públicas desplegadas alrededor del mundo en entornos y condiciones diferenciadas. Así, se fue formando un instrumental metodológico lo suficientemente amplio como para admitir la existencia de distintos tipos de síntesis rigurosa de evidencia, las cuales hoy

día son utilizadas en estudios con propósitos diversos.

La presente guía de estudio ofrece una mirada integral a la utilidad y relevancia de las metodologías de síntesis rigurosa de evidencia más comúnmente utilizadas, su uso para la investigación y su función en el contexto de la formulación y diseño de políticas públicas de nivel local, regional, nacional y global.

La primera sesión introduce la motivación de estudiar la aplicación de metodologías rigurosas para la síntesis de evidencia relacionada a intervenciones sociales provenientes del sector público o del tercer sector (ONGs) y sus implicancias en materia de política pública. La segunda sesión explora en detalle algunos temas de la primera sesión y se adentra en el proceso de revisión sistemática, así como de metodologías alternativas de síntesis rigurosas. La tercera sesión aborda aspectos conceptuales de la evaluación de programas como son la importancia de contar con una teoría de cambio del programa correctamente diseñada a la par de tener preguntas de investigación claras, significativas y relevantes. La cuarta sesión se concentra en proporcionar herramientas para la búsqueda sistemática de literatura, la identificación de fuentes y el manejo de sistemas de generación, búsqueda y sistematización de información. La quinta sesión expone los criterios de inclusión y exclusión para la selección y descarte de estudios como una fase fundamental del proceso de evaluación y análisis de síntesis rigurosa. La sexta y la séptima sesión se complementan mutuamente: la sexta cubre los aspectos cuantitativos de las revisiones sistemáticas, llamado meta-análisis, cuya cruzada es medir tamaños de efecto y controlar la posible influencia de diversos tipos de sesgos. Por su parte, la séptima sesión abarca los aspectos cualitativos, llamado meta-síntesis, los cuales se centran en comprender las causas, los mecanismos y el contexto donde ocurren tales efectos. Finalmente, la octava sesión cierra el curso relacionando el proceso de síntesis rigurosas con la labor de incidencia e impacto en la formulación y diseño de mejores políticas públicas basadas en evidencia.

Fórmula	
Sesión	Tema
1	Introducción a las metodologías rigurosas para la síntesis de evidencia
2	Otras metodologías de síntesis rigurosa de evidenci
3	Teoría de cambio y preguntas de investigación
4	Búsqueda sistemática de literatura, identificación de fuentes y sistemas de generación de mecanismos de búsqueda
5	Criterios de Inclusión y Exclusión de Pautas Para la Evaluación Crítica de Estudios
6	Extracción de datos: meta-análisis (cuantitativo)
7	Extracción de datos: meta-síntesis (cualitativo)
8	Síntesis rigurosas y su incidencia en las políticas públicas

Experiencia de aprendizaje del trabajo práctico

Este curso se llevó a cabo a través de un conjunto de jornadas de capacitación en un formato presencial de cinco días de duración en los que se desarrollaron sesiones de carácter teórico y práctico. Los participantes se organizaron en equipos para elaborar de forma simultánea y progresiva una propuesta de proyecto de investigación de una síntesis rigurosa aplicada a un tema o pregunta en específico, a medida que los contenidos del curso se iban cubriendo. Algunos equipos tuvieron la oportunidad de continuar el proceso de mentoría y profundizar en sus respectivas propuestas de proyecto de investigación para consolidarlo en un protocolo de síntesis rigurosa durante un período de hasta seis meses.

En tal sentido, recomendamos como una actividad complementaria importante para el aprendizaje significativo que la provisión de este curso venga acompañada de jornadas de trabajo práctico que permitan aplicar y afianzar los conocimientos adquiridos. En ausencia de un componente práctico que permita relacionar los elementos conceptuales del curso con los retos que impone la realización de una propuesta de síntesis rigurosa, el presente documento sólo pretende servir como una guía sintética y esquemática de consulta rápida.

Sesión 1. Introducción a las metodologías rigurosas para la síntesis de evidencia

Objetivos de la sesión

- Proponer una mirada introductoria al concepto, utilidad, relevancia, uso y gestión de las síntesis rigurosas de evidencia para la toma de decisiones en las políticas públicas.
- Mostrar la importancia de una adecuada definición del problema de estudio en una síntesis rigurosa de evidencia, los tipos de preguntas que es capaz de responder y las diferencias existentes con las revisiones de literatura tradicionales.

Síntesis rigurosas de evidencia

Las síntesis rigurosas de evidencia son un tipo de revisión de literatura que identifica y analiza de forma crítica y sistemática múltiples investigaciones y publicaciones relacionadas a un tema en concreto usando métodos que son seleccionados a partir de las preguntas de investigación de interés. Es una práctica comúnmente aplicada en las ciencias médicas y de la salud, que se ha extendido progresivamente a la evaluación y análisis de políticas e intervenciones educativas, sociales, económicas y ambientales, donde el uso y la publicación de datos en cuerpos de literatura especializada internacional permiten juzgar su calidad metodológica desde una perspectiva rigurosa y sistemática.

Su objetivo principal es congrega y mostrar la evidencia científica existente en torno a un tema específico. La evidencia es una información o un conjunto de datos que sustenta o contradice una aseveración o conclusión. Esta casi siempre tiene un componente probabilístico. Es decir, rara vez es absoluta o definitiva; sobre todo en el caso de las ciencias sociales. Contra intuitivamente, la evidencia suele producir desacuerdos en términos de lo que se considera un hecho o algo “evidente”. En otras palabras, muchas evidencias son refutables, a veces por la calidad de los datos y de la metodología utilizada para producirlos, pero también por el surgimiento de nueva evidencia.

Uso e importancia de las síntesis rigurosas de evidencia

Las síntesis rigurosas de evidencia son de vital importancia cuando se busca analizar una creciente y a veces caótica masa de evidencia, que puede, en muchos casos, arrojar descubrimientos o conclusiones contradictorias. La síntesis de evidencia permite sistematizar y simplificar a través del método científico esa masa abrumadora e inmanejable de conocimiento. En este sentido, contribuye al proceso

de toma de decisiones basado en evidencia o informado a partir del conocimiento producido por la investigación científica; en contraposición a basar estas decisiones en estudios aislados de dudosa reputación o, peor aún, en intuiciones, percepciones, ideologías o prejuicios.

Más aún, basar las decisiones de política en estudios individuales o aislados puede desencadenar los siguientes inconvenientes relacionados al sesgo de la evidencia sobre un tema en concreto:

- Presentar una visión limitada o sesgada del balance de evidencia de investigación disponible.
- Dar luces solamente sobre algún aspecto específico del problema de política que nos concierne.
- Los resultados pueden tener validez sólo para el grupo que conforma la muestra, o para un tiempo o contexto específico.
- Muchas veces las evidencias pueden ser de calidad pobre o carecer de credibilidad.

Al respecto, las síntesis rigurosas de evidencia resuelven muchos de estos inconvenientes dado que cuentan con las siguientes características:

- No poseen sesgos de cobertura, es decir, incluyen toda la literatura disponible, publicada o no publicada, en cualquier idioma, sobre el tema de interés (que puede muy delimitado).
- Son rigurosas en el sentido de que identifican toda la evidencia disponible y utilizan críticamente toda la literatura incluida y excluida de la revisión.
- Son transparentes porque los métodos utilizados para la revisión son reportados de forma clara y suficiente para que la revisión pueda ser replicada por otros investigadores (peer reviews).
- Son un bien público global ya que las organizaciones vinculadas al tema se pueden beneficiar del conocimiento producido.
- Deben mantenerse actualizadas a medida que surge nueva evidencia.

Tipos de síntesis rigurosa de evidencia que utilizan métodos de revisión sistemática

Todos los tipos de síntesis siguen las mismas reglas básicas en cuanto a rigurosidad, y transparencia, aunque varían en cuanto a sus objetivos, el alcance de sus conclusiones, el uso que se puede hacer de ellas, sus costos y su duración. Entre las más comunes tenemos a los mapas de brechas de evidencia o “evidence gap maps”, las revisiones rápidas de evidencia o “rapid evidence assessments”, las revisiones de alcance “Scoping reviews” y las revisiones sistemáticas o “systematic reviews”.

La explicación a profundidad de cada tipo de síntesis será expuesta en la sesión 2 de esta guía.

Pasos para el desarrollo de una síntesis rigurosa de evidencia

Uno de los elementos característicos de las síntesis rigurosas de evidencia es que son elaboradas a partir de un protocolo en el cual se especifican todos los parámetros de la síntesis:

1. Se define claramente el problema a analizar
2. Se define en el protocolo una pregunta que sea posible responder y los métodos planteados para responderla.
3. Se determina el alcance de la evidencia disponible:
4. Se definen criterios para la búsqueda de información (terminología y fuentes a consultar; metodología de búsqueda)
5. Se incorporan criterios de inclusión y exclusión de estudios (PICOS) para determinar qué será finalmente incluido en la revisión.
6. Se define la metodología a partir de la cual se llevará a cabo la síntesis

Una vez elaborado el protocolo:

7. Se conducen búsquedas exhaustivas para identificar estudios publicados y no publicados, sea cual sea su origen o idioma (de acuerdo a los parámetros y alcance de búsqueda definidos previamente).
8. Se efectúa una evaluación crítica de la calidad de los estudios.
9. Se extrae y organiza de forma sistemática los datos disponibles en los estudios incluidos
10. Se desarrolla una síntesis de evidencia, la cual expone resultados obtenidos a lo largo de una cadena causal claramente planteada.
11. Se interpretan y comunican los resultados a diferentes actores y comunidades de práctica y se formulan de recomendaciones de política.

Fórmula		
Antes de elaborar el protocolo	1	Definir claramente el problema a analizar
	2	Definir una pregunta posible responder y métodos para responderla
	3	Construir una teoría de cambio de acuerdo al tipo o eje de intervención
	4	Determinar el nivel y alcance de evidencia disponible
	5	Definir criterios para la búsqueda de información
	6	Incorporar criterios de inclusión y exclusión de estudios
	7	Definir la metodología a partir de la cual se realizará la síntesis
Después de elaborar el protocolo	8	Conducir búsquedas exhaustivas para identificar estudios
	9	Efectuar una evaluación crítica de la calidad de los estudios
	10	Extraer y organizar sistemáticamente los datos disponibles en los estudios
	11	Desarrollar una síntesis de evidencia que muestre resultados
	12	Interpretar y comunicar los resultados a los distintos actores

Diferencias entre las síntesis rigurosas y las revisiones de literatura convencionales

Se puede afirmar que ambas poseen objetivos distintos, pero principios comunes. Ambas están basadas en el método científico y son realizadas para mapear el conocimiento que ha sido producido por la comunidad científica especializada a lo largo del tiempo en torno a un tema. En líneas generales, las síntesis rigurosas buscan abarcar toda la evidencia disponible sobre un tema bajo un enfoque metodológico más exhaustivo y transparente que el de las revisiones de literatura convencionales ya que requieren, entre otros elementos, de la realización de un protocolo o documento orientador que especifique claramente los criterios que van a determinar la revisión de ciertos tipos de estudios. En el caso de las revisiones de literatura el proceso de selección de estudios suele suceder de una manera más emergente o arbitraria, ya que carece de un protocolo. Por tanto se dice que las primeras son metodológicamente más rigurosas .

Las revisiones de literatura podrían beneficiarse en la medida en que incorporen algunos de los criterios de transparencia y rigurosidad característicos de las síntesis rigurosas.

A continuación se presentan una serie de diferencias en base a sus características:

Tabla n° 1: Diferencias entre síntesis rigurosas y revisiones de literatura convencionales

Característica	Síntesis rigurosas	Revisiones de literatura
Definición	Revisión exhaustiva de investigación primaria que identifica, selecciona, sintetiza y evalúa toda la evidencia disponible en torno a una pregunta	Resumen cualitativo de evidencia en torno a un tema usando métodos informales y subjetivos para recolectar e interpretar estudios
Pregunta	Pregunta acotada y contestable	Puede ser de un tema en general o de una pregunta específica
Perspectiva	Postura o posición neutral	Postura o posición definida
Componentes	Especificación de criterios de elegibilidad Estrategia de búsqueda sistemática Evaluación de la validez de hallazgos Interpretación de resultados Lista de referencias	Introducción Métodos Discusión Conclusión Lista de referencias
Cobertura	Exhaustiva (literatura publicada y no publicada)	Selectiva (literatura publicada)
Audiencia	Hacedores de política	Académicos
Proceso toma de decisiones	Protocolo claro de inclusión y exclusión, extracción de datos, análisis y reportes	Menor transparencia en la toma de decisiones por la ausencia de protocolo
Evaluación crítica	Consideración a profundidad de aspectos de validez interna y externa	Sólo se menciona diseño y hallazgos de los estudios
Equipo de estudio	Conducido en equipo (doble codificación)	Conducido individualmente

Fuente: Kysh, Lynn (2013)

Importancia de definir adecuadamente el problema a investigar

En una buena síntesis rigurosa tiene que haber una adecuada definición del problema en torno al cual el investigador centra su atención, ya sea para sintetizar evidencia o para presentar los hallazgos de un nuevo estudio. Para ello es necesario definir, primeramente, el trasfondo (background) del problema. Esta parte en cierta medida equivale a una revisión de literatura usual basándose en literatura clave (key papers) identificada a través de búsquedas preliminares y/o de nuestro propio conocimiento de un tema de una manera más intencional. Luego, se hacen explícitas la(s) forma(s) en que se ha buscado solucionar el problema en cuestión y la teoría de cambio correspondiente a esas intervenciones.

La definición del problema y la formulación de las preguntas determinarán qué tipo de literatura incluir en la síntesis. Algunas preguntas, como las de impacto, nos llevarán a incluir estudios cuantitativos experimentales. Otras preguntas como las de proceso, diseño y/o factores asociados, nos llevarán hacia estudios cualitativos.

En las síntesis rigurosas no existe un estándar de oro en cuanto a qué tipo de estudios o evidencia incluir (i.e. no necesariamente se debe incluir solo metodologías experimentales). Como ya se ha mencionado, la relevancia de estudios conducidos a partir de metodologías diferentes se determina en función de las preguntas de investigación.

El análisis de calidad de los estudios, ya sean cuantitativos o cualitativos, se determina al momento de definir y evaluar el 'riesgo de sesgo' y de llevar a cabo la 'evaluación crítica' de los estudios incluidos, la cual será explicada en detalle en sesiones posteriores de esta guía.

Uso de síntesis rigurosas de evidencia para la práctica, la toma de decisiones de política y el debate público

El cómo y el por quién será usada la síntesis de evidencia son aspectos que se deben tomar en cuenta desde el primer momento de la investigación. Así, luego de realizar el planteamiento del problema e identificar posibles preguntas de investigación, es preciso hacerse las siguientes preguntas:

- ¿Para quiénes podría ser útil y relevante la revisión?
- ¿Qué podemos hacer para fomentar el uso de los hallazgos de la revisión?
- ¿Con qué perspectivas pueden contribuir los potenciales usuarios a la definición del problema y de las preguntas, así como a la lectura crítica de la revisión?

El reconocimiento y la gestión inteligente de los grupos de interés, actores involucrados o “stakeholders” constituyen un elemento clave para todo el proceso de síntesis rigurosa. Al respecto, es ideal establecer desde el primer momento una estrategia comunicacional y de involucramiento de actores que especifique a quienes incluir (tomadores de decisiones, profesionales, expertos globales) y cómo comunicar los hallazgos (conducción de reuniones, presentaciones, difusión de informes y “policy briefs”, etc).

Consideraciones finales

La rigurosidad metodológica es fundamental pero no siempre sinónimo de una mejor síntesis rigurosa. Por un lado, existen síntesis rigurosas metodológicamente impecables pero cuyo aporte al debate público y a la conceptualización de los problemas sociales es mínima o nula. Por otro, existen revisiones de literatura muy bien conducidas cuyo aporte ha sido vital para cerrar brechas de conocimiento en sus disciplinas.

En líneas generales, tanto en una buena síntesis rigurosa como en una buena revisión de literatura, es esencial analizar los hallazgos críticamente y sopesar los argumentos a la luz de la teoría y el instrumental metodológico disponible de manera clara y transparente.

Bibliografía de la sesión

- Deep, A., Thillairajan, A., John, P., & Mahalingam, A. (2015). Impact of Private Sector Participation on access and quality of services: systematic review of evidence from the electricity, telecommunications and water supply sectors. *Journal of Development Effectiveness*, Vol. 7, 64-89.
- Anderson, A. A. (2009). *The Community Builder's approach to Theory of Change: A practical guide to theory development*. New York : The Aspen Institute.
- Cirera , X., & Lakshman, R. (2014). *The impact of export processing zones on employment, wages and labour conditions in developing countries: systematic review*. Londres: International Initiative for Impact Evaluation (3ie).
- Guerrero, G., Cueto, S., Leon, J., & Zapata, M. (2013). *Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries*. *Journal of Development Effectiveness*.
- Ortiz , A., & Rivero, G. (2007). *Desmitificando la teoría de cambio*. Washington, DC: Pact Organization.
- Snilstveit, B. (2016). *The impact of education programmes on learning and school participation in low and middle-income countries*. Londres: International Initiative for Impact Evaluation (3ie) Systematic Review Summary 7.
- Vogel, I. (2013). *ESPA guide to working with Theory of Change for research projects*. Londres: Ecosystem Services for Poverty Alleviation (ESPA).
- Kysh, L (2013). *Difference between a systematic review and a literature review*. Presentation from Medical Library Group of Southern California & Arizona (MLGSCA).

Sesión 2. Otras metodologías de síntesis rigurosa de evidencia

Objetivos de la sesión

- Introducir otros métodos rigurosos de síntesis de evidencia diferentes a las revisiones sistemáticas
- Discutir la aplicación de cada uno de estos métodos mediante ejemplos.
- Caracterizar cada uno de estos métodos alternativos e identificar diferencias entre ellos.

Metodología de síntesis rigurosa de evidencia

La primera sesión introdujo la temática del curso para así generar discusiones preliminares y de expectativas entre participantes y docentes. Estas conversaciones permitieron reconocer que nos movemos en un contexto de creciente y abrumadora evidencia en términos de lo que funciona o no de intervenciones públicas, programas y políticas sociales. Internacionalmente, las decisiones de políticas se basan cada vez más en evidencia, la cual, como hemos mencionado, suele proceder de 3 tipos de fuentes:

- Revisión de uno o pocos estudios donde no se deja claro cuáles fueron los criterios de elección
- Revisiones de literatura convencionales
- Revisiones sistemáticas basadas en métodos rigurosos

Ante la existencia de esta masa de evidencia en muchos casos confusa, dispersa, contradictoria y de reputación cuestionable, la solución más incluyente y rigurosa es conducir revisiones sistemáticas. Esto se debe a que cubren toda la literatura disponible -tanto publicada como no publicada-, conllevan una evaluación y de riesgo muy estricta de cada estudio incluido, poseen altos niveles de transparencia y reproductibilidad, y ofrecen un balance completo y no sesgado de la evidencia empírica al sólo considerar evidencia de alta calidad.

No obstante, las revisiones sistemáticas tampoco deben considerarse la panacea. También involucran limitaciones retadoras, como lo son los largos plazos que se requieren para llevarlas a cabo -entre 12 y 18 meses-, los cuantiosos recursos económicos requeridos por la necesidad de contar con un equipo numeroso compuesto por especialistas en sistemas de información, el tema de interés y el proceso de análisis (estadísticos y econométricos, entre otros). Además, las revisiones sistemáticas pueden ser restrictivas en el sentido de que sólo consideran estudios muy rigurosos y comparables donde se presentan abundantes resultados de manera de poder proceder con la realización del meta-análisis y/o meta síntesis.

Métodos alternativos de síntesis rigurosa de evidencia

En la presente sesión se discuten tres métodos alternativos de síntesis rigurosa de evidencia que se utilizan comúnmente en la cooperación internacional y en las organizaciones abocadas al desarrollo: 1) el mapa de brechas de evidencia, 2) la revisión rápida de evidencia, y 3) la revisión de alcance de evidencia. Del mismo modo, se expresan las ventajas y desventajas de cada uno:

1. **“Evidence gap maps” o mapas de brechas de evidencia:** este método hace revisiones de evidencia en un tema, sector, o intervención particular pero con características similares. Se conducen mapeando de manera sistemática y rigurosa dos tipos de estudios: estudios primarios (evaluaciones de impacto) y estudios secundarios (revisiones sistemáticas), usando como sustento la teoría de cambio de una intervención o política y sus resultados intermedios y finales.

La construcción de este mapa de resultados refleja indicadores en forma de globos que dan cuenta, a través de su tamaño y color, del número relativo y la calidad de estudios conducidos alrededor de un tema en específico. Es decir, mientras más grande el círculo, mayor volumen existente de evidencia para ese tipo de estudio. El color rojo suele usarse para estudios de confiabilidad baja, el naranja para confiabilidad media y el verde para confiabilidad alta. Además, el gris suele usarse para evaluaciones de impacto y el azul para estudios que tengan diseñado un protocolo para la realización futura de una revisión sistemática. El tamaño de un mapa de brechas de evidencia puede variar ampliamente en la medida en que se incluyan más intervenciones y más variables de resultado en sus correspondientes ejes.

Tiene la ventaja de proveer una apreciación general o resumida de la evidencia de calidad dado que se identifica de manera visual y clara dónde hay suficiente evidencia y dónde hay vacíos, mostrando lo que se sabe y no se sabe de los efectos de determinadas intervenciones. Además son útiles para hacedores de política a la hora de tomar decisiones estratégicas con un mayor grado de conocimiento sobre el tema en cuestión y para investigadores y donantes que buscan desarrollar agendas de investigación en la búsqueda de evidencia rigurosa.

Por otro lado, poseen la desventaja de no proporcionar muchas luces o respuestas completas al problema en vista de que no se comparan resultados con otros estudios, ni ofrecen evidencia sobre los factores que promueven u obstaculizan el buen desenvolvimiento de la intervención. Su foco se centra, en esencia, en proveer evidencia en torno a la efectividad de manera rápida, visual y simple desde una perspectiva general. Sirve como un buen inicio o complemento a otras síntesis de mayor profundidad.

A continuación se presenta un ejemplo de mapa de brecha de evidencias extraído de la plataforma del International Initiative for Impact Evaluation (3ie). En el eje vertical se presentan distintos tipos de intervenciones dirigidas a niños, mientras que en el eje horizontal se expresan variables resultado típicamente de interés en temas infantiles.

Figura n° 1: Ejemplo de “evidence gap map” para intervenciones orientadas a niños



Fuente: International Initiative for Impact Evaluation (3ie)

2. “Rapid evidence assesments” o revisiones rápidas de evidencia: no existe una definición formal para las revisiones rápidas de evidencia. Según Tricco, et al (2015) “una revisión rápida es una forma de sintetizar conocimiento o evidencia en la que se usan de manera estratégica los componentes de la revisiones sistemáticas de forma simplificada, omitiendo algunos pasos, debido a restricciones de tiempo”. En otras palabras, es una herramienta para sistematizar la evidencia sobre una política o intervención de manera tan exhaustiva como sea posible con restricciones de tiempo. Estas restricciones de tiempo suelen ser impuestas por las demandas de los hacedores de política, en plazos que oscilan de 2 a 6 meses. Bajo esta metodología, los temas y las preguntas de investigación son más focalizados o específicos que los cubiertos por las revisiones sistemáticas. A continuación presentamos algunas características distintivas de las revisiones rápidas de evidencia:

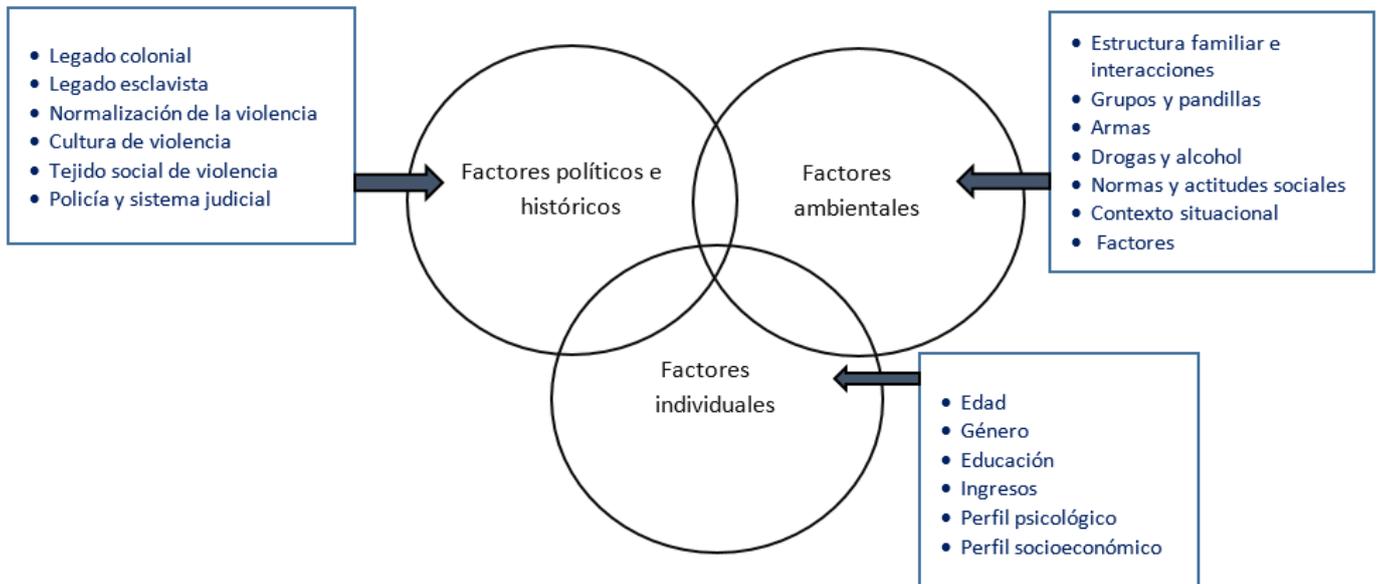
- Buscan seguir los principios de una revisión sistemática pero su alcance es menor, en cantidad y exigencias de rigurosidad.
- Están siendo crecientemente empleadas usando terminologías algo diferentes; entre ellas las llamadas “revisiones rápidas” desarrolladas mayormente en Europa para temas de salud y para sintetizar evidencia de intervenciones.
- Utilizan métodos de búsqueda de información similares a las revisiones sistemáticas pero de manera menos exhaustiva (menor número de fuentes).
- Suelen incluir un protocolo de búsqueda y criterios de inclusión y exclusión.
- Suelen incluir criterios de calidad pero menos rigurosos. Por ejemplo, no usan dos revisores por estudio.
- Presentan síntesis cuantitativas y cualitativas pero no incluyen meta-análisis.

Del mismo modo, presentamos un ejemplo ilustrativo y esquemático de una revisión rápida de evidencia orientada a comprender los determinantes de los niveles de violencia y criminalidad en Sudáfrica, titulado “Why is crime in South Africa so violent?”, acompañado de un diagrama que muestra los factores políticos, históricos, sociales y demográficos identificados para esa problemática en específico:

Tabla n° 2: Ejemplo de revisión rápida de evidencia sobre los determinantes de la violencia en Sudáfrica. Davies, et al (2011)

Objetivo/problema	Entender las causas de los altos niveles de violencia y criminalidad en Sudáfrica.
Encargado por	Presidente de la República de Sudáfrica, año 2011.
Justificación	Necesidad de entender las causas de la violencia para diseñar políticas para combatirla. Sudáfrica tiene una de las tasas más altas de crimen y violencia en el mundo.
Resultados	Se reportan síntesis de causas del crimen violento en el país como una combinación (sin meta análisis) de factores político-históricos (i.e. Colonia y apartheid) que “normalizan” la violencia, y producen desconfianza en las leyes y en las autoridades. Factores del contexto social: familias fragmentadas, prevalencia de pandillas y grupos irregulares violentos, acceso a drogas, alcohol y armas, ideología de género, etc. Factores individuales: edad, género, nivel educativo y socioeconómico del individuo y de los padres, acceso a servicios básicos, etc. En base a estas revisiones se reportan resultados mostrando variaciones por regiones del país en términos de cartelización y patrones de crimen diferenciados de acuerdo a los factores sociales e individuales

Figura n° 2: Ejemplo de revisión rápida de evidencia: ¿Por qué el crimen en Sudáfrica es tan violento? Explicaciones teóricas



Fuente: Davies & Macpherson (2011)

Una síntesis rápida de evidencia se considera un buen punto de partida para una revisión sistemática. Entre sus fortalezas más importantes están que contemplan el uso de protocolos especificando criterios de búsqueda e inclusión y que persiguen responder una pregunta de política específica, lo cual las hace extremadamente útiles para “policy makers”. No necesariamente es el caso para académicos e investigadores, los cuales suelen aproximar estos temas desde un enfoque más amplio. Por otro lado, poseen las deficiencias de no ser exhaustivas y de presentar una elevada heterogeneidad en términos de rigurosidad metodológica por el hecho de carecer de una metodología o procedimiento claramente definido.

3.

“Scoping reviews” o revisiones de alcance de la evidencia: mientras que una revisión sistemática típicamente se centra en una pregunta de investigación claramente definida, una revisión de alcance busca determinar si existe evidencia suficiente sobre un determinado tema, que pueda, por ejemplo, justificar la realización de una síntesis rigurosa. Como en el caso de las síntesis rápidas, estas revisiones también contemplan el desarrollo de un protocolo pero donde las búsquedas de información no sean tan exhaustivas y no se analice el contenido del material encontrado, sino su amplitud y cobertura. Se consideran un buen primer paso para la realización de una revisión sistemática, en especial cuando no es del todo claro si existe suficiente evidencia rigurosa para responder a la pregunta de investigación; pero también porque permiten afinar las preguntas de investigación.

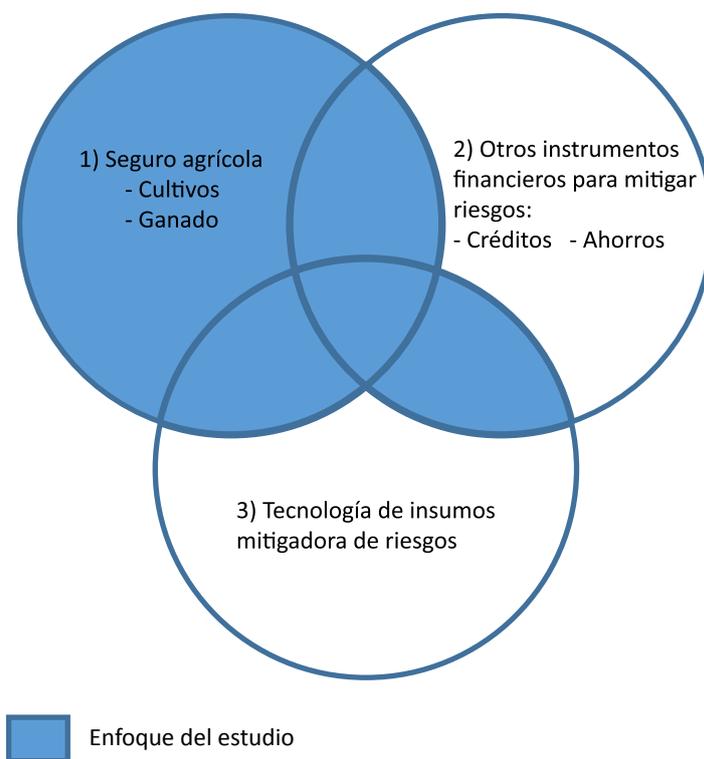
Entre sus debilidades tenemos que no posee una metodología claramente definida, lo cual implica que su nivel de rigurosidad es variable. Además, a veces aplicar la metodología más rigurosa puede aumentar el nivel de complejidad y volverse una carga de trabajo excesiva. La idea es que la revisión de alcance se restrinja a ser un paso previo y funcione como un insumo para realizar una revisión sistemática. A continuación se presentan una serie de pasos que son aplicables a cualquier revisión de alcance:

- a) Identificar qué se necesita explorar. Pensar en un enfoque más amplio que el de una revisión sistemática.
- b) Plantear las preguntas de la investigación en base a las necesidades identificadas.
- c) Buscar estudios y evidencia relevante de manera sistemática a través de mecanismos usuales como bases electrónicas, “snowballing”², páginas web, etc.
- d) Seleccionar estudios de acuerdo a criterios de inclusión (PICO) y de calidad.
- e) Sintetizar la información mediante cuadros descriptivos y tablas.
- f) Comparar, concluir y producir respuestas más generales que las de una revisión sistemática

De la misma manera, se presenta como ejemplo una revisión de alcance producida por 3ie (Barooah et al, 2017) que sintetiza exclusivamente evaluaciones de impacto asociadas a intervenciones orientadas a mitigar sólo algunos aspectos del riesgo financiero agrícola; concretamente lo relacionado al aseguramiento agrícola para el manejo cosechas y ganado. El documento provee un mapeo de la situación actual de la problemática para identificar los vacíos de conocimiento a través de la delimitación clara de los tipos de intervenciones a ser analizados, la formulación de una teoría de cambio comprehensiva para estas intervenciones (que será revisada como ejemplo más adelante), el establecimiento de criterios de inclusión y calidad para dichos estudios y la realización de un mapa de brecha de evidencias.

² El “snowballing” o muestreo de bola de nieve es una técnica de muestreo no probabilístico utilizada por los investigadores para identificar a elementos potenciales en estudios en donde son difíciles de encontrar. Es un mecanismo en cadena y se utiliza si la muestra para el estudio es muy rara o si está limitada a un sub-grupo muy pequeño de la población.

Figura n° 3: Ejemplo de revisión de alcance de evidencia: análisis de intervenciones para mitigar el riesgo agrícola



Fuente: Barooah, et al (2017)

De esta manera se puede observar cómo cada una de las revisiones mencionadas en este capítulo de alguna manera contiene a la anterior, es decir, van aumentando en profundidad y complejidad. La tabla n°3 muestra este aumento de complejidad y detalle a medida que se sube en la escala jerárquica de las revisiones o síntesis rigurosas, mientras que la tabla n°4 muestra una tipología esquemática en base a sus características.

Tabla n° 3: Orden ascendente de las metodologías rigurosas de síntesis

Término		Pasos involucrados
1	Revisión narrativa	Revisión de literatura realizada de forma no sistemática mínimamente extraída para responder a una pregunta de investigación amplia que puede estar vagamente planteada
2	Mapa de brechas de evidencia	Agrega: preguntas explícitas, búsqueda sistemática de evidencia y presentación de tablas descriptivas resumen de los hallazgos
3	Revisión de alcance	Permite identificar si existe o no evidencia suficiente sobre un determinado tema que permita luego llevar a cabo una síntesis rápida o revisión sistemática. Permite afinar las preguntas de investigación.
4	Meta-análisis	Agrega: síntesis cuantitativa de la evidencia basada en la aplicación de herramientas estadísticas y econométricas al pool de estudios seleccionados y sus respectivos hallazgos.
5	Revisión sistemática	Se basa en una pregunta muy puntual. Agrega: evaluación de la calidad de la evidencia a través de protocolos y produce recomendaciones basadas en síntesis cualitativas de toda la evidencia.

Fuente: elaboración propia

Tabla n° 4: Resumen metodologías de síntesis rigurosas

Metodología	Tema amplio	Evidencia incluida			Búsqueda sistemática	Evaluación crítica	Análisis
		RSs	EIs	Otras			
Mapa de brechas	Sí	Sí	Sí	No	Sí	Sí	No
Revisión de alcance	Sí	Puede	Puede	Puede	Limitada	No	Sí
Revisión rápida	No	Puede	Sí	Puede	Limitada	Limitada	Sí
Revisión sistemática	No	No	Sí	Puede	Sí	Sí	Sí

Fuente: elaboración propia. RSs: Revisiones Sistemáticas, EIs: Evaluaciones de Impacto

Bibliografía de la sesión

- Deep, A., Thillairajan, A., John, P., & Mahalingam, A. (2015). Impact of Private Sector Participation on access and quality of services: systematic review of evidence from the electricity, telecommunications and water supply sectors. *Journal of Development Effectiveness*, Vol. 7, 64-89.
- Anderson, A. A. (2009). *The Community Builder's approach to Theory of Change: A practical guide to theory development*. New York : The Aspen Institute.
- Cirera , X., & Lakshman, R. (2014). *The impact of export processing zones on employment, wages and labour conditions in developing countries: systematic review*. Londres: International Initiative for Impact Evaluation (3ie).
- Guerrero, G., Cueto, S., Leon, J., & Zapata, M. (2013). *Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries*. *Journal of Development Effectiveness*.
- Ortiz , A., & Rivero, G. (2007). *Desmitificando la teoría de cambio*. Washington, DC: Pact Organization.
- Snilstveit, B. (2016). *The impact of education programmes on learning and school participation in low and middle-income countries*. Londres: International Initiative for Impact Evaluation (3ie) Systematic Review Summary 7.
- Vogel, I. (2013). *ESPA guide to working with Theory of Change for research projects*. Londres: Ecosystem Services for Poverty Alleviation (ESPA).
- Davies, P., & Macpherson K. (2011). *Why Is Crime In South Africa So Violent? A rapid evidence assessment*. Oxford Evidentia.
- Barooah, B., Kaushish, B., Puri, J. (2017). *Understanding financial risks forsmallholder farmers in low-and middle-income countries: what do we know and not know?* International Initiative for Impact Evaluation (3ie).

Sesión 3. Teoría de cambio

Objetivos de la sesión

- Establecer la utilidad de la teoría de cambio para las síntesis rigurosas de evidencia
- Definir conceptualmente a la teoría de cambio, sus características, utilidad práctica, etapas en su proceso de diseño y su relación con la investigación social aplicada.
- Resaltar diferencias existentes entre las herramientas de teoría de cambio y de marco lógico.
- Proporcionar ejemplos de teorías de cambio empleadas en intervenciones reales.

Preguntas de investigación

Todas las preguntas de investigación están constituidas por un conjunto de elementos comunes. Es posible caracterizarlas diciendo que poseen un verbo (i.e. medir, evaluar, estimar), consideran un objetivo (i.e. impacto, tamaño del efecto, efectividad, eficiencia, costo-efectividad, así las posibles explicaciones de estos resultados), se enfocan en una intervención relacionada a un tema, la cual puede ser específica (i.e. transferencias condicionadas, anemia, aprendizajes) o genérica (i.e. pobreza, salud, educación), busca comprender relaciones entre variables y delimita un universo o población de interés o de influencia.

Un diseño inteligente y coherente de las preguntas de evaluación debe pasar por la construcción de la lógica de la intervención que se pretende analizar o evaluar. Esto es, cómo las actividades realizadas mediante la intervención se van a traducir en cambios permanentes de largo plazo. Este paso es fundamental ya que distintas partes involucradas en la intervención pueden tener comprensiones divergentes sobre lo que se debe hacer y cuáles se espera que sean sus efectos. Procedemos, pues, a adentrarnos en el concepto de teoría de cambio.

Teoría de cambio

También llamada “ruta de cambio”, “teoría de acción” o “cadena causal”, es un método para planificar, desarrollar y evaluar intervenciones o programas, estudios primarios o revisiones sistemáticas. Constituye una herramienta para desarrollar soluciones a problemáticas complejas a través de una visión global de las partes requeridas para conseguir cierto resultado o cambio.

La teoría de cambio presenta las diversas relaciones entre las acciones y los resultados en los distintos niveles posibles de cambio. Muestra la ruta planteada para

conseguir cierto cambio. La figura nº4 muestra la jerarquía de cambios existentes en una teoría de cambio. La consecución de actividades y productos en las partes bajas permiten la consecución de objetivos finales o de largo plazo.

En el campo de la investigación, la teoría de cambio representa tanto un proceso como un producto de constante reflexión en el entendido de guiar el diseño de la investigación ya que permite identificar la problemática abordada considerando el contexto y los grupos de interés, a la vez que proporciona sustento a las preguntas de investigación. En materia de revisiones sistemáticas, la teoría de cambio ayuda a explicitar la historia o la narrativa detrás de las intervenciones que se quiere evaluar.

Figura nº 4: Distintos niveles de cambio: de actividades a resultados



Fuente: Anderson (2006)

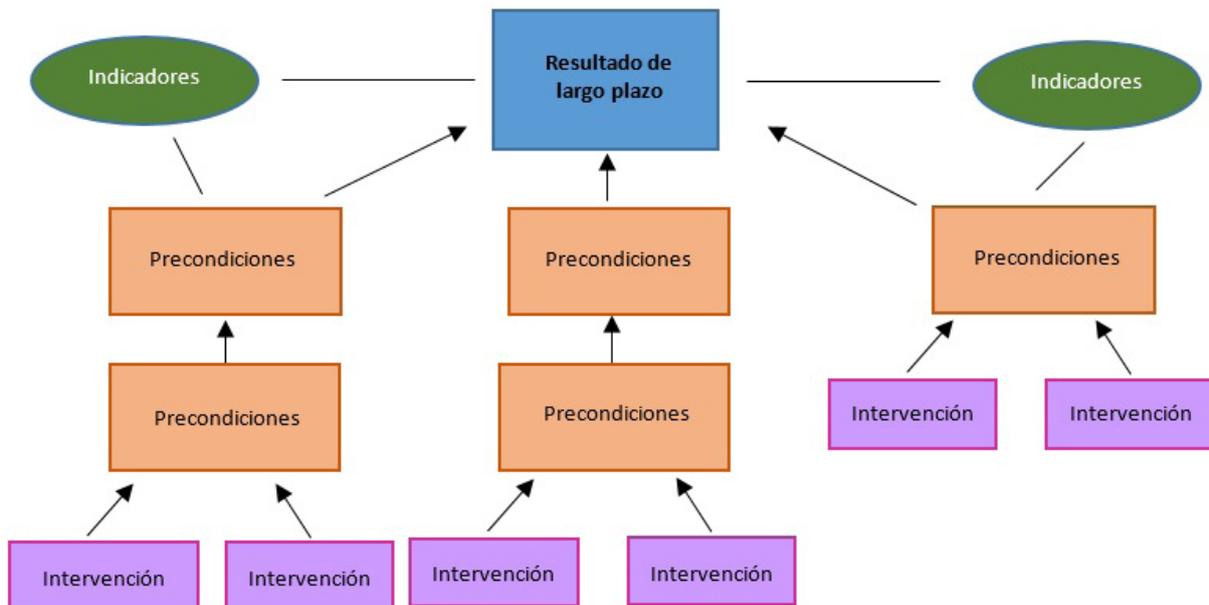
Elementos de una teoría de cambio

Toda teoría de cambio debe incluir un conjunto de elementos que se describen a continuación y pueden visualizarse en la figura nº5:

- **Visión de éxito:** propone una meta de largo plazo, que constituye el cambio más significativo del proceso (Ejemplo: empleo de largo plazo para adultos jóvenes que han sido desplazados por violencia política).
- **Precondiciones:** resultados necesarios (y suficientes) para conseguir la visión de éxito (Ejemplos: A) Jóvenes cuentan con un centro de información y capacitación. B) Contratos de prácticas pre profesionales firmados con empresas locales).
- **Intervenciones:** estrategias/intervenciones, simples o complejas, necesarias para lograr cada precondición. (Ejemplos: A) desarrollar un CV para capacitar a jóvenes en carreras técnicas de corto plazo. B) Coordinar con empresas locales para facilitar prácticas).

- **Supuestos:** situaciones dadas, no controlables, que influyen las precondiciones (Ejemplos: A) empresas locales desean contar con practicantes jóvenes. B) La economía en su conjunto mantiene un nivel de estabilidad y crecimiento).
- **Indicadores:** medidas para conocer el éxito o logro de cada precondición y de la visión de éxito (Ejemplos: A) tasa de jóvenes empleados: 10%. Meta: 90% se mantiene en su nuevo trabajo durante al menos 6 meses. B) número de jóvenes capacitados: Línea Base: 0 Meta: 250 jóvenes).

Figura n° 5: Elementos de una teoría de cambio



Fuente: Anderson (2006)

Utilidad de una teoría de cambio

La teoría de cambio facilita el entendimiento de una serie de procesos fundamentales en el diseño y la evaluación de intervenciones:

- Permite explicitar nuevas preguntas o áreas temáticas a considerar por los grupos de interés.
- Fuerza a cuestionar y reflexionar sobre las relaciones entre cada acción y resultado.
- Proporciona una mirada holística de los resultados que se necesitan para conseguir el cambio deseado.
- Facilita una adecuada asignación de recursos para alcanzar los resultados (precondiciones) identificados.
- Promueve el consenso y una comprensión conjunta de las metas entre los grupos de interés.

Etapas para el desarrollo de una teoría de cambio

La formulación de una teoría de cambio incluye un conjunto de etapas que son aplicables independientemente de la temática en particular. Entre estas, se incluyen:

1. Identificar una meta de largo plazo (visión de éxito) de los programas a incluir en la síntesis.
2. Conducir un “backwards mapping”, también llamado mapeo en reverso o hacia atrás, orientado hacia identificar los objetivos intermedios de las intervenciones y las acciones que las intervenciones proponen para lograr la meta de largo plazo.
3. Construir y validar indicadores para el objetivo final y los objetivos intermedios, que serán usados para evaluar el desempeño de distintas dimensiones de cada intervención.
4. Identificar los supuestos, riesgos o fenómenos exógenos asociados a las precondiciones.
5. Graficar, diagramar y esquematizar la teoría de cambio (TdC).
6. Escribir una narrativa que resuma las diferentes partes de la teoría de cambio, siguiendo una lógica de causalidad de efectos.

La teoría de cambio y las síntesis rigurosas de evidencia

La herramienta de teoría de cambio es muy importante a la hora de realizar síntesis rigurosas de evidencia debido a la necesidad imperante de tener claridad desde el comienzo del proceso de síntesis acerca del tipo de intervención que se pretende revisar. Esta noción proviene del hecho de que pueden existir numerosas intervenciones alrededor de un mismo tema o problemática implementadas en diversos entornos o contextos donde los mecanismos causales de cambio y de transmisión de efectos sean completamente diferentes o hayan sido diseñados

desde distintos enfoques o aproximaciones conceptuales. En tal sentido, es preciso acotar o delimitar claramente el alcance, tipo y mecanismos de cada intervención en base a lo expuesto en sus teorías de cambio respectivas, de manera de no confundir intervenciones que superficialmente se parezcan pero que desde el punto de vista de sus componentes o líneas de acción tengan diferencias sustantivas. Por tanto, la teoría de cambio funciona como una herramienta orientadora en la fase de selección de estudios y por ende debe formar parte central de todo el proceso de revisión de síntesis rigurosa. A continuación se presentan una serie de ejemplos de teorías de cambio aplicadas a problemáticas tan diversas como el desempeño escolar, la pobreza infantil o el aseguramiento agrícola.

Figura n° 6: Ejemplo de teoría de cambio. “Desayuno y rendimiento escolar”. Cueto (2001)

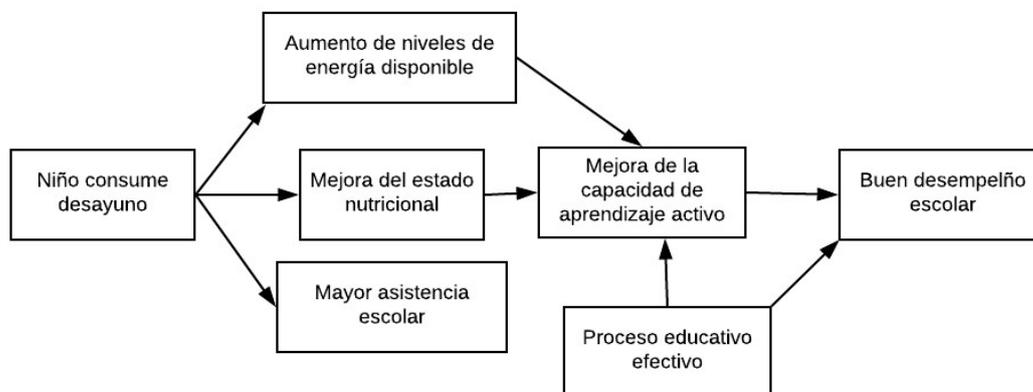
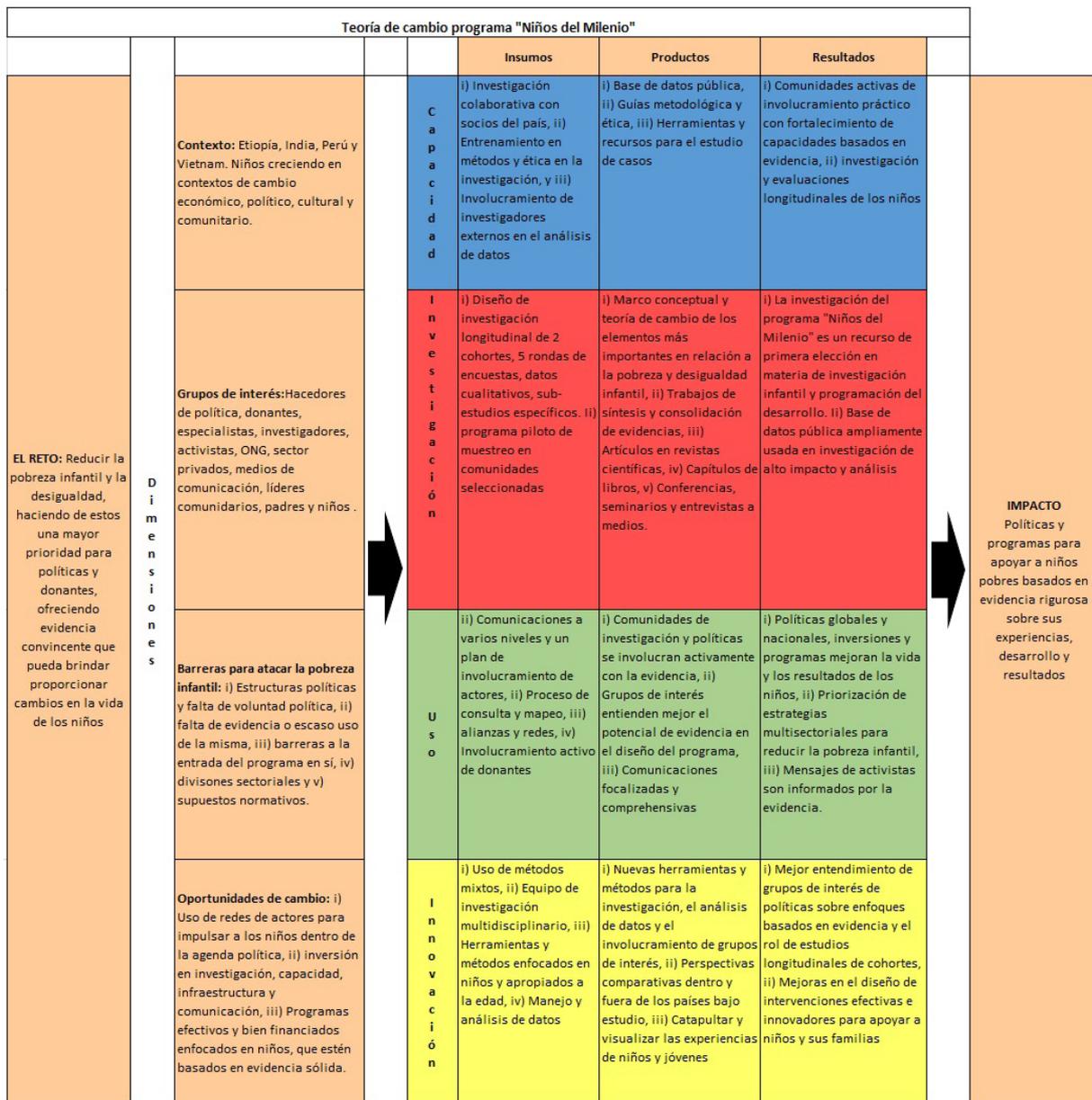


Figura n° 7: Ejemplo de teoría de cambio: Programa “Young Lives”



El marco lógico y su relación con la teoría de cambio

El marco lógico es una herramienta usada de forma ubicua en la planificación y gerencia de proyectos. Tiene indudablemente cierto parentesco con la teoría de cambio, aunque difieren en aspectos significativos. A grandes rasgos, se puede afirmar que la teoría de cambio es el argumento, visión o estructura visual sobre la cual se sustenta el marco lógico. En añadidura, el marco lógico no siempre recoge la totalidad de los matices o sutilezas que puede expresar la teoría de cambio de una organización. A continuación se presentan 2 tablas: la tabla n° 5, la cual muestra características tanto del marco lógico como de la teoría de cambio y la tabla n° 6, la cual da cuenta de sus diferencias en torno a varias dimensiones de interés.

Tabla n° 5: Características del marco lógico y la teoría de cambio

Marco lógico	Teoría de cambio
Características	
Herramienta para el análisis lógico, usada sobre todo en planeamiento de proyectos.	Enfatiza todas las relaciones entre los niveles de resultados.
Instrumento de planificación que identifica distintos elementos en un proceso de cambio.	Toma en consideración teorías y estudios previos.
Desarrolla la meta, el propósito, los resultados, las actividades y los insumos del proyecto.	Reconoce la participación de diferentes actores y su influencia en la ruta del cambio.
Presenta indicadores de medición, fuentes de verificación y supuestos.	Propone un pensamiento a nivel de sistemas

Fuente: elaboración propia

Tabla n° 6: Diferencias entre el marco lógico y la teoría de cambio

Dimensión	Marco lógico	Teoría de cambio
Plazos	Ideal para proyectos de corto plazo con objetivos, indicadores y resultados claramente definidos. Funciona mejor cuando existe una lógica lineal simple, en donde las actividades causan resultados. Es poco flexible o adaptable a cambios que aparecen sobre la marcha.	Refleja de forma explícita el conocimiento a profundidad del contexto, el cual guía las estrategias de intervención. Favorece estrategias flexibles y diversas y puede ser usada para rastrear la contribución a los procesos de cambio de alta complejidad. Brinda indicadores de largo plazo para medir cambios sociales en el tiempo.
Indicadores de éxito	Preferible de aplicar usando indicadores cuantitativos basados en actividades o productos específicos y resultados de corto plazo.	Hace uso de indicadores de éxito de corto y largo plazo, cuantitativos y cualitativos, dentro del marco de una ruta de cambio. Monitorea cambios en el comportamiento, actitudes y funciones.
Niveles de cooperación interinstitucional	Se centra en resultados específicos de una sola organización. No es muy útil para identificar resultados compartidos ni para valorar la contribución de relaciones de colaboración en el logro de resultados de largo plazo.	Promueve la identificación de resultados compartidos y permite que diferentes actores dentro de los grupos de interés puedan planificar y realizar seguimiento de sus contribuciones y acciones colaborativas.
Rendición de cuentas	Funciona mejor como una herramienta para informar “hacia arriba” a quienes financian la intervención y para monitorear si una organización ha llevado a cabo las actividades planteadas originalmente. Este suele construirse para fines internos, por lo que no es muy útil para informar a actores clave.	La TdC está firmemente arraigada al diálogo con las partes relevantes en torno a los resultados compartidos. Facilita dar cuenta de información que refleja el aprendizaje compartido.
Comparabilidad	Se implementa mejor en organizaciones aisladas con actividades, productos y resultados muy específicos; lo que reduce su comparabilidad en términos de desempeño con otras organizaciones similares.	Permite a las organizaciones controlar su capacidad para influir en el cambio. Puede producir información que permite comparar la efectividad de distintos enfoques y organizaciones.

Fuente: elaboración propia

Bibliografía de la sesión

- Deep, A., Thillairajan, A., John, P., & Mahalingam, A. (2015). Impact of Private Sector Participation on access and quality of services: systematic review of evidence from the electricity, telecommunications and water supply sectors. *Journal of Development Effectiveness*, Vol. 7, 64-89.
- Anderson, A. A. (2009). *The Community Builder's approach to Theory of Change: A practical guide to theory development*. New York : The Aspen Institute.
- Cirera , X., & Lakshman, R. (2014). *The impact of export processing zones on employment, wages and labour conditions in developing countries: systematic review*. Londres: International Initiative for Impact Evaluation (3ie).
- Guerrero, G., Cueto, S., Leon, J., & Zapata, M. (2013). *Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries*. *Journal of Development Effectiveness*.
- Ortiz , A., & Rivero, G. (2007). *Desmitificando la teoría de cambio*. Washington, DC: Pact Organization.
- Snilstveit, B. (2016). *The impact of education programmes on learning and school participation in low and middle-income countries*. Londres: International Initiative for Impact Evaluation (3ie) Systematic Review Summary 7.
- Vogel, I. (2013). *ESPA guide to working with Theory of Change for research projects*. Londres: Ecosystem Services for Poverty Alleviation (ESPA).

Sesión 4. Búsqueda sistemática de literatura, identificación de fuentes y sistemas de generación de mecanismos de búsqueda

Objetivos de la sesión

- Introducir mecanismos de búsqueda sistemática de literatura científica pertinentes para la realización de revisiones sistemáticas y otras síntesis rigurosas de evidencia
- Identificar fuentes relevantes y sistemas de información relacionados a la búsqueda sistemática de literatura científica.

Consideraciones generales

La búsqueda rigurosa de literatura para las revisiones sistemáticas y otras síntesis rigurosas de evidencia es un proceso que se da luego de la delimitación de la pregunta de investigación y que permitirá obtener un cierto número de estudios relacionados a ésta. En fases posteriores se procederá a la extracción y análisis de la información proveniente de los estudios incluidos. A diferencia de lo que ocurre en una revisión de literatura libre o convencional, la búsqueda para la revisión sistemática requiere otro tipo de consideraciones pues lo que se busca es eliminar, en la medida de lo posible, los 'sesgos de búsqueda. Así, las revisiones de literatura convencionales no requieren la documentación del proceso de búsqueda y se efectúan a partir de bases de datos comunes provenientes de bibliotecas universitarias y repositorios de "papers" como Google Scholar. Las revisiones sistemáticas, en cambio, requieren reportar todas las formas de búsqueda empleadas y conocer las diferentes fuentes de datos relevantes, por lo que el equipo de trabajo suele incluir especialistas en ciencias de la información. La generación de estos reportes de búsqueda será esencial al momento de evaluar la rigurosidad de una revisión sistemática u otro tipo de síntesis rigurosa.

Es preciso aclarar que el desarrollo de estrategias de búsqueda es más difícil en las ciencias sociales que en las ciencias médicas, donde este tipo de revisiones son más comunes y antiguas. Esto se debe a que en las ciencias sociales pocas veces existen consensos alrededor de los conceptos que definen un fenómeno, lo que genera un uso impreciso del lenguaje. Además, en muchos casos se presentan confusiones en la descripción metodológica expresada en los "abstracts" o resúmenes de los estudios, lo que dificulta la posterior selección de los mismos.

Figura n° 8: Flujograma del proceso de revisiones sistemáticas



Fuente: Lane Medical Library

Cómo realizar búsquedas rigurosas

A continuación se presenta una serie de pasos a seguir cuando se desea emprender una búsqueda de literatura en el marco de revisión sistemática o síntesis rigurosa de evidencia:

- 1. Seleccionar la pregunta de investigación y delimitar su ámbito y alcance:** Antes de empezar con las búsquedas, es necesario tener las preguntas claramente definidas. Las características que deben tener las preguntas de investigación ya han sido revisadas en la sesión anterior. Para fines de búsqueda, las preguntas pueden clasificarse en primarias y secundarias. Un ejemplo de pregunta primaria de índole cuantitativo podría ser: ¿Cuáles son los efectos de programas de capacitación vocacional y emprendimiento sobre la participación laboral femenina en puestos de empleo mejor remunerados, las condiciones laborales y/o el empoderamiento económico? Por su parte, un par de ejemplos de preguntas secundarias, serían: ¿Cuáles son los factores que obstaculizan y los factores que facilitan la efectividad de los programas de capacitación vocacional y emprendimiento? O ¿Cuáles son las características de diseño e implementación de programas asociados a la capacitación vocacional y el emprendimiento para mujeres?
- 2. Construir las cadenas de búsqueda:** Luego de definir las preguntas, se deben de seleccionar los términos o palabras claves que ayudarían a responderlas. Para ello es recomendable utilizar un tesoro especializado

en ciencias sociales (<http://www.ilo.org/thesaurus/defaultes.asp> y <http://vocabularies.unesco.org/browser/thesaurus/es/index/>) donde se busca la terminología adecuada para responder a nuestra pregunta de investigación. Luego, seleccionamos las palabras clave encontradas en los tesauros usados. Algunos ejemplos:

- **Participants:** woman, female, wife, mother, gender, occupational stratification, empowerment businesswoman, woman industry.
- **Employment:** employment, employability, employer, employee, work, workforce, job, vocation, career, occupation, livelihood, workplace, part-time, casual, informal, wages, labor market, school-to-work.
- **Intervention:** skill, upskill, life skills, training, retraining, mentor, apprentice, internship, prestige, expertise, professional, qualification, education, coaching, leadership, entrepreneur, human capital, business management, business support, capacity building.
- **Setting:** transitional, low-income, middle-income, third world, developing, underdeveloped, underserved nations/countries/populations/economies; LMIC, LAMI, low GDP, low GNP.

Dependiendo de los idiomas en los cuales se propone hacer las búsquedas habrá que definir cadenas similares de términos en otros idiomas.

3. Seleccionar las fuentes: A partir de la terminología recabada, usamos las diferentes fuentes que podamos identificar con la ayuda de especialistas. Estas pueden ser principalmente cuatro (4):

- Bases de datos de acceso privado o “paid- access databases”.
- Bases de datos de acceso público o “open- access databases”.
- Búsqueda manual en “journals” seleccionados o “hand search: selected journals”.
- Búsqueda manual de literatura gris³ o “hand search: gray literature”.

Considerando la clasificación anterior, la tabla n°7 expone algunos títulos de diversas fuentes de información como “journals” científicos, repositorios de archivos y bases de datos a las que se podría acudir en el caso del ejemplo planteado anteriormente de programas de entrenamiento y capacitación a mujeres.

³ La literatura gris puede involucrar cualquier tipo de documentos que no hayan sido publicados en journals o libros, como documentos de trabajo, tesis universitarias, etc.

Tabla n° 7: Clasificación de fuentes de información para la búsqueda sistemática de estudios sobre entrenamiento laboral femenino

Bases de datos de acceso privado	Bases de datos de acceso libre	Búsqueda manual: revistas seleccionadas	Búsqueda manual: literatura gris
1. Web of Science	1. Ideas-REPEC	1. American Economic Review	1. International Labor Organization (http://www.ilo.org)
2. Scopus	2. Campbell Library	2. Quarterly Journal of Economics	2. World Bank (http://www.worldbank.org)
3. ASSIA	3. 3ie Impact Evaluation Repository	3. American Economic Journal: Applied Economics	3. ELDIS (http://www.eldis.org)
4. IBSS	4. Labourdoc	4. Journal of Labour Economics	4. DfID R4D (http://r4d.dfid.gov.uk)
5. PAIS		5. Journal of Development Economics	5. ASEAN (http://www.asean.org)
6. ERIC		6. Journal of Development	6. IADB (http://www.iadb.org)
7. Econlit		7. Economic Development & Cultural Change	7. UNGEI (http://www.ungei.org)
8. Academic Search Complete		8. International Labour Review	8. UN-Women (http://www.unwomen.org)
9. Business Source Corporate		9. World Bank Economic Review	9. J-PAL (http://www.povertyactionlab.org)
10. Gender Studies Database		10. World Development	10. IPA (http://www.poverty-action.org)
		11. Journal of Development Studies	11. British Library for Development Studies (http://blds.ids.ac.uk/)
		12. Journal of Human Resources	12. World Bank DIME (http://www.worldbank.org/en/research/dime)
		13. Journal of Education and Work	13. World Bank enGENDER Impact (https://bit.ly/2IYKNKn)
		14. Journal of International Women's Studies	14. World Bank Policy Research Working Paper Series (http://documents.worldbank.org/curated/en/docsearch/document-type/620265)
		15. Gender & Development	
		16. Feminist Economics	
		17. Feminist Review	

4. Usar las cadenas de búsqueda en cada base de datos tomando en cuenta sus particularidades: A continuación presentamos un conjunto de “tips” o buenas prácticas al momento de realizar las cadenas de búsqueda:

- **Introducir palabras clave.** Usar sustantivos, verbos y nombres propios. De preferencia usar palabras compuestas. Por ejemplo: “menor de edad” en vez de “menor”. Emplear sinónimos de las palabras elegidas, o variantes gramaticales.
- **Introducir términos relacionados de significado parecido, o del mismo campo semántico.** Por ejemplo, según el tesoro de la OCDE, la palabra “malnutrición”, tiene como términos relacionados: “carencia vitamínica”, “deficiencia calórica”, “deficiencia proteínica”, “enfermedades diarreicas”, “enfermedades por carencia”, “escasez de alimentos” y “hambre”.
- **Utilizar diferentes tipos de operadores.** Estos se pueden clasificar en:
 - a. **Operadores lógicos:** “AND”, “OR” y “NOT”. “AND” Localiza documentos que contengan cada término de la búsqueda. Ejemplo: “public health AND tuberculosis, labour market AND women AND public sector”. “OR” busca por lo menos una de las palabras que aparece en el documento. Ejemplo: “mercado de trabajo OR mercado laboral, women OR woman OR female”. Por su parte, “NOT” excluye los documentos con el término indicado. Ejemplo: “public education NOT rural education”.
 - b. **Operadores de truncamiento:** (?), (*) y (#). (?) Aplica truncamiento al final de la palabra y se utiliza para recuperar todos los sufijos de una misma raíz. Ejemplo: “farm?” recupera “farm”, “farmer”, y “farmers”. Por su parte, (*) aplica truncamiento en la parte media de la palabra y se utiliza para plurales en inglés y variaciones idiomáticas. Ejemplo: “wom*n” recupera “women” y “woman”. Finalmente, (#) se utiliza cuando la palabra contiene un carácter adicional. Ejemplo: “lab#r” recupera “labor” y “labour”.
 - c. **Operadores de proximidad:** “WITH” (W) y “NEAR” (N). Se utilizan para concretar las búsquedas, pues permiten indicar lo cerca que están unos términos de otros. Cuanto más cerca estén, tendrán mayor relación temática. “WITH” (W) indica que el orden de las palabras empleadas debe mantenerse. Puede contener un número para especificar el número de palabras. Ejemplo: “cambio W5 climático” recupera documentos donde aparece primero el término “cambio” seguido del término “climático” si están a 5 palabras o menos de cercanía en el orden en que se introdujeron. Por su parte, “NEAR” (N) indica que no es necesario conservar el orden de las palabras. Es decir, se usa de la misma manera que “WITH” (W), solo que no importa el orden de las palabras buscadas.
 - d. **Anidamiento:** Se usa para unir con paréntesis distintas

estrategias de búsqueda. Permite agrupar o anidar expresiones de búsqueda utilizando paréntesis. Las bases de datos buscan primero la expresión que se encuentra en los paréntesis internos, y posteriormente continúan la búsqueda, moviéndose hacia los extremos de la expresión. Ejemplo: ((cambio climático AND recursos hídricos) AND (Perú OR Bolivia OR Ecuador)).

e. **Comillas:** Busca frases o conjunto de dos o más términos que tengan que ir juntos y en un orden específico. Ejemplo: “Cambio climático”.

- **Utilizar filtros.** Estos permiten refinar la búsqueda de información. Se puede filtrar por textos completos o parciales, trabajos revisados por pares, idioma, período de publicación y/o área temática.
- **Utilizar la combinación de los términos con los distintos operadores.** Para esto sugerimos revisar los mensajes de ayuda. No todas las bases utilizan el mismo conector lógico. Por ejemplo: el conector “AND” puede ser reemplazado por el “*” (asterisco) en varios casos.
- **Traducir los términos al inglés.** En la medida de lo posible, trabajar siempre con el inglés.
- **Refinar la búsqueda en función de los resultados** y siempre realizar más de una búsqueda.
- **Posterior a la realización de búsquedas.** Se deben guardar las cadenas de búsquedas, así como los archivos de los “papers” en el formato predeterminado por cada base de datos donde se buscó (formato RIS, por ejemplo). Finalmente, importar los resultados en el programa o software para su codificación.

Uso de bases de datos Vs. Google Scholar

Las desventajas comunes de usar Google Scholar están relacionadas sobre todo a la abundante cantidad de información que dificulta su análisis y selección y a la confiabilidad de la información recuperada. Al no explicitar los criterios a partir de los cuales Google Scholar identifica la literatura no se pueden descartar los sesgos de búsqueda. Asimismo, se ha encontrado evidencia de la manipulación de datos e indicadores bibliométricos de herramientas proporcionadas por Google. Existen experimentos donde se han creado perfiles de investigadores con documentos inexistentes, los cuales han citado otros documentos que han ido incrementando su número de citas. Por esto, se recomienda usar bases de datos especializadas en los temas que se estén trabajando. En la tabla n° 8 se presentan algunas diferencias entre la búsqueda de literatura en bases de datos especializadas y en Google Scholar.⁴

⁴ Existen programas informáticos especializados como Eppi-Reviewer que facilitan considerablemente el registro de los procesos de búsqueda de información y su posterior análisis.

Tabla n° 8: Diferencias entre el uso de base de datos especializadas y Google Scholar

Bases de datos especializadas	Google Scholar
Existe un proceso de selección de las revistas a indexar	Indexa todo documento proveniente de un dominio académico o científico
Las revistas seleccionan los trabajos a publicar	Cualquier documento indexado emite y recibe citas
Existe un control de enlace entre citante y citado	No es posible denunciar comportamientos fraudulentos o errores en el enlazado de citas
Se persiguen y penalizan comportamientos fraudulentos	

Bibliografía de la sesión

- Ho, G et al (2016). Development of a search strategy for an evidence based retrieval service. Youngstown: Youngstown State University.
- Moncada-Hernández, S. (2014). Cómo realizar un búsqueda de información eficiente: foco en estudiantes, profesores e investigadores en el área educativa. Investigación en Educación Médica, 3(4), 106-115.
- Universidad de Alcalá. (2016). Estrategias de búsqueda de información. Obtenido de: http://www3.uah.es/bibliotecaformacion/BMED/AlfaBuah/13_estrategias_de_bsqueda_de_informacin.html.

Sesión 5. Criterios de Inclusión y Exclusión de Pautas Para la Evaluación Crítica de Estudios

Objetivos de la sesión

- Entender cómo establecer criterios de inclusión y exclusión para la evaluación crítica de los estudios.
- Conocer que herramientas y capacidades son necesarias para hacer una evaluación crítica.
- Conocer los pasos que se requieren para hacer una evaluación crítica de los estudios.

La evaluación crítica de estudios

La evaluación crítica de estudios representa una parte esencial de toda revisión sistemática de literatura. Es el proceso por el cual se utilizan criterios explícitos y transparentes para incluir y excluir estudios en la fase de síntesis. Estos criterios suelen variar en función de la naturaleza de los estudios, ya sean cualitativos o cuantitativos, y su objetivo principal es separar estudios de alta calidad de los de baja calidad de forma sistemática.

Es importante resaltar que los criterios mediante los cuales se va a regir la síntesis rigurosa deben hacerse explícitos en un documento o “protocolo” antes de emprender la tarea investigativa. Más aún, este protocolo debe construirse a partir del consenso entre investigadores, asesores y potenciales usuarios. Mínimamente, la decisión de inclusión o exclusión se debe basar en al menos dos personas o codificadores.

Criterios “PICO”

La selección de estudios requiere maximizar el volumen y la relevancia mediante un proceso de búsqueda eficiente. Para encontrar estudios con alta relevancia relativa, estos deben ser seleccionados, filtrados y codificados a partir de criterios comúnmente denominados “PICO”: “Population”, “Intervention”, “Comparisons”, y “Outcomes”.

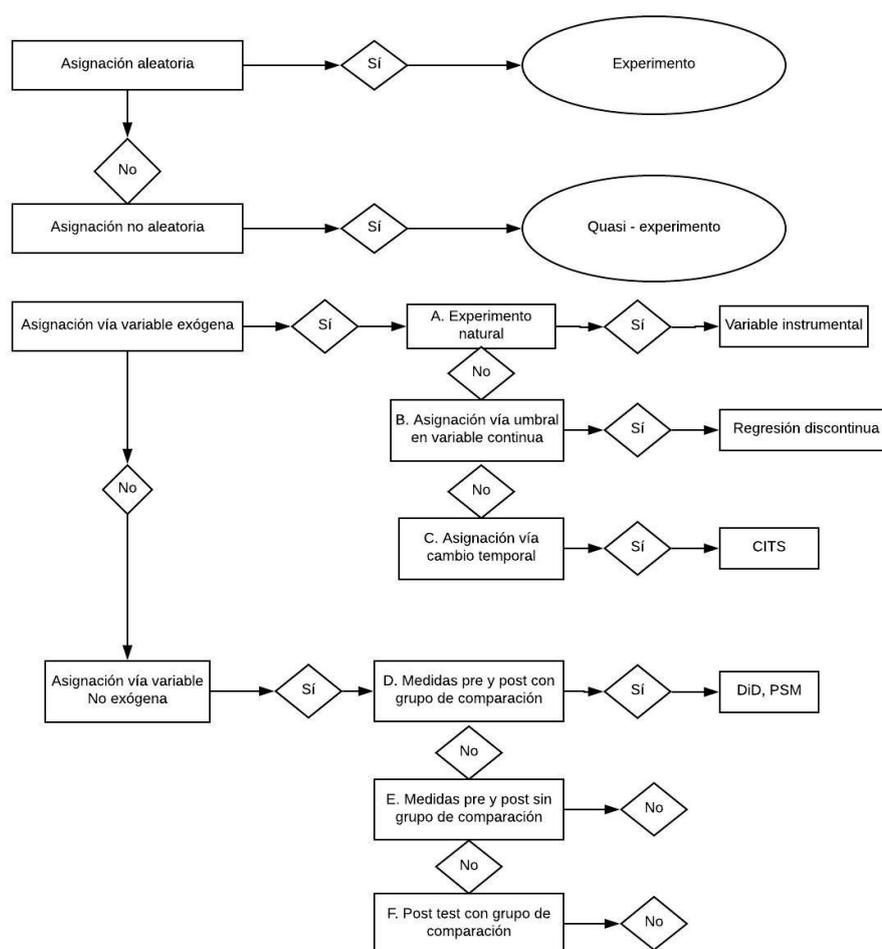
- **“Population”**: el primer criterio es el tipo de población. Esta podría dividirse por sexo (hombres, mujeres, hombres y mujeres), grupos de edad (menores o mayores de 18 años, niños, adultos, ancianos) o grupos con determinadas características o atributos (situación laboral, ingresos, nivel socioeconómico o educativo, etc.). Se presentan un par de ejemplos:

- Ejemplo 1: “Capacitación profesional y comercial para aumentar la participación de las mujeres en ocupaciones de alto nivel en países de bajos y medianos ingresos”. Población: mujeres, 18 o más, que antes de entrar al programa podía estar empleadas o desempleadas en países de bajo y mediano ingreso según la definición del Banco Mundial del 2016
 - Ejemplo 2: “¿Cuál es el impacto de los programas educativos en el aprendizaje y la participación escolar de los niños?” Población: Niños(as) en edad escolar de primaria y secundaria en educación general en países de bajos y medianos ingresos.
- **“Intervention”**: la priorización de estudios ocurre en base a lo descrito originalmente en la teoría de cambio en cuanto al objetivo de las intervenciones. Se deben describir en detalle los objetivos de dichas intervenciones, así como sus diferentes componentes y actividades. Así como es importante determinar qué tipo de intervenciones serán incluidas, también lo es para las intervenciones que serán excluidas. Más aún, se recomienda dar algunos ejemplos de intervenciones que serán excluidas y sus razones. Algunos ejemplos:
 - Ejemplo 1: “Capacitación profesional y empresarial para aumentar la participación de las mujeres en ocupaciones de alto nivel en países de bajos y medianos ingresos”. Intervención: Programas de capacitación vocacional y formación empresarial. Suponiendo que el programa intervenía hombres y mujeres, el estudio fue incluido sólo si los efectos de los hombres eran reportados por separado o si el 80% de los beneficiarios eran mujeres.
 - Ejemplo 2: “¿Cuál es el impacto de los programas educativos en el aprendizaje y la participación escolar de los niños?” Intervención: enfoque en distintas líneas de intervención asociadas a variables de éxito y desempeño escolar.
- **“Comparisons”**: consiste en comprender mejor la calidad metodológica del estudio a partir de qué manera se conformó el grupo de comparación (si existe) que se usó para estimar determinados efectos causales. Algunos ejemplos:
 - Ejemplo 1: “Formación profesional y empresarial para...” Comparación: No intervención, o intervenciones con diferentes niveles de intensidad.
 - Ejemplo 2: “Cuál es el impacto de los programas educativos...” Comparación: No intervención, o intervenciones educativas de distintos tipos (enfocados a nivel de alumnos, hogares, escuelas, maestros, de currículo y de sistema).
- **“Outcomes”**: consiste en evaluar qué variables finales o de resultado deben incluirse en la revisión. Este proceso parte por listar todas las variables de resultado, discriminando entre resultados intermedios y finales, para luego definir cuáles se incluyen en el criterio de selección y cuáles no.

- Ejemplo 1: “Capacitación profesional y comercial para aumentar la participación de las mujeres en ocupaciones de alto nivel en ingresos de países de bajos y medianos ingresos (ingresos, o salarios)”. Resultado: solo se incluyeron estudios que lograron medir ingresos salariales de mujeres a nivel individual a través de registros administrativos y datos auto-reportados. Se excluyeron estudios que se aproximaron a la variable de ingreso a través de encuestas de consumo a menos que estas diferenciaron el nivel de gasto o consumo por sexo. Variables: ingreso, salario real, condiciones laborales, brecha salarial, diferencias salariales de género, empoderamiento económico femenino, etc.

- **Otros criterios:** pueden usarse otras dimensiones para filtrar estudios relevantes como el año de publicación del estudio, el idioma de publicación, el contexto o entorno donde ocurrió el estudio y su calidad metodológica. Con respecto al último criterio, en la figura nº9 se exponen jerarquías o niveles de deseabilidad desde el punto de vista metodológico de estudios cuantitativos.

Figura nº 9: Jerarquía de métodos para la estimación de efectos causales



Fuente: elaboración propia

Jerarquía de métodos cuantitativos para la estimación de efectos causales

1. **Diseños experimentales:** son aquellos que usan asignación aleatoria a la intervención, es decir, individuos o clúster de individuos son asignados de manera aleatoria a los grupos de intervención y de comparación (o grupo de control). Los estudios con diseños experimentales son considerados los más “rigurosos” para determinar relaciones causales. Sin embargo, para determinar el nivel de rigurosidad del estudio es importante evaluar cómo se implementó la asignación así como la validez interna del estudio en base a otros criterios descritos más adelante.

- Ejemplo 1: Individuos interesados en capacitación vocacional, son asignados de manera aleatoria al grupo de intervención (3 meses de capacitación vocacional) o al grupo de control (0 meses de capacitación).
- Ejemplo 2: Escuelas primarias ubicadas en una misma región, son asignadas de manera aleatoria al grupo de intervención (recibir paquete de materiales educativos y capacitación en prácticas pedagógicas) o al grupo de control (no reciben nada).

2. **Diseños cuasi-experimentales:** contemplan asignación no aleatoria pero utilizan métodos que pueden abordar de manera creíble el sesgo de selección, dado que la asignación a las condiciones es por medio de auto selección. Se pueden jerarquizar a su vez en 2:

- a. Uso de reglas de asignación conocidas, tales como asignación a partir de un umbral en una variable continua o variación geográfica en la asignación del programa. Ejemplos: diseños de regresión discontinua (RDD) o experimentos naturales.
- b. Uso de mediciones de la variable de resultado de interés antes (pre) y después (post) de que se inicie la intervención. Entre estos métodos podemos encontrar al modelo de series temporales interrumpidas o “Comparative Interrupted Time Series (CITS)”, modelos de diferencia en diferencias (DiD), de preferencia en combinación con métodos de emparejamiento como “propensity score matching (PSM)”, método de variables instrumentales, o modelo de selección de Heckman.

Se suelen excluir estudios que no hayan llevado a cabo un levantamiento de información de línea de base o medición previa a la intervención de la variable de resultado de interés. Por tal razón es que se observa una línea roja en la figura n° 11. Por debajo de la línea, el estudio no es lo suficientemente riguroso como para aseverar con confianza que se está identificando un efecto causal o, en el caso de intervenciones, una evaluación de impacto.

Validez de los estudios cuantitativos

Los estudios seleccionados pueden variar en calidad. Por ello, es importante evaluar la validez de los estudios en torno a diferentes aspectos:

- **Validez interna:** Hace referencia a la validez de las inferencias sobre si la relación entre el tratamiento y la variable de resultado representa efectivamente una relación causal.
- **Validez de la conclusión estadística:** Hace referencia a la validez de las inferencias sobre la correlación y el grado de correlación entre el tratamiento, la variable de resultado y los supuestos identificadores del modelo.
- **Validez externa:** Hace referencia a la generalización o validez de las inferencias de la muestra del estudio a una población más amplia (diferentes personas, situaciones, contextos geográficos y variables de resultado).
- **Validez de los constructos:** Hace referencia a la validez de las inferencias obtenidas a partir de individuos particulares en la muestra a los constructos de orden superior que representan (personas, tratamiento, variables de resultado, o contextos).

Resumen del proceso multietápico o de embudo

El proceso de selección, revisión y clasificación de estudios para la elaboración de síntesis rigurosas de evidencia se puede desglosar en 3 etapas bien definidas:

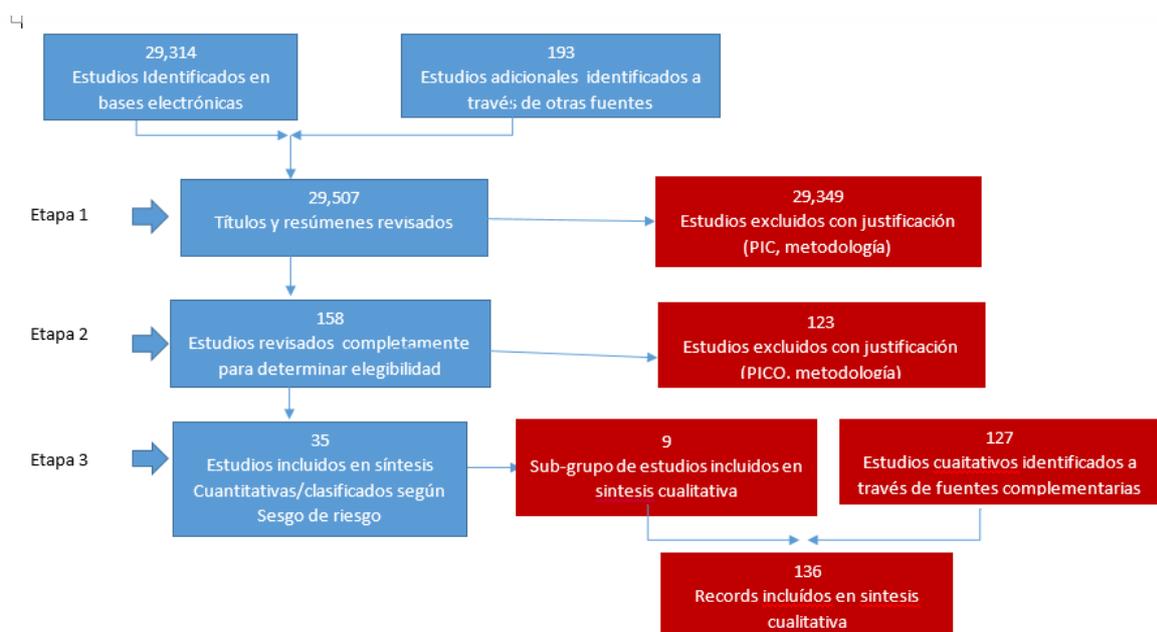
- **Etapas 1.** Selección de los estudios en base a títulos y resúmenes (abstracts): esto implica filtrar en base a “PIC” y metodología empleada de manera inclusiva. No es recomendable filtrar en base a “outcomes” o variables de resultado en esta etapa. Al menos 2 codificadores deben clasificar los estudios de manera independiente. Los codificadores pueden ser bachilleres propiamente entrenados para filtrar los estudios en base a los criterios de inclusión y de exclusión. Idealmente, los codificadores entienden de métodos cuantitativos de investigación. Los desacuerdos en la codificación se deben resolver mediante discusión y supervisión de un investigador senior. Los investigadores senior deberán supervisar el proceso, idealmente muestreando un subgrupo de estudios incluidos y excluidos que validen la codificación. Para esta etapa es recomendable usar el software EPPI.
- **Etapas 2.** Lectura completa de los estudios que pasaron la etapa 1: esto involucra filtrar en base a PICO y metodología empleada. La población y la Intervención hacen referencia a las características de la intervención(es), mientras que la comparación, las variables de resultado, la metodología, el idioma y el año de publicación hacen referencia a las características del estudio de evaluación de impacto. Con los estudios que pasaron la etapa 2 se debe llenar la información de PICO en una hoja de cálculo o Excel para crear una base de datos con información básica de los estudios seleccionados. Del mismo modo, al menos 2 codificadores deben clasificar cada estudio de manera independiente. Los codificadores deben tener

sólida formación en métodos y conocimiento del contenido temático. Idealmente, los codificadores tienen estudios de post-grado y son parte del equipo de investigadores senior. Como de costumbre, los desacuerdos en la codificación se deben resolver mediante discusión consensuada

- **Etapa 3.** Clasificación de los estudios que pasaron la etapa 2 según sesgo de riesgo (Risk bias analysis). Involucra un proceso mediante el cual los estudios que pasaron la etapa 2 son codificados y clasificados según riesgo en diferentes tipos de sesgo, usando un formato estructurado. Luego se utiliza dicha clasificación para efectuar análisis de sensibilidad en la fase de meta-análisis. El formato codifica los estudios según los siguientes tipos de sesgo:

- **Sesgo de selección:** estrategia de identificación del impacto del programa, equivalencia de los grupos de tratamiento y de comparación, y evaluación de riesgo de atrición diferencial.
- **Sesgo de performance:** efectos derrame (spillovers) que potencialmente contaminen el grupo de control.
- **Sesgo de reporte:** ocurre a nivel de variables de resultado y análisis, particularmente en el uso de variables de control que sean endógenas, métodos inusuales de análisis, y/o ausencia a la hora de reportar resultados no significativos.
- **Otros sesgos:** tales como sesgo de cortesía y de deseabilidad social (courtesy and social desirability bias), fuerte participación del equipo de investigación en la implementación de la intervención, efectos tipo “Hawthorne” y “John Henry”.
- **Otros aspectos:** Nivel de agregación o agrupación de los datos (clustering), tamaño de las muestras, poder estadístico, etc.

Figura n° 10: Ejemplo de proceso multietapa



Ejemplo: Chinen, et. al (2017)

A continuación se propone un conjunto no excluyente de preguntas que se pueden plantear a la hora de mapear los riesgos para luego ponderar los distintos tipos de sesgo de cada estudio:

- ¿Están definidas las variables principales de resultado?
- ¿Están definidas las variables secundarias de resultado?
- ¿Están definidas las variables explicativas y de control?
- ¿Cuáles son los métodos usados para la recolección de datos?
- ¿Cuál es la frecuencia de recolección de datos?
- ¿En qué nivel se realizó la asignación al grupo de tratamiento y control?
- ¿El estudio muestra valores de línea de base de los resultados de interés (como se definen en el protocolo) para beneficiarios y no beneficiarios?
- Si los valores basales del resultado de interés no están disponibles en la línea de base, ¿muestra el estudio valores basales de características de beneficiarios y no beneficiarios que no son susceptibles de ser afectados por la intervención?
- ¿Son los valores promedio o las distribuciones de las covariables al inicio del estudio estadísticamente diferentes entre los beneficiarios y los no beneficiarios?
- Si existen diferencias estadísticamente significativas entre los beneficiarios y los no beneficiarios, ¿se controlan estas diferencias como covariables en los análisis de evaluación del impacto?

Consideraciones finales

La definición de criterios de inclusión y la posterior selección de estudios no son procesos sencillos. Existe una gran cantidad de factores a considerar en ambas fases y la abundancia de información puede resultar abrumadora para el recopilador de datos o investigador. En este sentido, es realmente útil clasificar y esquematizar los estudios con la ayuda de tablas resumen de elementos clave como en la tabla n°9.

Tabla n° 9: Hoja resumen “PICOs”

Característica del estudio	Criterio de inclusion
Población	
Intervención	
Comparación	
Outcomes, variables de resultado	
Metodología, Tipo de estudio	
Período	
Idioma	

Bibliografía de la sesión

- Cook, T., Shadish, W., & Wong, V. (2008). The conditions under which experimental & observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis & Management*, 27(4), 724-750.
- Waddington, H. (2012). How to do a good systematic review of effects in international development: a toolkit. *Journal of Development Effectiveness*, Vol. 4, 359-387.

Sesión 6. Extracción de datos: meta-análisis (análisis cuantitativo)

Objetivos de la sesión

- Entender el concepto y las condiciones necesarias para realizar un meta-análisis.
- Proporcionar una mirada introductoria sobre la medición del tamaño del efecto en un meta-análisis, así como los criterios metodológicos que se deben cumplir.
- Comprender cómo se examinan e interpretan los resultados.
- Brindar herramientas para verificar la existencia de sesgo de publicación de los efectos usados en el meta-análisis

La presente sesión pretende profundizar en temas de análisis cuantitativo aplicado específica y exclusivamente a la elaboración de revisiones sistemáticas de evaluaciones de impacto con altos estándares de rigurosidad metodológica. En este sentido, se requiere que el lector posea un conocimiento mínimo de lo que implica realizar una evaluación de impacto en relación a otros tipos de evaluaciones que poseen requerimientos metodológicos y de datos considerablemente más flexibles.

Objetivos del meta-análisis

Es un conjunto de técnicas analíticas que tienen como objetivo:

- Sintetizar los resultados de diferentes estudios y variables de resultado que sean comparables. Por ejemplo, se podría sintetizar los resultados de diferentes estudios que investiguen el impacto de programas educativos sobre el rendimiento escolar.
- Resumir el proceso de la revisión sistemática a través de un estimador estadístico que contenga el impacto promedio obtenido a partir de varios estudios.
- Dar una solución a la incertidumbre que puede ocurrir por la existencia de resultados de impacto distintos en los estudios que superaron la etapa 2.

Cálculo del tamaño de efecto

El meta-análisis se centra en obtener y calcular el “tamaño del efecto” (TA) estandarizado o “standardized effect sizes” de cada estudio y de cada variable de resultado. El TA estandarizado es la unidad de estudio del meta-análisis y se

mide en unidades de desviación estándar de la variable de resultado de interés. El TA es una medida de la magnitud de la relación o efecto de una variable sobre otra (Cohen, 1988). En el caso de las evaluaciones de impacto experimentales o cuasi-experimentales, es la diferencia en la variable de resultado entre el grupo de tratamiento y grupo de control. El TA se puede calcular usando la fórmula de Cohen's d o Hedges' g. En algunos casos, el TA es reportado directamente en el estudio. De lo contrario, se debe calcular o contactar a los autores para obtener los insumos que permitan calcularlo.

Si el estudio bajo revisión reporta resultados para diferentes variables de resultado, se debe estimar el TA para cada variable de resultado. Bajo la misma lógica, si el estudio presenta múltiples tratamientos, se debe estimar el TA para cada tratamiento. Luego, si el estudio presenta diferentes tipos de análisis estadísticos, se debe identificar la estimación "final" y estimar su TA o hacerlo para cada modelo presentado. Para cada TA, se procede a calcular el intervalo de confianza (IC) con 95% de certeza: El IC brinda información sobre la precisión del TA: a mayor IC, menor precisión. Cuando se promedian los TA de diferentes estudios, el meta-análisis cuantitativo pondera cada TA en función del tamaño de la muestra del estudio.

Los tipos de TA varían de acuerdo al tipo de variable de resultado que se esté analizando. En el caso de revisiones sistemáticas aplicadas a evaluaciones de impacto, el interés está en el tamaño de efecto para variables continuas. Así, los TA más usados para este tipo de variables son la diferencia simple o la diferencia estandarizada de los promedios. La manera de calcular ambos se presentan en el siguiente cuadro:

Tabla n° 10: Tamaños de efecto más usados en estudios de investigación

Fórmula	
Diferencia simple de los promedios	$Y_{Tratamiento} - Y_{Control}$
Diferencia estandarizada de los promedios	$\frac{Y_{Tratamiento} - Y_{Control}}{S_{pooled}}$

Como se ha mencionado, el que se usa para meta-análisis es la diferencia estandarizada de los promedios porque este permite tener los diferentes tamaños de efecto en la misma métrica (desviaciones estándar); aspecto que permite realizar las comparaciones entre los diferentes estudios. Sin embargo, se cuenta con diferentes tipos de efectos estandarizados y que son usados al momento de realizar un meta-análisis. Estos efectos se muestran en el siguiente cuadro. La principal diferencia entre los tres está en la desviación estándar que se usa para la estandarización de la diferencia promedio. Se aprecia que tanto el d-Cohen (Cohen, 1962) y el g-Hedges (Hedges, 1981) usan diferentes formas de combinar

la desviación estándar de ambos grupos, mientras el Δ -Glass usa la desviación estándar del grupo de control dado que esta no estaría afectada por el tratamiento (Glass, McGraw, & Smith, 1981).

En cuanto a cual TA usar, se recomienda usar el g-Hedges dado que no asume que las varianzas de ambos grupos de estudio son iguales y pondera las varianzas de acuerdo al tamaño de cada grupo, a diferencia del d-Cohen. Finalmente, el Δ -Glass no es recomendable dado que no se cuenta con una clara especificación de porque se emplea la desviación estándar del grupo de control para estandarizar el efecto promedio de un programa (Richardson, 1996: 14).

Tabla n° 11: Tamaños de efecto más usados en evaluaciones de impacto

Tipo	Tamaño de Efecto	Desviación Estándar
d - Cohen	$d = \frac{\bar{X}_T - \bar{X}_C}{S_{pooled}}$	$S_{pooled} = \sqrt{\frac{(s_T^2 + s_C^2)}{2}}$
g - Hedges	$g = \frac{\bar{X}_T - \bar{X}_C}{S_{pooled}}$	$S_{pooled} = \sqrt{\frac{(s_T^2(n_T - 1)) + (s_C^2(n_C - 1))}{(n_T + n_C - 2)}}$
Δ - Glass	$\Delta = \frac{\bar{X}_T - \bar{X}_C}{S_{Control}}$	$S_{Control} = \frac{\sum(X_i - \bar{X}_C)}{n - 1}$

Para calcular el TA, se debe extraer la siguiente información de los estudios seleccionados:

- La media para cada grupo.
- La desviación estándar para cada grupo.
- El número de observaciones para cada grupo.
- Los errores estándar asociados a cada grupo.
- La diferencia reportada como tamaño de efecto.
- Los errores estándar asociado a cada efecto.
- El "t" estadístico asociado a la diferencia.
- Los grados de libertad.
- La correlación entre la variable de resultado y el tratamiento.

Si bien es útil contar con la media, la desviación estándar y el número de observaciones para cada grupo (ver fórmulas en la tabla n° 10), también se pueden usar otros indicadores estadísticos que están disponibles en los estudios de investigación que pueden ser usados para el cálculo del tamaño del efecto, ya sea el d-Cohen o el g-Hedges.

Tabla n° 12: Fórmulas para calcular información para estimar d-Cohen o g-Hedges

Tipo	Tamaño de Efecto	Desviación Estándar
Error Estándar	Desviación estándar	$SD = SE \sqrt{n}$
Correlación	d-Cohen	$d = \frac{2r}{\sqrt{1-r^2}}$
	g-Hedges	$g = \frac{\frac{r}{\sqrt{1-r^2}}}{\sqrt{\frac{(n_T + n_C - 2)(n_T + n_C)}{(n_T n_C)}}}$
t-estadístico asociado a la diferencia	d-Cohen	$d = \frac{2t}{\sqrt{n_T + n_C - 2}}$
	g-Hedges	$g = t \sqrt{\frac{n_T + n_C}{n_T n_C}}$
Coefficiente de regresión no estandarizado	d-Cohen	$d = \frac{\beta}{SD_Y}$
g-Hedges	d-Cohen	$d = g \sqrt{\frac{N}{n_T + n_C - 2}}$

Nota: Para mayores detalles revisar los documentos de Nakagawa & Cuthill (2007) y Cooper (2009).

A continuación se presenta un ejemplo práctico donde se puede observar la aplicación de todas las fórmulas que se muestran, usadas para el cálculo del tamaño de efecto:

Ejemplo Práctico. En función de la media, desviación estándar y número de observaciones estime el tamaño del efecto de Cohen.

Se ha podido extraer la siguiente información (media, desviación estándar y número de observaciones) de cinco estudios que se presenta en el siguiente cuadro.

	Grupo Tratamiento			Grupo Control		
	Media	DE	N	Media	DE	N
Estudio 1	562	83	282	511	119	287
Estudio 2	539	80	290	464	82	290
Estudio 3	595	83	186	528	112	200
Estudio 4	516	92	95	477	116	88
Estudio 5	508	99	581	457	87	488

Una vez obtenida esta información, podemos estimar la diferencia estandarizada entre el grupo tratado y control usando el d de Cohen. Para esto aplicamos las siguientes formulas:

$$d = \frac{\bar{X}_T - \bar{X}_C}{S_{pooled}}, \text{ donde el } S_{pooled} = \sqrt{\frac{(s_T^2 + s_C^2)}{2}}$$

Apliquemos las fórmulas para obtener las diferencias promedio estandarizadas para los dos primeros estudios.

Estudio 1

$$S_{pooled} = \sqrt{\frac{(83^2 + 119^2)}{2}} = 103, \quad d = \frac{562 - 511}{103} = 0.50$$

Estudio 2

$$S_{pooled} = \sqrt{\frac{(80^2 + 82^2)}{2}} = 81, \quad d = \frac{539 - 464}{81} = 0.93$$

De esta manera, usando la información extraída de los estudios, se pudo calcular la media estandarizada de Cohen (d).

Combinando los Tamaño de Efecto (Pooled Effect Sizes)

Una vez estimado los TA para cada uno de los estudios, el siguiente paso es combinarlos. Para lograr este objetivo es necesario contar con el error estándar de cada TA estimado. Para esto es necesario contar con el número de observaciones para cada uno de los grupos de estudio (tratamiento y control), además del tamaño de efecto estimado, como se puede ver en la siguiente formula:

$$ee_d = \sqrt{\frac{(n_T+n_C)}{(n_T n_C)} + \frac{d^2}{2(n_T+n_C)}}$$

Ejemplo Práctico. Estimar el Error Estándar de cada Tamaño de Efecto

Usando el ejemplo mostrado anteriormente, vamos a calcular el error estándar para el primer estudio usando la siguiente formula:

$$ee_d = \sqrt{\frac{(n_T+n_C)}{(n_T n_C)} + \frac{d^2}{2(n_T+n_C)}}$$

Estudio 1

$$ee_d = \sqrt{\frac{(282+287)}{(282*287)} + \frac{0.50^2}{2(282+287)}} = 0.09$$

De esta manera, usando la información extraída de los estudios, se pudo calcular el error estándar para la media estandarizada de Cohen (d).

El siguiente paso es usar el error estándar estimado como un ponderador para combinar los TA de los diferentes estudios y obtener un tamaño de efecto promedio de la relación en estudio. A continuación presentamos las fórmulas que se emplean para combinar los diferentes TA estimados:

Tabla n° 13: Fórmulas para el cálculo del tamaño de efecto promedio

Indicador	Fórmula
Ponderador	$w = \frac{1}{SE_{TE}^2}$
Tamaño de efecto combinado	$\overline{TE} = \frac{\sum(w_i TE_i)}{\sum w_i}$
Error estándar del TE	$SE_{TE} = \sqrt{\frac{1}{\sum w_i}}$

Ahora, usemos las formulas vistas anteriormente para calcular el peso y el Tamaño de Efecto promedio para el ejercicio práctico que se viene realizando.

Estudio	TE (d)	EE (d)	Peso ($w=1/EE^2$)	Peso*TE
Estudio 1	0.50	0.09	137.98	68.59
Estudio 2	0.93	0.09	130.97	121.26
Estudio 3	0.68	0.10	91.12	61.93
Estudio 4	0.37	0.15	44.91	16.73
Estudio 5	0.55	0.06	255.73	139.95

Ahora para estimar el Tamaño de Efecto promedio hacemos lo siguiente:

$$\overline{TE} = \frac{68.59 + 121.26 + 61.93 + 16.73 + 139.95}{(137.98 + 130.97 + 91.12 + 44.91 + 255.73)} = 0.62$$

Luego, estimamos el error estándar asociado al Tamaño de Efecto promedio

$$SE_{TE} = \sqrt{\frac{1}{\sum w_i}} = \sqrt{\frac{1}{660.69}} = 0.04$$

De esta manera, el Tamaño de Efecto promedio es 0.62 y el error estándar es 0.04.

La forma visual y sintética de mostrar los tamaños de efecto de un conjunto de estudios es especificando, a través de filas, el valor del coeficiente estimado de una misma variable por cada estudio y su p-valor correspondiente, lo que no es más que la probabilidad de detectar un efecto cuando en realidad no lo hay. Por tanto, cuando el p-valor es menor a la regla de decisión definida, la cual suele ser 0,05 o 0,1, se dice que el coeficiente bajo estudio es estadísticamente significativo. En las figuras nº 11 y 12 se muestran resúmenes de tamaños de efecto para revisiones sistemáticas de evaluaciones de impacto de intervenciones dirigidas a incrementar la asistencia docente y su posterior efecto en el desempeño escolar de los alumnos.

Figura nº 11: Ejemplo de tamaño de efecto. Asistencia docente. Guerrero, et al (2013)

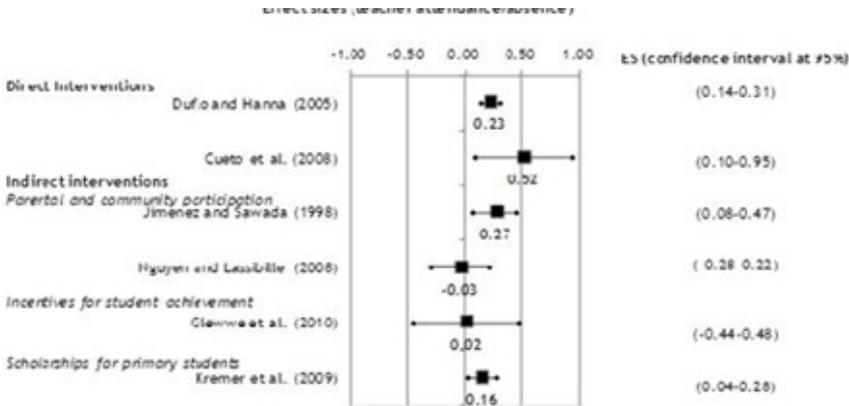
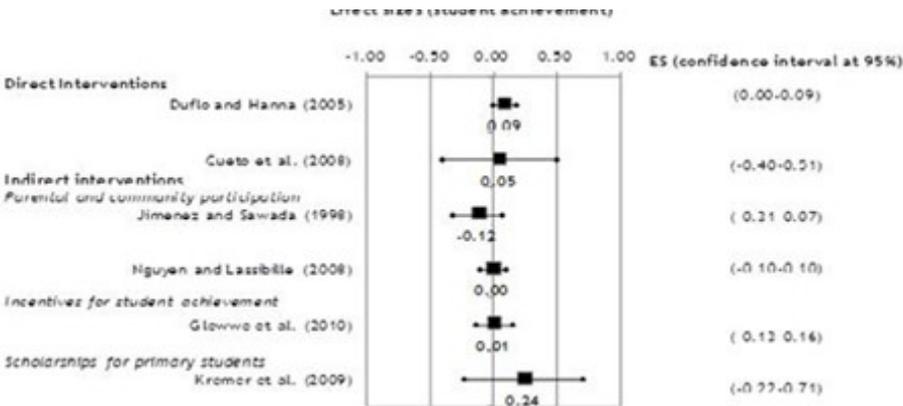


Figura nº 12: Ejemplo de tamaño de efecto. Logro escolar del estudiante. Guerrero, et al (2013)



Análisis de sensibilidad, heterogeneidad y “forest plot”

El análisis de sensibilidad involucra una actividad de comparación donde se contrastan diferentes tipos de diseños de investigación de acuerdo a la metodología empleada sobre un mismo tema o mismas variables de resultado; para ver cómo se distribuyen los hallazgos en términos de coeficientes de interés estimados. Una vez separados por método, se ponderan los posibles tipos de sesgos para determinar su nivel de riesgo de no reportar un efecto causal. Por su parte, el análisis de heterogeneidad contrasta diferencias en los estudios en términos de enfoque, componentes y aspectos contextuales de carácter institucional, geográfico, climático y cultural; los cuales tienen una incidencia en las estimaciones. El proceso pasa por clasificar los estudios sub-grupos a partir de su ámbito y región o zona geográfica de influencia.

Una herramienta visual realmente útil es el llamado “forest plot”. Este presenta de forma visual los resultados para cada estudio con sus correspondientes variables de resultado similares. Reporta los tamaños de efecto, intervalos de confianza, valores promedio ponderaciones o pesos. Las figuras n° 7 y 8 a continuación pueden usarse como un ejemplo de análisis de sensibilidad. Se especifica el nombre de cada programa vocacional cuyo estudio ha utilizado un diseño experimental y cuasi-experimental para medir el efecto de programas vocacionales de entrenamiento laboral sobre el nivel de empleo y se grafican conjuntamente los tamaños de efecto respectivos, su nivel de significancia y la ponderación aplicada en función de parámetros de calidad del estudio. Las figuras n° 9 y 10 presentan 2 ejemplos de análisis de heterogeneidad. En la primera se clasifican los estudios de programas vocacionales por regiones de influencia: África y Asia en comparación con el resto del mundo. En la segunda la clasificación se hace en función de si el programa contó con un enfoque de género o no.

Figura n° 13: Ejemplo de “Forest Plot”. Estudios experimentales - Efecto de programas de entrenamiento vocacional sobre el empleo

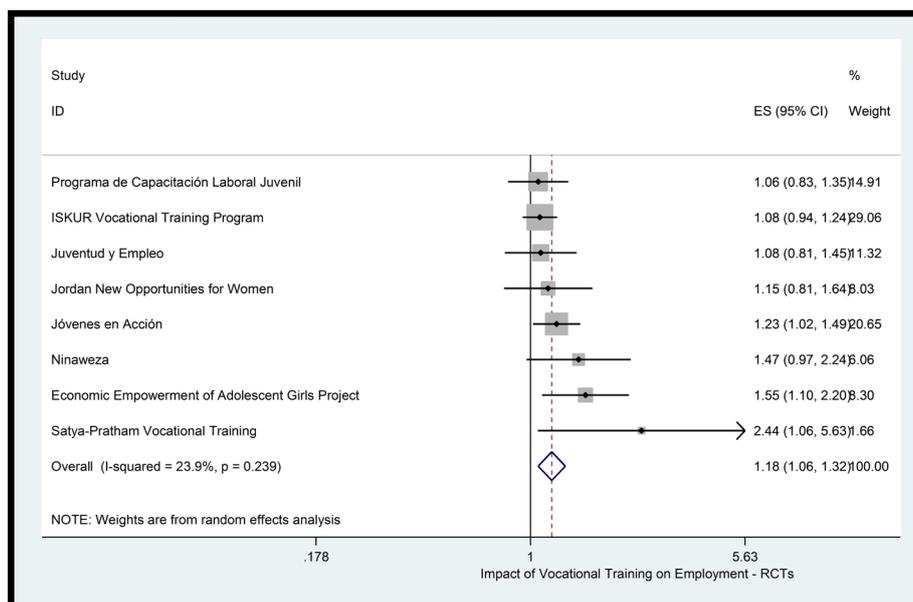


Figura n° 14: Ejemplo de “Forest Plot”. Estudios cuasi experimentales - Efecto de programas de entrenamiento vocacional sobre el empleo

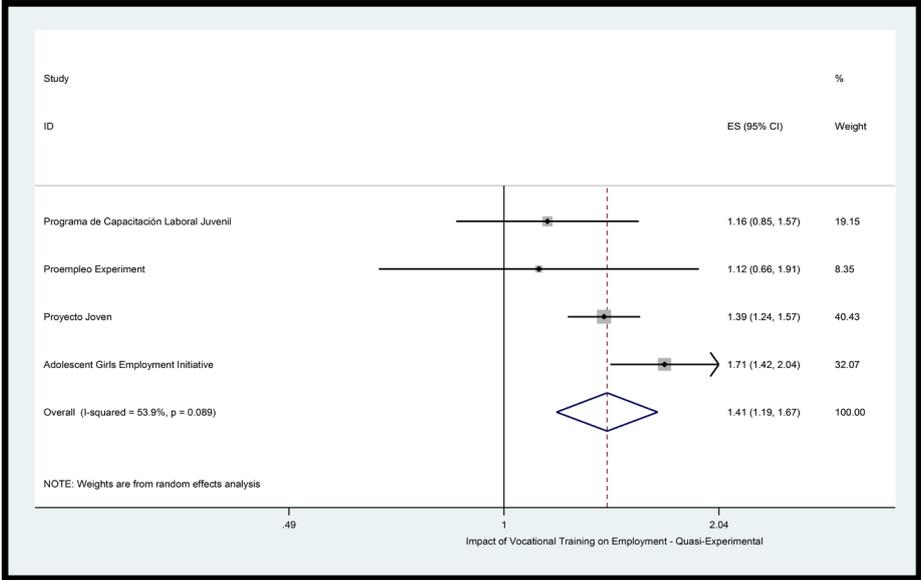


Figura n° 15: Ejemplo de análisis de heterogeneidad. Estudios experimentales - Programas vocacionales en África/Asia Vs. otros continentes

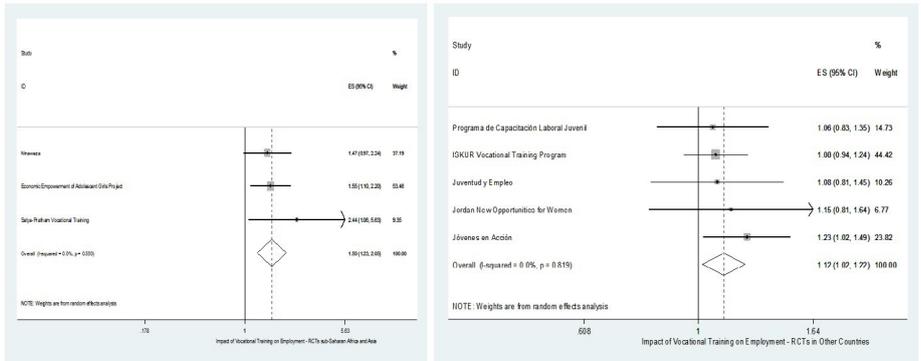
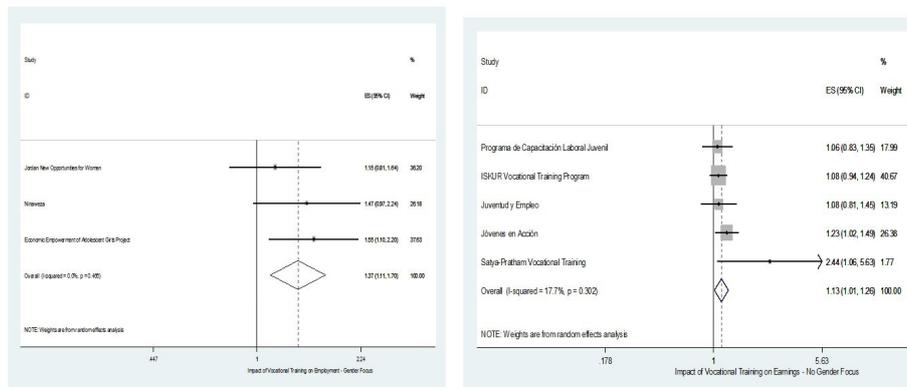


Figura n° 16: Ejemplo de análisis de heterogeneidad. Estudios experimentales – Programas vocacionales con enfoque de género Vs. sin enfoque de género



Sesgo de Publicación

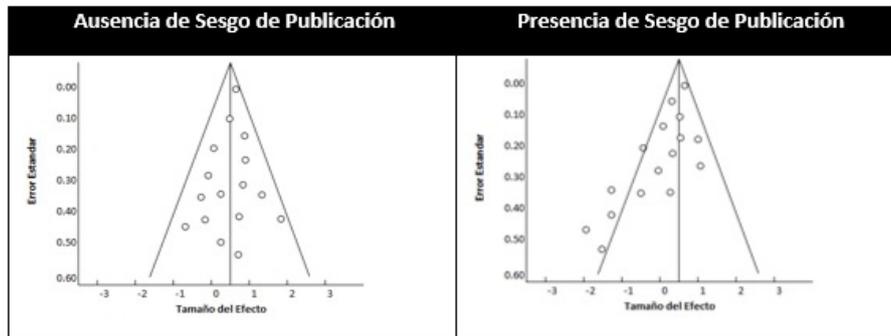
Una vez estimado el tamaño de efecto promedio del tipo de programa que se está estudiando, se tiene que verificar la existencia o no de sesgo de publicación en el grupo de estudios identificados para el meta-análisis.

El sesgo de publicación hace referencia a que la decisión de publicar un estudio está en función de sus resultados; es decir, la mayoría de revistas electrónicas tienden a publicar estudios que tienen efectos “significativos” y no aquellos estudios que no encuentran efectos (Rothstein et al, 2006). Así, la presencia de sesgo de publicación origina que se sobreestime el tamaño de efecto promedio del tipo de programa o intervención que se estudia. Por este motivo, es importante maximizar la búsqueda de los estudios que puedan ser parte de la revisión sistemática y posterior meta-análisis. Otras fuentes de datos para estudios de investigación pueden ser “google scholar”, publicaciones gubernamentales, repositorios de tesis universitarias, resúmenes de conferencias o congresos, entre otras.

Para identificar la presencia de sesgo de publicación, se debe usar tanto estrategias gráficas o visuales como test estadísticos. En el caso de las estrategias visuales, se usa el funnel plot. La idea es ordenar a los TA de cada estudio de acuerdo a su nivel de precisión que viene dado por su error estándar estimado. Para obtener el funnel plot, se hace un gráfico de dispersión donde en el eje de las Y está el nivel de precisión de cada estudio (p.ej.: tamaño de la muestra, error estándar) y en el eje de las X el TA. Se espera que se forme un triángulo simétrico al realizar el gráfico, donde en la parte inferior (o superior) se encuentren los estudios con mayor precisión (o muestras grandes) y en la parte superior (o inferior) los estudios con menor precisión (o muestras pequeñas).

El siguiente gráfico muestra casos hipotéticos donde no existe sesgo de publicación y donde se presenta el sesgo de publicación.

Figura n° 17: Funnel Plot para evaluar el sesgo de publicación



Bibliografía de la sesión

- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology*, 65, 145-153.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates
- Cooper, H. (2009). *Research synthesis and meta-analysis: A step-by-step approach*. 4th edition, Sage publications.
- Egger M, Davey Smith G, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*; 315:629-634
- Glass, G. V., McGraw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Beverly Hills: Sage Publications.
- Guerrero, G., Cueto, S., Leon, J., & Zapata, M. (2013). Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries. *Journal of Development Effectiveness*.
- Hedges, L. V. (1981). Distributional theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics* 6, 107-128
- Nakagawa, S., & Cuthill, I. C. (2007). Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biological reviews*, 82(4), 591-605.
- Richardson, J. T. (1996). Measures of effect size. *Behavior Research Methods, Instruments, & Computers*, 28(1), 12-22.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.

Sesión 7. Extracción de datos: meta-síntesis (análisis cualitativo)

Objetivos de la sesión

- Comprender las características de las síntesis rigurosas de estudios cualitativos y las diferencias con las de estudios cuantitativos.
- Identificar criterios y estrategias para evaluar la calidad, validez, confiabilidad, credibilidad e impacto de las síntesis rigurosas de estudios cualitativos.

Uso e importancia de las síntesis rigurosas de estudios cualitativos

Los tipos de estudios cuantitativos que se han abordado en esta guía se caracterizan por tener el propósito último de medir efectos causales o impactos atribuibles a alguna intervención o programa social. En otras palabras, estos estudios proporcionan información sobre si el programa en cuestión está generando el efecto esperado o no.

En contraposición, las síntesis rigurosas de índole cualitativo, también llamadas comúnmente síntesis narrativas, meta-etnografías o meta-síntesis cualitativas, se concentran en responder preguntas primarias o secundarias para saber por qué un programa funciona, en qué contextos y para quiénes.

Citando Snilstveit et al, (2012), “a diferencia de lo que ocurre con la síntesis cuantitativa, donde el objetivo es convertir las diferentes mediciones en una métrica común y sintetizar los datos para ‘testear’ una teoría a través del meta-análisis estadístico, la síntesis cualitativa busca sintetizar información cualitativa, que suele ser de naturaleza textual. Por tanto, estas revisiones abordan la síntesis desde un enfoque narrativo, y no estadístico, que busca generar nuevas ideas y recomendaciones tratando de ir más allá de los usuales resúmenes de hallazgos que caracterizan a las síntesis narrativas”. Al respecto, el “estándar de oro” que propone las evaluaciones aleatorizadas deja de ser útil y por tanto necesitamos otras formas de conocimiento que den cuenta del contexto específico de la intervención, su aproximación y sus mecanismos de implementación.

En añadidura, este tipo de síntesis se centra en la interpretación, revelando información sobre experiencias y percepciones de los beneficiarios, así como consecuencias inesperadas – positivas o negativas – de las intervenciones más allá de los resultados esperados.

Las síntesis rigurosas cualitativas son versátiles en el sentido de que incluyen estudios de distintos tipo. Entre ellos podemos encontrar a los estudios cualitativos -los cuales son claramente la fuente principal-, luego se tienen los estudios cuantitativos

no considerados evaluaciones de impacto -los cuales se usan como datos de corte narrativo-, luego existen los estudios con métodos mixtos y finalmente estudios sobre programas con información extraída de fuentes diversas como estudios incluidos y excluidos, páginas web institucionales, búsquedas en internet, entre otros.

La naturaleza altamente contextual de los hallazgos hace que estos no puedan ser simplemente extraídos y comparados. Por el contrario, estos deben analizarse sin perder de vista el contexto metodológico, conceptual y empírico en el que fueron generados. A diferencia de lo que ocurre en los estudios cuantitativos, las interpretaciones cualitativas están más explícitamente mediadas por las perspectivas de los investigadores.

Criterios y estrategias para evaluar la calidad de la investigación cualitativa

A continuación se presentan un conjunto de criterios, estrategias y herramientas que, al ser aplicadas a la investigación cualitativa en síntesis rigurosas de evidencia, pueden contribuir notablemente a incrementar su calidad, validez, confiabilidad y credibilidad, así como su potencial impacto en el diseño de políticas públicas basadas en evidencia:

Criterios

Las investigaciones deben mostrar los siguientes criterios en el reporte de la perspectiva teórica, metodológica y analítica:

- Transparencia
- Rigurosidad
- Coherencia interna
- Relevancia
- Comprensibilidad y verificabilidad intersubjetiva
- Relevancia del diseño para las preguntas de investigación
- Hacer explícito el sustento empírico (groundedness) de las ideas planteadas
- Utilización de códigos bien definidos, discutidos, consensuados
- Elaborar el análisis en grupos
- Ponderación y análisis de casos contradictorios o divergentes
- Fundamentación de las conclusiones en términos de la triangulación de datos, métodos y casos divergentes
- Capacidad explicativa de las interpretaciones presentadas
- Contener un apartado que tome en cuenta consideraciones éticas
- Hacer explícitas las dificultades, desafíos y vicisitudes que surgieron durante todo el proceso de investigación
- Capacidad reflexiva del investigador

Checklist “CASP”

Del mismo modo, se incluye una lista de chequeo de los elementos que deben estar presentes en toda síntesis cualitativa, la cual fue recuperada del “Critical Appraisal Skills Programme (CASP)”, una organización inglesa orientada a promover buenas prácticas de evaluación basadas en evidencia en el sector salud:

1. El propósito u objetivo de la investigación está claramente definido
2. La metodología cualitativa empleada es la apropiada
3. El diseño de investigación es apropiado para abordar el propósito de la investigación
4. La estrategia de reclutamiento es apropiada para el propósito de la investigación
5. El proceso de recolección de datos hace posible responder al problema de investigación
6. La relación entre el investigador y los participantes es considerada adecuadamente
7. Se hacen explícitas consideraciones éticas relacionadas al problema de investigación
8. El análisis de datos alcanza altos estándares de rigurosidad
9. Los hallazgos de la investigación están claramente definidos
10. La investigación y sus aportes son altamente valiosos

Estrategias

Según Snilstveit et al, (2012), una característica clave que diferencia a las diferentes estrategias es “el grado en que los diferentes métodos ponen a prueba, exploran o generan teorías y el grado en el que interpretan la evidencia surgida de los diferentes estudios. Las diferencias en los métodos se encuentran en algún punto de un continuo que va desde lo meramente “agregativo” -visto desde la perspectiva de los resúmenes-, a lo “configurativo” -donde el foco está en los temas comunes que se analizan en el proceso de configurar y generar una nueva teoría-”.

Se puede afirmar que este proceso es altamente similar al proceso de procesamiento de información cualitativa en general. Este implica la codificación, agregación, sistematización y análisis de una abundante masa de datos que se va progresivamente complejizando a través del surgimiento de elementos paralelos de un proceso iterativo y configurativo donde las categorías analíticas van evolucionando conforme se avanza en el proceso de investigación. En este sentido, se busca permanentemente “reducir” la cantidad de información sin perder valor contextual y complejidad analítica de tal manera que se produzca una interpretación que no solo plantee una descripción lo que hay, sino también la generación de nuevo conocimiento.

Bibliografía de la sesión

- Cook, T., Shadish, W., & Wong, V. (2008). The conditions under which experimental & observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis & Management*, 27(4), 724-750.
- Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative Approaches to Systematic Review and Synthesis of Evidence for International Development Policy and Practice. *Journal of Development Effectiveness*, 4, 409-429.
- Waddington, H. (2012). How to do a good systematic review of effects in international development: a toolkit. *Journal of Development Effectiveness*, Vol. 4, 359-387.

Sesión 8. Síntesis rigurosas y su incidencia en las políticas públicas

Objetivos de la sesión

- Relacionar el propósito de las síntesis rigurosas de evidencia con los objetivos de incidencia en el diseño de políticas públicas.
- Discutir ejemplos de síntesis rigurosas con profunda relevancia y alto impacto en incidencia llevadas a cabo tanto nacional como internacionalmente.
- Mostrar algunas estrategias comunicacionales, de incidencia y de involucramiento de actores para la conducción de síntesis rigurosas de evidencia.

Cuando nos referimos a políticas basadas en evidencias, estamos hablando del proceso político que ayuda a los planificadores a tomar decisiones mejor informadas, poniendo la mejor evidencia disponible en el centro del debate. Lo que llamamos “evidencia” puede incluir la información producida mediante sistemas de evaluación, investigación académica, experiencias, opinión de expertos y buenas prácticas correspondientes a un sector en específico.

Ventajas y limitaciones de las síntesis rigurosas de evidencia en términos de incidencia

Hacemos un breve recuento conceptual de algunos tipos de investigaciones abordadas en esta guía. Como se ha discutido, las revisiones narrativas son investigaciones no sistemáticas que comparan y analizan los resultados de varios estudios seleccionados por el autor sin criterios explícitos definidos. Por el contrario, las revisiones sistemáticas son investigaciones altamente reproducibles que permite comparar y analizar los resultados de varios estudios. Estas cuentan con un protocolo que define los pasos de la revisión, la estrategia de búsqueda de información y los criterios para evaluar la calidad de estudios incluidos. Por su parte, el meta-análisis es una revisión sistemática en la cual se combinan los resultados de varios estudios y que busca resumir en un valor numérico toda la evidencia relacionada a un tópico específico.

La gestión pública basada solo en evidencias estaría deliberadamente divorciada del pensamiento crítico dado que típicamente no responde el “¿por qué?” del fenómeno bajo estudio. Por tal motivo, en la mayoría de los casos puede requerirse una combinación del trabajo basado en evidencias, el análisis crítico basado en una teoría y la elaboración de revisiones realistas que incluyan una dimensión tanto cualitativa como cuantitativa para ir más de allá de saber si el tipo de intervención en cuestión funciona o no.

Al respecto, es necesario que las revisiones incluyan un análisis explicativo sobre

aspectos detallados del tipo de intervención que se está estudiando. Por ejemplo, el análisis debe centrarse en qué funciona, para quiénes, en qué circunstancias y cuáles son los mecanismos de acción o transmisión de efectos mediante los cuales la misma está surtiendo su efecto. Para ello, la intervención debe contar con una teoría de cambio bien diseñada y combinar la comprensión teórica con la evidencia empírica para encontrar la relación entre el contexto en el cual la intervención es aplicada, los mecanismos por el cual trabaja y los resultados que se podrían desprenderse de ella. De esta manera, será posible aportar información a los tomadores de decisión sobre cómo implementar intervenciones más efectivas.

No obstante, la escasa disponibilidad de tiempo, la falta de capacidad técnica y logística, y/o barreras ideológicas en la interacción de organizaciones son indudablemente factores predominantes que pueden limitar la ejecución exitosa de una síntesis rigurosa. Estas pueden suceder dado que las síntesis rigurosas son profundamente minuciosas y exhaustivas, por lo que son por naturaleza de muy larga duración y deben ser necesariamente llevadas a cabo por un personal técnico altamente especializado.

Por otro lado, la labor de incidencia en las síntesis rigurosas suele realizarse en un entorno donde interactúan una serie de elementos condicionantes y factores inherentes a las organizaciones involucradas en el proceso. No son escasos los retos que se presentan en el camino, los cuales pueden poner en riesgo el proceso de difusión e incidencia de una revisión de síntesis rigurosa en el diseño de políticas públicas. Elementos que son considerados determinantes, como la conflictividad o divergencia de ideas, la existencia de intereses contrapuestos y una estructura institucional compleja dificultan la consecución de objetivos comunes y por ende deben ser ampliamente comprendidos en fases tempranas del estudio.

Figura n° 18: Contexto en el cual busca incidir las revisiones sistemáticas



A continuación se presentan algunos ejemplos internacionales donde se ha hecho uso de síntesis rigurosas para avanzar en agendas de política y realizar mejoras sustantivas en el diseño y manejo de intervenciones:

Ejemplos internacionales del uso de evidencia rigurosa

Internacionales

- **MINEDU LAB y MIDIS LAB (Perú):** son laboratorios de innovación costo-efectiva que funcionan como una herramienta para identificar innovaciones de bajo costo en las políticas educativas y sociales que puedan ser piloteadas y evaluadas haciendo uso de datos administrativos existentes. Iniciativas como estas hacen uso extensivo de evidencia basada en síntesis rigurosas.
- **Programas de transferencias condicionadas o “Conditional Cash Transfers (CCTs)”:** se han llevado a cabo numerosas evaluaciones de impacto en este tema alrededor del mundo. Hace ya algunos años se concluyó que lograban algunos efectos positivos en el consumo y en la tasa de asistencia escolar.
- **Programas de alimentación escolar:** Estos programas suelen caracterizarse por tener una evidencia escasa y contradictoria.
- **Programas nutricionales:** se considera que la calidad en su implementación ha ido evolucionando positivamente sobre la base de la evidencia disponible.
- **Programas de entrenamiento laboral y emprendimiento:** evidencia encontrada a través de revisiones de que esta clase de intervenciones ayudan a atenuar la problemática de desigualdad de género y fomentan el empoderamiento femenino.

Requisitos conducentes a lograr altos niveles de incidencia

La realización de la síntesis rigurosa y su posterior fase de difusión e incidencia son procesos separados que poseen requerimientos diferenciados. A continuación se mencionan algunos criterios o requisitos que contribuyen notablemente a una mayor influencia de la investigación en las políticas:

- **Relevancia:** Abordar una pregunta con implicancias de política interesantes y relevantes
- **Contextualización:** Tomar en consideración el contexto social, económico, político, organizacional y cultural.
- **Claridad:** Elaborar mensajes claros y que conduzcan a acciones o soluciones concretas
- **Factibilidad:** proveer lecciones y recomendaciones factibles, costo-efectivas, sencillas y políticamente implementables.
- **Oportunidad:** Buscar ventanas de oportunidad para la cooperación, la colaboración y el aprendizaje mutuo.

Para aumentar la probabilidad de que una síntesis rigurosa efectivamente logre o contribuya a un cambio en las políticas públicas, esta debe contar con una serie de

documentos que simplifiquen el mensaje central de la revisión, faciliten su difusión y multipliquen su influencia en entornos gubernamentales y de alta dirección que tengan la capacidad ejecutiva de llevar a cabo tales cambios. Una síntesis con altos niveles de incidencia suele contar con 3 tipos de documento para hacer llegar su mensaje:

- **Informe completo:** suele constar de aproximadamente 200 páginas. Incluyen fuentes, bases de datos y anexos.
- **Informe resumen tipo artículo con fines de publicación:** Carece de componentes técnicos y posee una longitud aproximada de 50 páginas.
- **“Policy briefs”:** condensan evidencia global y local derivada de revisiones sistemáticas para informar al diálogo de políticas y el debate público. Están compuestos por una descripción del problema de política, un resumen condensado de lo que sostiene la evidencia, posibles alternativas para solventar el problema, consideraciones sobre potenciales barreras de implementación de tales alternativas y estrategias para sortearlas.

Plan de comunicaciones, incidencia e involucramiento de actores

Es fundamental desarrollar estos planes desde el inicio de la síntesis rigurosa. Estos deben considerar el contexto, los riesgos y las oportunidades que se despliegan de las interacciones con grupos de interés, beneficiarios y hacedores de política. En este sentido, debe involucrarse en la medida de lo posible a los actores o “stakeholders” relevantes. Esto implica hacer primero una tipología o clasificación de potenciales interesados, para posteriormente incluir la participación de al menos 1 grupo de cada tipo, de manera de que se escuchen y canalicen sus intereses y preocupaciones por igual.

El plan de comunicaciones puede ser una herramienta profundamente útil para esta causa. Un buen plan de comunicaciones debe incluir preguntas y mensajes clave centrados en responder aspectos como: ¿Por qué es útil la revisión?, ¿Cuáles son los posibles resultados y recomendaciones de política que pudieran derivarse?

Del mismo modo, la estrategia de comunicación debe hacer uso de múltiples canales de comunicación:

- Eventos y reuniones con los grupos de interés
- Presentaciones, talleres y seminarios
- Policy briefs
- Reportes cortos en lenguaje sencillo y comprensible para todos
- Blogs y redes sociales
- Videos
- Artículos en medios de comunicación social
- Boletines informativos por medio de difusión electrónica

La importancia y el rol del comité consultivo

El comité consultivo es una instancia organizacional de carácter técnico creada con fines de asesoría y consulta. Suele estar conformado por un conjunto de entre 5 y 10 especialistas, expertos, financiadores, hacedores de política y propulsores o “paladines” relacionados al problema social que las intervenciones contribuyan a solucionar.

El comité consultivo, o estructuras de tipo similar, suele ser una instancia requerida por donantes importantes o instituciones reguladoras o financiadores del sector público como ministerios y demás áreas de alta dirección gubernamental, en tanto que representa un mecanismo formal y continuo para indicar las necesidades usuarios clave y provee consejo experto alrededor de la temática, contextos y metodologías empleadas.

Adicionalmente, el consejo consultivo puede contribuir en la definición de las preguntas de la investigación, comentar el protocolo y reportes preliminares, identificar estudios, proporcionar retroalimentación en el análisis y ejercer una influencia en el plan de comunicaciones y en el proceso final de difusión. A continuación se enumera una lista de preguntas que podrían considerarse clave para los miembros de tal comité, así como un ejemplo de conformación de comité consultivo para una revisión sistemática realizada por 3ie sobre programas de capacitación laboral y vocacional en mujeres en la tabla n° 14 (Chinen et al, 2017):

- ¿Será interesante y útil para las audiencias principales (académica y para incidencia)?
- ¿En qué intervenciones nos debemos enfocar? ¿En qué variables de resultado?
- ¿Qué grupos poblacionales incluir?
- ¿Qué grupos específicos debemos analizar? ¿Es relevante considerar variables de género, etnicidad o raza?
- ¿Qué información será relevante para los hacedores de política? ¿Cuáles aspectos cualitativos se deben considerar?
- ¿Qué actores clave debemos contactar? ¿Cómo difundir los resultados? ¿Dónde? ¿De qué forma?

Tabla n° 14: Ejemplo de Comité Consultivo. Chinen, et al (2017)

Nombre	Institución	Cargo	Rol
Mona Marshy	Global Affairs Canada	Asesor de políticas	Especialista Senior en igualdad de género
Dragana Vujic	Global Affairs Canada	Analista Senior de políticas	
Patience Stephens	UN Women	Director	Topic expert
Shravanti Reddy	UN Women	Especialista en Evaluación	Topic expert
Seemin Qayum	UN Women	Policy Advisor	Líder/influyente
Fathema Khatton	BRAC - Bangladesh	Asociado Senior de investigación	Experto temático
Hugo Ñopo	IADB/GRADE	Investigador principal	Experto temático
Carolina Trivelli	IEP	Investigador principal	Hacedora de política/ funcionaria pública

Bibliografía de la sesión

- Chinen, M., Balarin, M., Alcázar, L., De Hoop, T., & Sennett, J. (2017). Vocational and business training to improve women's labour market outcomes in low- and middle-income countries. Campbell Collaboration.